



**HAL**  
open science

# High-throughput sequencing techniques for the detection and survey of marine non-indigenous species : a valuable addition to traditional methods ?

Marjorie Couton

► **To cite this version:**

Marjorie Couton. High-throughput sequencing techniques for the detection and survey of marine non-indigenous species : a valuable addition to traditional methods ?. Ecosystems. Sorbonne Université, 2020. English. NNT : 2020SORUS039 . tel-03180937

**HAL Id: tel-03180937**

**<https://theses.hal.science/tel-03180937>**

Submitted on 25 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sorbonne Université

Ecole doctorale 227

Sciences de la nature et de l'Homme : évolution et écologie

*UMR7144 Adaptation et Diversité en Milieu Marin*

*Equipe Dynamique de la Diversité marine*

## **High-throughput sequencing techniques for the detection and survey of marine non-indigenous species: a valuable addition to traditional methods?**

*Le séquençage haut-débit pour détecter et étudier les espèces marines non-indigènes : un complément pertinent aux méthodes traditionnelles ?*

Par Marjorie Couton

Thèse de doctorat en Ecologie Marine

Sous la direction de Frédérique Viard et Thierry Comtet

Présentée et soutenue publiquement le 29 octobre 2020

Devant un jury composé de :

Pr TURON Xavier, Professeur, Centre d'études avancées de Blanes

Rapporteur

Dr RODRIGUEZ-EZPELETA Naiara, Directrice de recherche, AZTI

Rapporteuse

Dr NUNES Flavia, Directrice de recherche, Ifremer

Examinatrice

Pr DESTOMBES Christophe, Professeur, Sorbonne Université

Examinateur

Dr VIARD Frédérique, Directrice de recherche, CNRS

Directrice de thèse

Dr COMTET Thierry, Chargé de recherche, CNRS

Co-directeur de thèse



## Remerciements

En tout premier lieu, j'aimerais remercier mes superviseurs, Frédérique Viard et Thierry Comtet pour leur soutien et leurs encouragements. Toutes nos discussions, formelles et informelles, ainsi que leurs conseils et remarques avisés, m'ont permis de ne pas me reposer sur mes acquis et de toujours me remettre en question. Tous les progrès réalisés et la maturité acquise lors de ces trois dernières années n'auraient pas été possibles sans vous. Bien plus qu'un support scientifique, vous m'avez permis de réaliser cette thèse dans une ambiance bienveillante et amicale, ce qui représente, à mon sens, la majeure partie d'un travail réussi.

Contrairement à ce qui pourrait transparaître au travers d'un manuscrit, une thèse est une aventure collective et aucun des résultats décrits ici n'auraient pu être acquis sans l'aide d'un grand nombre de personnes. C'est le cas de Claire qui m'a épaulé tout au long de ce périple, de l'échantillonnage jusqu'au séquençage. Ses compétences et son expérience au laboratoire m'ont évité pas mal de faux pas. C'est également le cas de Laurent qui a réalisé toutes les identifications morphologiques présentées et dont l'expérience du terrain a été d'une aide précieuse. Enfin Erwan m'a également accordé de son temps pour jeter un œil avisé de bioinformaticien sur mes travaux et me conseiller dans mes analyses. A vous trois un, grand merci.

J'aimerais également rendre hommage à tous ceux qui m'ont aidée pour les échantillonnages. Merci aux plongeurs, Mathieu, Wilfried et Yann, pour leur indéfectible bonne humeur, merci aux marins, François, Gilles et Noël pour leur gentillesse, malgré des conditions parfois difficiles, et merci également à Stéphane et Jérôme pour patience et leur efficacité. Je suis aussi reconnaissante envers Marion et Aurélien pour leur aide au laboratoire, envers l'équipe du CRBM et notamment Gaëtan pour la mise en place de matériel pour la remontée des panels de recrutement, ainsi qu'à la plateforme Genomer, et surtout Gwenn et Erwan, pour leur aide lors des séquençages.

Enfin, parce que les interactions sociales sont très importantes pour un travail efficace (cf. le confinement), je tiens à remercier tous ceux avec qui j'ai pu passer des moments de discussions informelles et de détente pendant ces trois ans. Merci à toute l'équipe ABiMS pour leur accueil chaleureux, leurs pauses café/gâteaux de 10h30, les blagues du mercredi et surtout pour ne pas m'avoir retiré de l'alias igm3 pour que je puisse continuer à venir m'incruster de temps en temps après avoir déménagé dans un autre bâtiment. Merci à Delphine pour nos randonnées, parfois un peu plus longues que prévues à cause de notre mauvais sens de l'orientation, merci à Anaëlle et Anaïs pour les discussions autour d'un café et pour m'avoir supportée (dans les deux sens du terme) lors de la dernière ligne droite, et merci à Romain pour nos balades quotidiennes qui m'ont permis de me changer un peu les idées pendant la rédaction de ce manuscrit. Et parce que travailler pendant le confinement n'a pas été une mince affaire, merci à mes sœurs, Gaëlle et Solène, pour m'avoir permis de conserver ma santé mentale pendant ces deux mois compliqués.

Et pour tous ceux que j'ai pu oubliés et qui ont participé, de près ou de loin, à la création de ce travail, MERCI !



# Table of contents

<b>Introduction</b>	<b>1</b>
<hr/>	
1. Biological invasions: one facet of the global change .....	3
a. Biological introductions – definition and global overview .....	3
b. A dynamic process (from introduction to invasion, and failure) .....	5
c. Consequences and management of biological invasions .....	7
2. Some important properties of marine biological introductions .....	10
a. Introduction vectors responsible for specific patterns and processes .....	10
b. Marinas and ports as invasion hubs .....	13
3. Molecular tools for the biomonitoring of non-indigenous species .....	16
a. A furnished DNA-based toolbox with many applications .....	17
b. NIS identification through DNA barcoding: general principle .....	20
c. Producing and processing metabarcoding data .....	22
Thesis objectives .....	32

## **Chapter I : High-Throughput Sequencing on preservative ethanol is effective at jointly examining infra-specific and taxonomic diversity, although bioinformatics pipelines do not perform equally**

---

Preamble .....	39
Abstract .....	41
Introduction .....	42
Materials and methods .....	43
a. Case study & sampling .....	43
b. Sanger sequencing on individual zooid (SSIZ) .....	45
c. High-throughput sequencing on assemblages (HTSA) .....	45
<i>Sample processing</i> .....	45
<i>Reads processing</i> .....	47
d. Data analyses .....	48
<i>Assignment</i> .....	48
<i>Haplotype comparison</i> .....	48
<i>Diversity indices</i> .....	48
Results .....	49
a. Sanger sequencing on individual zooid (SSIZ) .....	49
b. Species assignment .....	49
c. Pipeline performance for HTSA-based haplotype detection .....	50
d. Population diversity indices .....	54
Discussion .....	55

a. Ethanol-based DNA is a valid non-destructive alternative to bulkDNA, even after several months of storage .....	55
b. Specific primers can improve the quantitative use of HTSA data .....	56
c. Careful choice of bioinformatics pipeline is needed to examine genetic diversity .....	57
d. Improving haplotype detection – a matter of compromise .....	58
References .....	60

## **Chapter II : How effective is metabarcoding for studying communities living in marinas?** **65**

---

Preamble .....	67
----------------	----

### **Chapter II.1: Non-indigenous and native species from biofouling communities in marinas: the (un)detected fraction from environmental DNA metabarcoding .....**

Abstract .....	69
Introduction .....	69
Materials and methods .....	71
a. Sampling .....	71
b. DNA extraction .....	73
c. Library preparation and sequencing .....	74
d. Reads processing and filtering .....	76
e. Taxonomic assignment .....	76
f. Comparison between quadrats and metabarcoding datasets .....	77
Results .....	78
a. Reads processing and taxonomic assignment .....	78
b. Comparison of quadrats (morphology-based) and eDNA metabarcoding datasets .....	80
c. Effect of the sampling strategy on taxa detection .....	84
Discussion .....	86
a. Taxon detection with eDNA metabarcoding is effective but major flaws challenge NIS survey .....	87
b. The added value of environmental DNA metabarcoding .....	89
c. Improving sampling strategy and data processing to overcome limitations .....	91
d. Some issues are still unresolved .....	93
References .....	95

**Chapter II.2 : Marina communities are not homogenous at a regional scale nor over time, as jointly shown by eDNA metabarcoding and quadrat analyses .....100**

Abstract .....	101
Introduction .....	101
Material and methods .....	104
a. Quadrat and eDNA datasets .....	104
b. ASVs clustering and taxonomic assignment .....	106
c. Alpha and beta-diversity analyses .....	107
d. Non-indigenous species diversity and contribution .....	109
Results .....	109
a. OTUs distribution and taxonomic assignments .....	109
b. Alpha diversity patterns .....	110
c. Community dissimilarities .....	112
d. Non-indigenous species contribution to community patterns .....	117
Discussion .....	122
a. Spatial differences suggest that natural processes are not fully overcome by human-mediated processes .....	122
b. Marinas are highly disturbed habitats that may lead to fast community turn-over .....	124
c. The impact of non-indigenous species in marina community structure .....	125
d. eDNA metabarcoding as a monitoring tool in marinas .....	126
References .....	129

**Chapter III : Can metabarcoding detect hidden non-indigenous species in the wild? 133**

---

Preamble .....	135
----------------	-----

**Chapter III.1 : Are fouling communities in marinas, and particularly the non-indigenous fraction, singularly different from communities on natural rocky habitats? .....136**

Abstract .....	137
Introduction .....	137
Materials and methods .....	140
a. Survey design .....	140
b. Processing of settlement plates .....	141
c. DNA extraction, amplification and sequencing .....	142
d. Denoising and taxonomic assignment .....	143
e. Diversity analyses .....	144
Results .....	145



a. Identification of taxa across methods .....	145
b. Taxa found only in the marina .....	149
c. Alpha- and beta-diversity at the bay scale .....	151
Discussion .....	157
a. Marina assemblages are not equivalent to communities observed in natural environments .....	157
b. Are non-indigenous species observed outside the marina? .....	159
References .....	162

### **Chapter III.2 Metabarcoding on planktonic larval stages: an efficient approach for detecting and investigating life cycle dynamics of benthic aliens .....166**

Abstract .....	167
Introduction .....	168
Materials and methods .....	171
a. Sampling and DNA extraction .....	171
b. Molecular analyses .....	171
c. Data processing .....	172
d. Taxonomic assignment .....	173
e. Tests for amplification failure .....	175
f. Cross-validation with traditional methods .....	175
Results .....	177
a. Overall taxonomic assignment .....	177
b. Identification of non-indigenous species .....	180
c. Temporal variations .....	182
d. Comparison of the metabarcoding results with other datasets .....	184
Discussion .....	186
a. Zooplankton metabarcoding: an efficient tool to detect non-indigenous benthic species .....	187
b. An efficient method to detect both long and short dispersers .....	189
c. Detection of previously unreported species .....	190
d. Read counts as a proxy for larval abundance and NIS reproductive success .....	192
References .....	195

## **Conclusion and perspectives**

**205**

1. Is metabarcoding ready for marine NIS detection? .....	208
2. Added value of amplification-based HTS approaches in studying the introduction process .....	212

3. The future of molecular methods for the study of biological introductions .....215

References from the general introduction and conclusion .....218

List of tables .....231

List of figures .....234

Appendices.....241



# Introduction



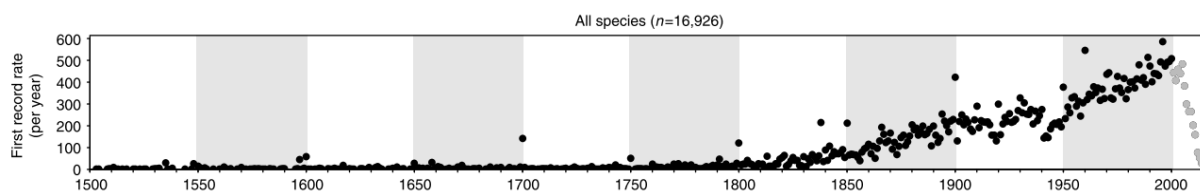


# 1. Biological invasions: one facet of the global change

The constant increase in human population and the overexploitation of natural resources have led to the destruction or disturbance of ecosystems throughout the world. Human-mediated perturbations are associated with climate change, biodiversity loss, or changes in species distributions, and it has become clear that measures need to be taken in order to stop, or at least slow down this global environmental change. Public awareness is growing on several aspects, such as pollution or deforestation, and more efforts are put into the limitation of human impacts on these elements. Other facets, however, such as biological introductions, receive much less attention from the general public despite being a worldwide issue and a major cause of global environmental change (Vitousek, et al. 1997).

## a. Biological introductions – definition and global overview

The transport of species outside of their natural geographic range by human activities have started thousands of years ago, with the beginning of human migrations and commerce. Domesticated species, such as cereals or cattle, were displaced from their original distribution area over distances far greater than possible through natural dispersal (Mack, et al. 2000). Then, the intensity of human travels and international trade has started to increase substantially from the 15<sup>th</sup> century, with the discovery of America, and the number of biological introductions has subsequently risen from this time. Finally, the last two centuries have seen the development of new, fastest ways of travel and the growing globalization has accelerated the rate of new introductions (Fig.1; Seebens, et al. 2017). Nowadays, this phenomenon is so widespread that few habitats on earth remain free of introduced species (Mack 2000).



**Figure 1** Global temporal trend in first record rate (dots) with the total number of established alien species during the time period considered given in parentheses. Data after 2000 (grey dots) are incomplete because of the delay between sampling and publication. Figure from Seebens, et al. (2017).

Through the history, numerous species were intentionally transported outside of their natural distribution range. Some of them were utilized as food resources such as corn (*Zea mays*), which originates from Central America and is now the most cultivated cereal throughout the world (Mangelsdorf 1983). Others were used for biological control, and were introduced to regulate the populations of other species. This is, for example, the case of the generalist parasitoid fly *Compsilura concinnata*, which was repeatedly introduced in North America from 1906 to 1986 as a biological control agent for 13 different species of insects (Boettner, et al. 2000). Species were also introduced voluntarily for recreational purposes, such as many ornamental flower species. Nevertheless, a higher proportion of biological introductions were done unintentionally. In these cases, species can be transported, without people's knowledge as stowaways on boats, planes or with any kind of goods. A famous example is the introduction of the Asian tiger mosquito in the United States in the 1980s with imported automobile tires (Hawley, et al. 1987). Many of these accidental introductions are also "hitchhikers" of deliberate introductions, such as parasites and pathogens, as was the case with the introduction of the Pacific oyster in Europe (Wolff and Reise 2002).

Contrary to intentional introductions for which information on the date of first introduction and the native range of the species is most often available, no such data can be collected for inadvertently displaced species. In such cases, determining their status as native or introduced species can be difficult, especially when they are found in many places throughout the world (i.e. cosmopolitan species). Moreover, the long history of species displacement can be an additional obstacle to their status identification because their arrival in new areas can predate biodiversity surveys in this region. Carlton (1996) proposed to name the species for which the native *versus* introduced status is uncertain, cryptogenic species. The same species can, however, have an introduced status in some part of the globe but have a cryptogenic status in other parts. For example, the calanoid species *Acartia (Acanthacartia) tonsa*, which is present in every ocean around the world, was reported for the first time in the Lagoon of Venice (Mediterranean) in 1985. It is thus classified as an introduced species there, despite having a cryptogenic status in other parts of the globe because its native range could not be resolved (Camatti, et al. 2019). Similarly, the tunicate species *Corella eumyota*, first described from Valparaiso, Chile (Traustedt, 1882) is widespread in the Southern hemisphere, where it has a cryptogenic status but has been reported in the late 20<sup>th</sup> century in Europe where it has an introduced status (Dupont, et al. 2007).

## b. A dynamic process (from introduction to invasion, and failure)

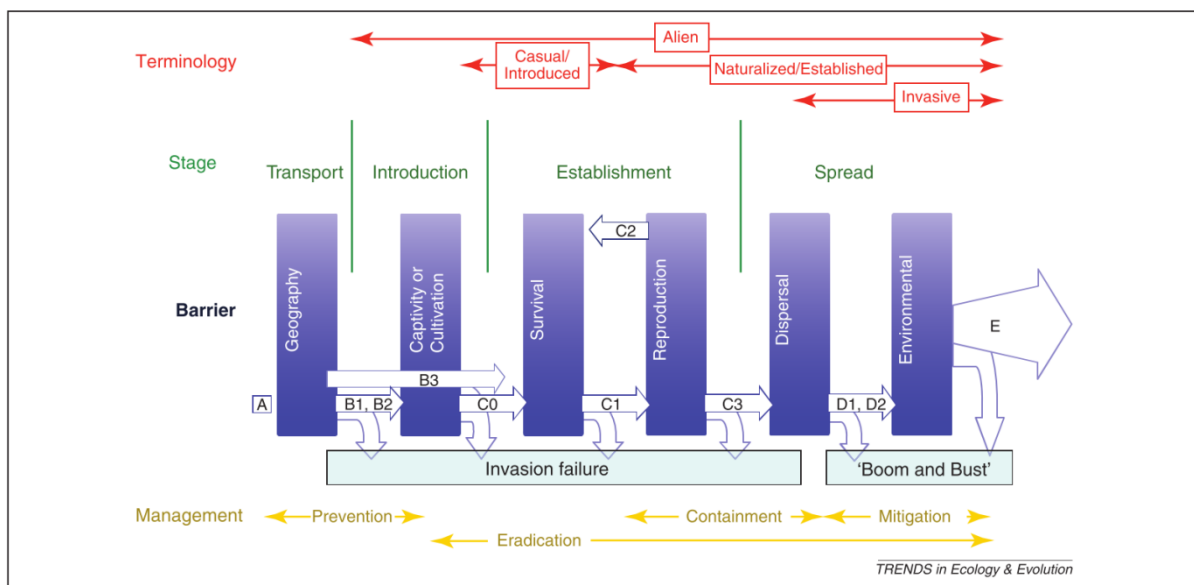
Introduced species face many obstacles in the colonization of new habitats and the process by which they might become established is dynamic and reversible. Among the many introduced organisms around the globe, only a very small amount will thrive in their new environment. This statement has been supported by empirical surveys and studies, and is in agreement with the statistical prediction, known as the “tens rule”, proposed by Williamson and Fitter (1996), in which each transition from one stage to the next (escaping, establishing, becoming a pest) has a 10% probability of realising. Entering a specific stage is, however, not final and the process can, at any time, end up in an invasion failure. In addition, time matters: some of the introduced species can stay at a low abundance for a very long time before starting to expand rapidly. The Brazilian pepper (*Schinus terebinthifolius*), for example, introduced in Florida in the 19<sup>th</sup> century, only started to widely spread in the early 1960s and is now very abundant (Morgan and Overholt 2005). Others might show a rapid population “burst” before drastically declining or even disappearing. This phenomenon, sometimes called “boom and bust” cycles (Strayer, et al. 2017), has been observed for several introduced species. The yellow crazy ant *Anoplolepis gracilipes*, for example, considered as a highly invasive species throughout the world, is introduced in Australia. Cooling and Hoffmann (2015) demonstrated the decline of four populations of this species and the extinction of three others without human intervention.

Understanding the mechanisms leading to the success of few introductions and to the failure of many others has been the focus point of several studies. Some have investigated species invasiveness, which is a set of features that might grant introduced species a high invasive potential (e.g. Nyberg and Wallentinus 2005 or Fournier, et al. 2019). Others have explored habitats invasibility which corresponds to environmental characteristics that might make them more vulnerable to invasions (Catford, et al. 2012 and references therein). Cassey, et al. (2018) showed that propagule pressure (i.e. the amount of individuals released in the environment during the introduction and the number of introduction events) is correlated to the success of an introduction. Moreover, environments subject to major anthropogenic disturbances are thought to be more susceptible to be invaded, and the impact of climate change on ecosystems might favour future invasions (Aronson, et al. 2007; Hellmann, et al. 2008). Other processes have been suggested to either explain introduction failures in some ecosystems such as biotic resistance (i.e. high native richness that prevent the establishment of introduced species; Stachowicz, et al. 1999) or to justify invasion success such as the invasional meltdown (i.e. the fact that the



presence of already established non-indigenous species might favour the establishment of other introduced species; Simberloff and Von Holle 1999). All these hypotheses, however, need additional research in various environments and for different taxonomic groups in order to be fully validated.

Conceptualizing a theoretical framework that encompasses all aspects of the complexity and dynamic of the introduction process is challenging. Several attempts have been made by scientists working with different taxa in different environments (Gurevitch, et al. 2011 and references therein), which resulted in a multitude of terminologies and definitions. Blackburn, et al. (2011) proposed a unified framework for biological invasions. This framework is applicable to all types of introductions, whatever the organism or the environment considered, but is only suited for human-mediated species displacement and does not take into account natural dispersal or range expansions resulting from human activities (e.g. northward migration of temperate species with climate change). In their approach, these authors attempted to combine a “stage framework” and a “barrier framework” (Fig. 2), each of them mostly used by animal and plant invasion biologists, respectively, and proposed a categorization scheme combining stages and barriers.



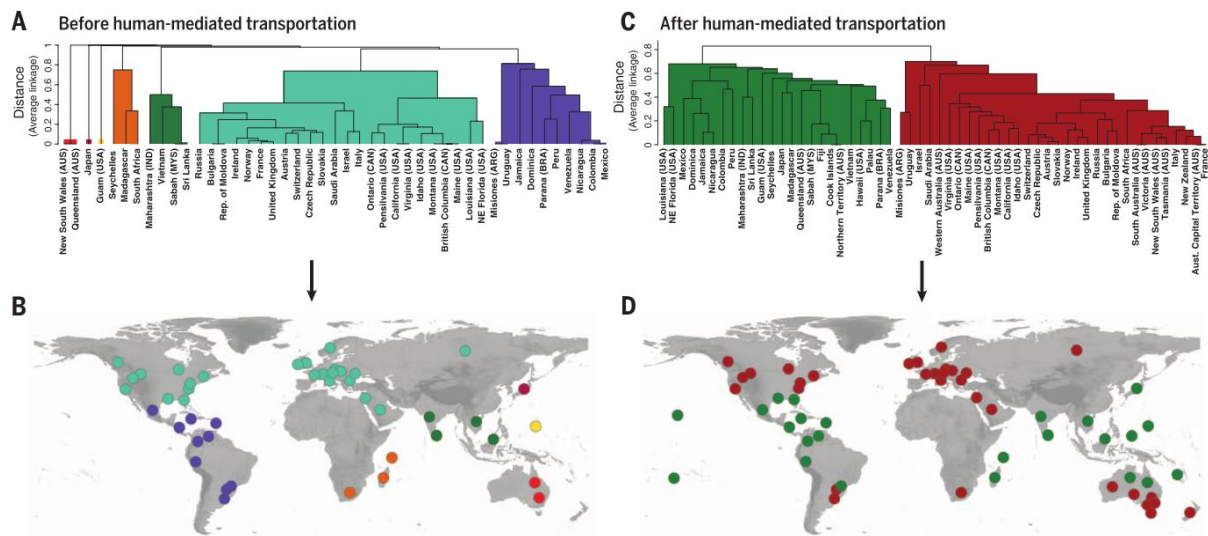
**Figure 2** Unified framework for biological invasions proposed by Blackburn, et al. (2011). An alphanumeric code (inside the white arrows) categorizes the species along an invasion pathway.

Various terminologies have been employed to describe introduced species in the literature, such as alien, exotic, neophytes, introduced or immigrant, sometimes referring to a particular stage of the introduction process. Evaluating the stage and categorizing the status of a particular species, with reference to the Blackburn, et al. (2011) framework, is difficult, because it requires an accurate assessment of its distribution, impacts etc. For this reason, the terms “introduced species” or “non-indigenous species (NIS)” will be used, in this thesis, to describe species in their introduced range, whatever their situation in the process. The term cryptogenic species (Carlton 1996) will be used if the native range of the species could not be determined and possibly include the study area. Finally, these terms will be opposed to the term native species for organisms in their native range.

### **c. Consequences and management of biological invasions**

When species are introduced into a new environment, they can have numerous ecological and economic impacts. Several studies have tried to classify species according to their harmfulness or the major disruptions they are responsible for (e.g. Ojaveer, et al. 2015; Bacher, et al. 2018) but these are subjective categories and can be quite different depending on the taxonomic group or the environment considered. Among the local consequences of biological introductions are the disruption of ecosystems functioning, such as the decrease in silicate due to the invasive gastropod *Crepidula fornicata* (Ragueneau, et al. 2002) or the creation of new habitats by ecosystem engineer as is the case for the Pacific oyster *Crassostrea gigas* (Markert, et al. 2009). Introduced species can also have local direct or indirect effects on native biota that might lead to biodiversity loss or even species extinction, such as competition for food or space with resident species (Gurevitch and Padilla 2004; Blackburn, et al. 2019). Non-indigenous species (NIS) have also been shown to impede the provision of ecosystem services in marine coastal areas. A recent pan-European review showed that 56 out of 87 (65%) NIS for which data were available alter ecosystem services (Katsanevakis, et al. 2014), especially food provisioning, ocean nourishment, recreation, tourism and lifecycle maintenance. At a global scale, one of the major impacts of biological introductions on biodiversity is the break of biogeographical barriers, and the redefinition of biogeographic boundaries. Distribution ranges are no longer only defined according to the species natural dispersal ability and by environmental factors (Capinha, et al. 2015). The constantly increasing rate of new introductions is responsible for a worldwide biotic

homogenization where past dissimilarities between, sometimes very distant, communities tend to decrease, as shown for several terrestrial communities (Fig. 3). This process can be observed in terms of similarity of species between assemblages (taxonomic homogenization) as well as in functional traits (functional homogenization) and molecular diversity (genetic homogenization) (Olden and Rooney 2006).

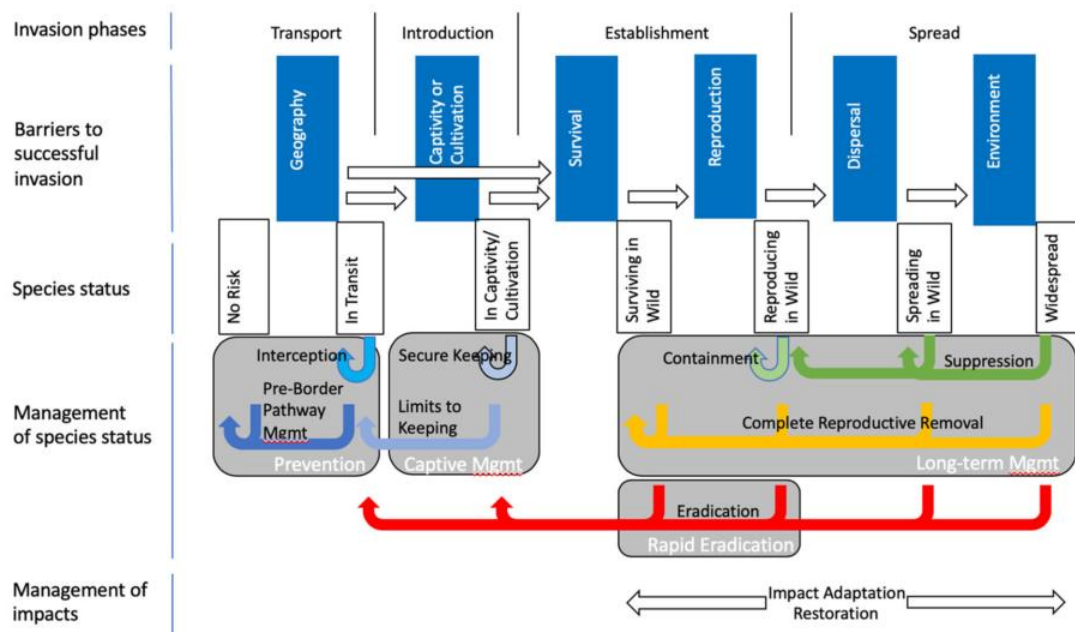


**Figure 3** Dendrogram and map of compositional similarities among lists of introduced terrestrial gastropods before (A and B) and after (C and D) dispersal by humans. Colors indicate main clusters identified by the dendrogram and their corresponding locations in the world map. Figure from Capinha, et al. (2015). This figure highlights a redistribution of the species assemblages, and thus biogeographic boundaries after human-mediated transportation.

As one of the major components of environmental global change, measures need to be taken in order to limit the impact of introduced species on native biota. When introduced populations have grown so largely that they become a threat to other organisms, it is mostly impossible to eradicate them from the environment. A more practical approach is thus to avoid any new introduction, both intentional and unintentional. Several countries require that a risk assessment must be performed in order to evaluate the invasiveness of a species before it is introduced voluntarily. The reliability of these evaluations is, however, quite limited because of the lack of generalization that can be made across taxonomic groups and environments, as well as important knowledge gaps regarding those introduced species that prevent accurate risks' assessment (Simberloff, et al. 2005; Ojaveer, et al. 2015). Measures

have also been taken to limit unintentional introductions such as quarantines but unexpected species arrivals are hard to contain (Mack, et al. 2000).

Since organisms are still being displaced around the world, detecting their arrival as fast as possible and monitoring their population dynamics is of the utmost importance in order to be able to anticipate their impact on ecosystems. It is easier to control NIS populations when they are in low abundance in the environment, and the biology of the introduced species must be sufficiently understood so that a valid management strategy can be planned. Prevention is the most effective way to limit the ecological and socio-economic impacts as well as the management costs of NIS (Simberloff, et al. 2013). Moreover, ecological baselines need to be set for the ecosystem in order to be aware of any impact that could be related to an introduced species. Finally, the reproductive and dispersal ability of NIS are criteria for defining their invasive stage (*sensu* Blackburn, et al. 2011) and the assessment of both characteristics is essential for an efficient management. Altogether, management actions require to be adapted to the stage of the introduction process, and to the likelihood of transition from one stage to the other, as recently proposed by Robertson, et al. (2020) (Fig. 4) with reference to the unified framework proposed by Blackburn, et al. (2011) (Fig. 2).



**Figure 4** Management actions specific to the stage of the introduction process, with coloured arrows highlighting expected changes in the species status following management actions. Figure from Robertson, et al. (2020).

The complexity of the introduction process can thus not be comprehended by the sole detection of NIS but needs to be globally evaluated, including the different life stages of the organisms of interest. Research is still needed to grasp the complexity of invasion biology which, in turn, could both support effective prevention, detection and management, and provide valuable insights into ecological, evolutionary, and biogeographic theories and concepts.

## 2. Some important properties of marine biological introductions

Many of the terminologies, concepts etc. highlighted above, obviously hold for marine system as well as for terrestrial ones. There are, however, specificities of marine systems, regarding introduction patterns and processes, important to highlight, especially regarding this thesis work.

### a. Introduction vectors responsible for specific patterns and processes

Since men have sailed the world's seas, marine species have intentionally or unintentionally been displaced. The Vikings, which were great seamen, could have been responsible for the introduction of the bivalve *Mya arenaria* from North America to Europe (Petersen, et al. 1992). With the development of international shipping traffic and new technical advances in marine vessels, the number of introduced species has dramatically increased in the last century. The number of non-indigenous species (NIS) reported in Europe, from unicellular algae to vertebrates, has almost reach 1,400 species (Nuñez, et al. 2014), more than half having established a self-sustaining population in their new range (Gollasch 2006; Katsanevakis, et al. 2013). The majority of the recorded NIS was invertebrates such as molluscs, arthropods, tunicates and annelids. The rate of introduction for marine species is incredibly high with an average, 18 years ago, of one new record every nine weeks in the world (Minchin and Gollasch 2002). This value varies greatly depending on the region considered and could even have reached a rate of one new record every three weeks in Europe for the period 1998-2000. Although not all newly introduced species are

able to establish, these alarming facts call for a rapid regulation of transportation vectors.

Carlton (2001) reported 14 categories of introduction vectors related to human activities for marine species. They include trading activities (e.g. shipping, aquarium pet industries), natural resources exploitation (e.g. drilling platforms), leisure activities (e.g. diving equipment, leisure boating) and education and research. Introduction pathways and vectors responsible for marine introductions, following the terminology of Ojaveer, et al. (2014) (Table 1), are thus extremely diverse. Nevertheless, shipping, canals and aquaculture have been regularly targeted as the most important ones (Molnar, et al. 2008; Nuñez, et al. 2014; Ojaveer, et al. 2014).

**Table 1** Pathways and vectors (i.e. physical mechanisms) of introduction in marine systems. Table from Ojaveer, et al. (2014).

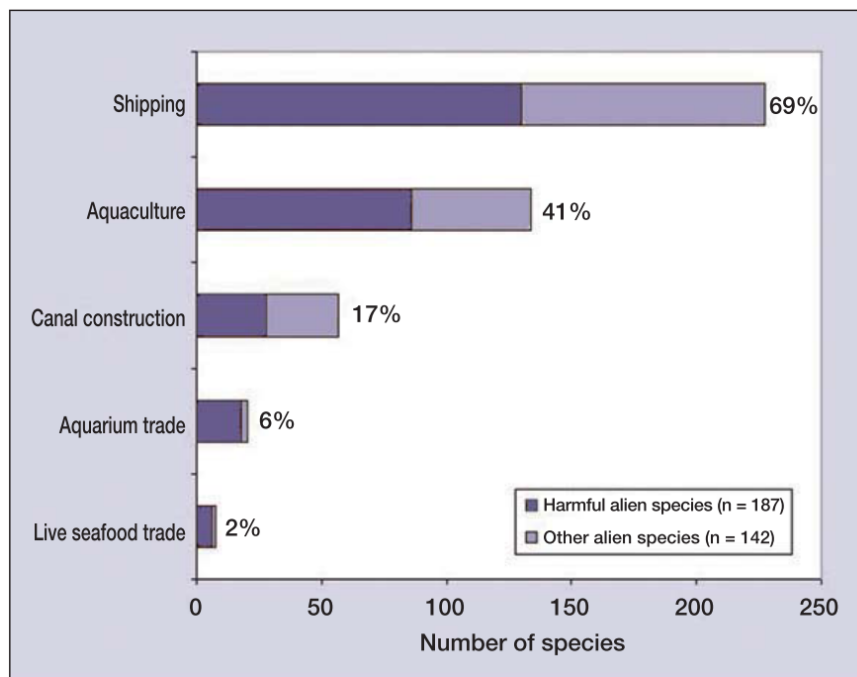
<b>Pathway</b>	<b>Vectors</b>
Vessels: ships, vessels, platforms	Ballast water and sediments, sea-chests, hull fouling
Canals	A specific canal (e.g. Suez Canal)
Aquaculture activities	Aquaculture equipment, packaging, stock movement
Aquarium trade/public aquaria	Transported water, waste discharge, direct release, packaging
Leisure activities	Angling baits, stocking, discharges, sport equipment
Live food trade	Intentional release, waste discharge, transported water
Management	Habitat management, biological control
Research and education	With equipment, intentional release, waste discharges
Wild fisheries	Fishing gear, discharges, stock movements

The vector being responsible for the highest number of introductions in marine environments is shipping, accounting for more than two third of the total amount of displaced species (Fig. 5; Molnar, et al. 2008). These numbers are not expected to decline since shipping trade is projected to be between 240 and 1209% greater in 2050, as compared to 2014 (Sardain, et al. 2019). Ships have used water as ballast since the end of the 19<sup>th</sup> century, for balance and stability of big international cargo ships. The large amount of water drawn in their ballast tanks can comprise

hundreds of species, from unicellular organisms to benthic metazoans, which will be transported over thousands of kilometres and then be released in completely new environments (Gollasch, et al. 2002). This also includes pathogens, such as the bacteria *Vibrio cholerae*, as shown by Ruiz, et al. (2000), in a seminal paper pointing the risk associated with ballast water in transporting microorganisms. The importance of ballast water as an introduction vector supported the adoption in 2004 of the International Convention for the Control and Management of Ships' Ballast Water and Sediments which, however, came into effect only in 2017. More recently, attention has been drawn on another vector related to shipping: ships' hull. Benthic sessile organisms, living fixed on hard substrates, can attach to the hull of boats as a part of biofouling. This happens particularly in "refuge areas" protected from the currents (Coutts and Dodgshun 2007). These species can thus be unintentionally transported from one location to another as stowaways (Sylvester, et al. 2011). Moreover, biofouling is not only a concern for large cargos, and may also develop on other kind of ships, such as fishing boats or recreational sailing boats (Clarke Murray, et al. 2011). The second most important vector for marine introductions, at a worldwide scale, is aquaculture (Naylor, et al. 2001). Numerous species are willingly transported into new regions to serve as food such as the Pacific oyster *Crassostrea gigas* which is now found in most parts of the globe (Molnar, et al. 2008). Nevertheless, the highest number of NIS being released from aquaculture is actually resulting from unintentional transport, such as epibionts living fixed on commercial organisms or parasites from farmed species (Streftaris, et al. 2005). Finally, the construction of canals is also an important human-mediated vector of dispersal in marine environments. For example, the opening of the Suez Canal between the Mediterranean and the Red Sea is responsible for more than half of reported NIS in the Mediterranean Sea (called Lessepsian migrants; Streftaris, et al. 2005; Galil, et al. 2018).

The particular characteristics of marine introduction vectors make their control very difficult, increasing the colonization pressure (i.e. the number of species introduced; Lockwood, et al. 2009). Most species are displaced unintentionally, and the frequency and repetition of introduction events, especially for shipping-related transports, are so important that many marine introductions are characterized by high propagule pressure (i.e. number of propagules introduced per species; Lockwood, et al. 2009). The importance of propagule pressure in marine introductions has been demonstrated by population genetics studies failing to find severe founder events in most introduced populations, and showing that the genetic diversity of marine populations is most often similar between their native and introduction range

(Roman and Darling 2007; Rius, et al. 2015; Viard, et al. 2016). For both genetic and demographic reasons, propagule pressure is expected to favour initial settlement and sustainable establishment (Rius, et al. 2015; Cassey, et al. 2018). Moreover, many marine organisms can be introduced by several vectors. Bivalves, for example, can be transported either as adults through hull fouling and *via* aquaculture, or as larvae in ballast water (Gollasch 2007). This makes prevention measures more difficult to be set-up, as different pathways and vectors need to be surveyed and regulated.



**Figure 5** Number of NIS known or likely to be introduced by the most common human-assisted pathways. Percent of total number of species in assessment (n=329) is indicated. Figure from Molnar, et al. (2008).

## b. Marinas and ports as invasion hubs

Shipping being responsible for the vast majority of marine introductions, ports and marinas (often located nearby large commercial ports) are points of entry for many non-indigenous species (NIS). The density of maritime traffic and the consistence of commercial routes increase the chances of establishment of new introduced species in these particular artificial habitats. Marinas and harbours are anthropogenic environments with particular abiotic features, such as high pollution



levels, high turbidity, reduced water flow or extensive shading (Rivero, et al. 2013). These characteristics might favour the establishment of species being more tolerant towards anthropogenic pressure over native species (Piola and Johnston 2009; Canning-Clode, et al. 2011; Lenz, et al. 2011). In the case of fouling organisms, marinas and ports also offer a wide range of artificial structures that might promote NIS settlement (Bulleri and Chapman 2004; Glasby, et al. 2007). In fact, establishment of sessile introduced species might be facilitated by contemporary adaptation to artificial habitats (here ports and marinas) that occurred in their native range, similarly to what have been suggested in terrestrial environments for agricultural pests (i.e. “Anthropogenically induced adaptation to invade” as coined by Hufbauer, et al. 2012). It could also rely on selection of particular traits during transports, adaptive to settlement on artificial substrates (Briski et al. 2018). Whatever the neutral (e.g. repeated introductions decreasing Allee effects) or selective (e.g. pre-adaptation) processes behind, numerous studies have shown that marinas and ports are composed of a high proportion of introduced species, and that their communities are very different from those observed in close natural habitats (Connell 2001; López-Legendil, et al. 2015).



**Figure 6** Pictures illustrating the species diversity (left) and population abundance (here the tunicate *Ciona intestinalis*, right) of fouling organisms attached to floating pontoons in French marinas. Photo credit: Wilfried Thomas – Station Biologique de Roscoff.

From a management point of view, ports and marinas, being invasion hubs and sustaining an important part of the introduction load in coastal ecosystems, are risk areas to survey with a high priority, and in which early detection (when

prevention fails) is critical to achieve. The detection of newly arrived species need to be done as early as possible and before the novel species had the time to reproduce, establish and increase in population density. Information on new NIS arrival is, however, not the only valuable insights that we can gain by examining marinas communities. All the specificities listed above make marinas and ports particularly interesting from a research perspective when studying the introduction processes. In addition, these habitats, despite being singularly different from close natural environments, share common features with other marinas and ports across the globe (Minchin 2006). As pointed above (pre-adaptation in the native range), this peculiarity might facilitate the establishment of NIS in new harbours from distant coastal areas, particularly when located in regions with broadly similar environmental conditions (e.g. located in temperate regions). Other ecological and evolutionary findings could also be acquired by studying marinas communities such as the understanding of the interactions between native and introduced species, the role played by previous NIS in facilitating the establishment of new NIS, or the genetic factors favouring or hindering NIS establishment.

So far, the focus has been made on the particularity of marinas to be points of entry for introduced species. However, marinas may also be major contributors to secondary spread and expansion of NIS at a regional scale (including across borders, as for instance between the French and English side of the Western English Channel; Bishop, Wood, Yunnice, et al. 2015). A large number of marine species have planktonic life stages (e.g. spores and gametes in seaweeds, larvae in invertebrates) that are the actors of their natural dispersal. When NIS survive and reproduce in their introduced range, they become able to colonize nearby environments through natural dispersal. In marine habitats, high environmental connectivity through the water currents promotes dispersal, rendering efforts to control biological invasions even more challenging. The enclosed structure of marinas, however, could limit the dispersal of planktonic organisms or larvae from marine invertebrates because of a reduced water flow. Nevertheless, marinas and ports may further serve as stepping stones for secondary dispersal *via* human-mediated transport that allows species to disperse in natural environments, or from one port to another. Marinas are drastically increasing in numbers, and are major contributors to the "ocean sprawl" (i.e. the proliferation of artificial infrastructures at sea; Duarte, et al. 2013; Firth, et al. 2016). Bugnot, et al. (2020) estimated, in 2018, that 9628 marinas are existing at a global scale, which are responsible for major physical footprints in coastal areas, through noise due to shipping. These authors, however, did not include biological footprint such as NIS spread in their study. At a more local scale, more than 470 marinas (including 114 in

Brittany and 140 along the Mediterranean Sea) are spread along the French coasts, representing more than 180 000 moorings for recreational boats. They are thus forming a dense network connected through leisure boating, which may influence the connectivity of populations established in these artificial habitats, in particular for species that can attach to hulls or ropes (Clarke Murray, et al. 2011). Connectivity patterns among marinas are supposed to contrast with those reported or expected among natural habitats, in which natural dispersal predominates. For instance, gradual spread and isolation by distance are not particularly expected among populations in marinas (Azmi, et al. 2014). These expectations were supported by results of population genetics studies that examined connectivity patterns of native and non-native species inhabiting marinas. For example, Hudson, et al. (2016) showed a low genetic structure between close or distant populations of the native tunicate *Ciona intestinalis* within the Western English Channel. Similarly, in the same region, Guzinski, et al. (2018) revealed a chaotic regional genetic structure of marinas' populations of the introduced seaweed *Undaria pinnatifida*. In these two cases, the observed patterns were explained by (almost) unpredictable human-mediated dispersal through leisure boating. It is, therefore, particularly important to evaluate the potential routes of secondary dispersal from a given marina, and better assess the presence of NIS in neighbouring natural habitats. This will allow a better understanding of the factors that may facilitate their spread and establishment away from marinas, such as particular dispersal or life-history traits, to ultimately anticipate any further spread of NIS in natural habitats.

### 3. Molecular tools for the biomonitoring of non-indigenous species

Identifying and counting non-indigenous species (NIS) are pre-requisites for addressing scientific questions and management issues. Various methods and tools have been proposed for marine biodiversity monitoring and surveys, especially in marinas, from full inventories, deployment of settlement panels examined in the laboratory, to Rapid Assessment Surveys (Lehtiniemi, et al. 2015). They all mostly rely on species identification based on morphological criteria. For monitoring purposes, time-limited *in situ* surveys (Rapid Assessment Surveys; e.g. Cohen, et al. 2005; Bishop, Wood, Lévêque, et al. 2015) are particularly used because they allow fast reports over a large area and in a relatively short time. However, many marine NIS may remain undetected because they are hidden, particularly at the early stages of

the introduction, during which they occur at low density. In addition, identification errors might occur during field surveys, particularly when morphological traits are difficult to identify with fast checking, as recently highlighted in several introduced colonial tunicates (e.g. *Didemnum vexillum*, Turon, et al. 2020; *Botrylloides diegensis*, Viard, et al. 2019). Even with laboratory work, identifying NIS based on morphological criteria is a difficult task. Many NIS indeed belong to taxonomic groups for which identification at the species level is challenging, and especially now that taxonomist expertise is becoming rare. Moreover, certain species or particular life stages (e.g. larval stages, early recruits) may not display morphological criteria allowing their identification. These difficulties have been the causes of many misidentifications that potentially led to new introductions being not detected. For example, *Watersipora subtorquata* (d'Orbigny, 1852) was misidentified by several authors in different parts of the world that were, in fact, invaded by the congeneric species *Watersipora subatra* (Ortmann, 1890) (Vieira, et al. 2014). In other cases, the same species was described several times in different regions and their names were, much later, synonymised. This is the case of the polychaete worm *Marphysa victori*, recently described from the Arcachon Bay (France) which was later found to be genetically identical to the Japanese species *Marphysa bulla* (Lavesque, et al. 2020). Finally, some species were considered as cosmopolitan but were, in fact, a complex of cryptic species with more localised distribution ranges (Darling and Carlton 2018). This is the case of the colonial ascidian *Diplosoma listerianum*, in which at least four cryptic lineages were identified; only one of them showing a wide distribution (Pérez-Portela, et al. 2013). The extent of these identification issues, together with the need to resolve some questions that could not be answered with morphological approaches, such as determining the source population of a particular NIS, called for the use of alternative methods using molecular techniques.

### **a. A furnished DNA-based toolbox with many applications**

The development of molecular approaches in the last decades and their use becoming more accessible for smaller-size laboratories has generated great advances in the field of biological introductions (Comtet, et al. 2015; Viard and Comtet 2015; Darling, et al. 2017; Zaiko, et al. 2018). From a research perspective, DNA-based approaches allow to investigate introduction patterns and processes such as determining introduction routes, estimating the extent of propagule pressure, examining dispersal patterns, or looking for adaptive changes in NIS new range

(Geller, et al. 2010; Lawson Handley, et al. 2011; Bock, et al. 2015; Viard and Comtet 2015; Viard, et al. 2016). These questions are central to the “invasion genetics” scientific field, which has increased in the last thirty years. DNA-based approaches have also provided tools to support specimens’ identification, thus overcoming some limitations of traditional approaches (in particular the lack of taxonomic expertise and the lack of morphological diagnostic features). Moreover, the advances in the field of high-throughput sequencing (HTS) is expected to free us from the time-consuming task of treating individuals separately, as done with traditional molecular techniques (e.g. molecular barcoding, see below) and to allow to gain insights into whole communities with one single sample. Methods currently in use and their applications for studies of biological introductions are listed in Table 2.

**Table 2** Overview of popular molecular techniques with some of their applications in the study of biological invasions, and NIS surveys. Modified from Viard and Comtet (2015), Darling, et al. (2017), and Zaiko, et al. (2018).

Technique		Target	Applications
PCR-based	End-point PCR	A single or a few species	Abundance and distribution data of the targeted species Active surveillance (targeted surveillance) for detection of targeted NIS (e.g. high risk NIS)
	qPCR		
	ddPCR		
	Microsatellites		Population-orientated studies (e.g. sources and pathways of introduction, evolutionary changes)
Sanger sequencing	DNA sequencing data DNA barcoding	A single or a few species	NIS diversity; reconstruction of introduction pathways Surveillance: NIS identification (e.g. confirming field identification)
High-throughput sequencing	DNA metabarcoding	Large taxonomic range	Community analyses Passive surveillance (broad taxonomic range), including NIS and associated native species
	metagenomic/metatranscriptomic	Large taxonomic range	Functional and genetic community analyses (so far, mostly available for unicellular organisms)
	SNPs from Genotyping-by-Sequencing (e.g. RAD-sequencing)	A single or a few species	Population-orientated studies (e.g. sources and pathways of introduction, evolutionary changes)

All these methods are valuable additions to more traditional techniques but they have their own limitations which should be taken into account when applying them to NIS biomonitoring. PCR-based methods rely on the amplification *via* Polymerase Chain Reaction of a DNA marker targeted by a set of primers designed to identify one particular species. While end-point PCR only allows the identification of the target species, quantitative PCR (qPCR) and digital droplet PCR (ddPCR) also allow the calculation of the amount of DNA from the target species initially present in the sample. These methods can be applied to detect NIS in a sample composed of a mixture of organisms or in an environmental sample (see Box 1). They are more sensitive than HTS methods and can offer, in the case of the qPCR and ddPCR, quantitative results (Wood, et al. 2019). These techniques, however, can only be applied to a few number of species at a time and are mostly dedicated to the detection and quantification of targeted NIS (e.g. to study the spread of a given NIS; Ardura and Zaiko 2018). DNA barcoding based on Sanger sequencing (see next section) can be a valuable addition to traditional methods for NIS identification and establishment of species inventories. It is a useful tool to confirm first identifications made in the field or for discriminating cryptic species. Its use for a high number of individuals is, however, not recommended as it can be very time consuming and costly (deWaard, et al. 2009; Porco, et al. 2013). Similarly, genotyping methods such as the amplification of microsatellites or the detection of single nucleotide polymorphisms (SNPs) from Genotyping-By-Sequencing techniques (such as RAD-sequencing, Mastretta, et al. 2015) are powerful tools for addressing many questions about the introduction process, such as tracing back the introduction history, testing for admixture or founding events, but they require a high optimization effort, sampling effort, expertise and only target one or a few species. For some research questions and monitoring objectives, the development of HTS techniques offered alternatives, notably because individual pools can be treated altogether at a reduced labour and financial cost. Whatever the type of nucleic acid targeted (DNA or RNA) or the particular method employed, data can theoretically be collected for the whole community. The efficiency and accuracy of these approaches is, however, still under evaluation. Some biases inherent to the sequencing and bioinformatics processes have already been pointed out. These biases and limitations deserve further assessment, in particular regarding NIS biomonitoring and studies of introduction processes.

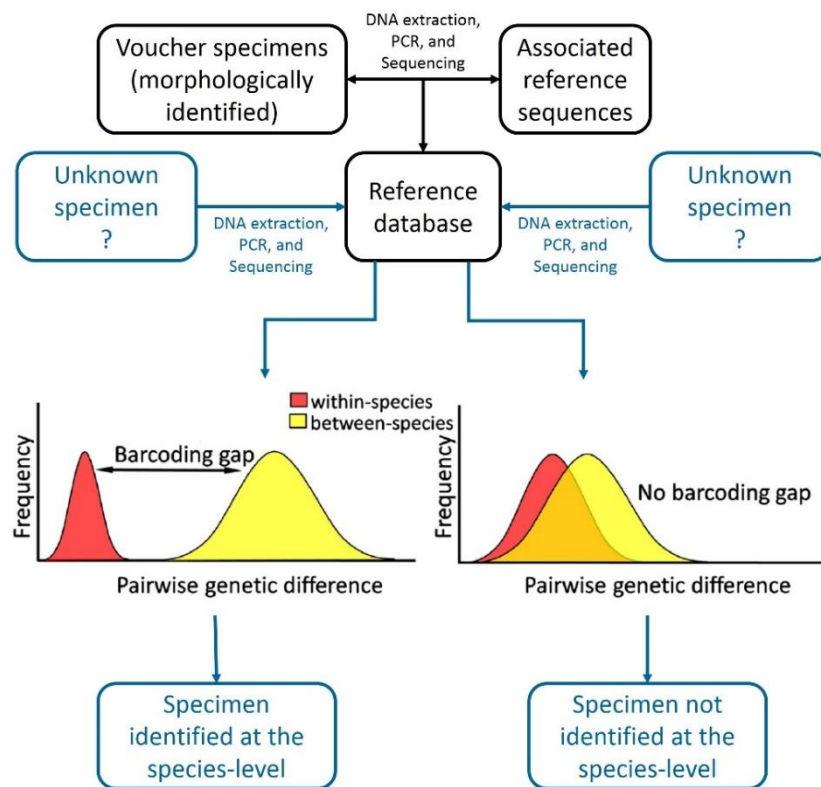
## b. NIS identification through DNA barcoding: general principle

Taxonomy is central to all fields of biology, including in ecology (Bortolus 2008), and issues related to species identification are particularly critical when dealing with biological introductions. Species misidentifications can cause the non-detection of a newly arrived non-indigenous species (NIS), which could lead to its uncontrolled proliferation in its new environment. This was the case for the dense-flowered cordgrass *Spartina densiflora* which was introduced from Chile to Humboldt Bay (California) and was misidentified as the native *Spartina foliosa*. It was later intentionally transplanted to Creekside Park in San Francisco, as part of a restoration program, still under the false assumption that it was the native species. The NIS was correctly recognised 30 years later but, by then, it had already spread along the US West Coast (Bortolus 2008). NIS misidentifications can also be responsible for the use of management strategies not adapted to the organism's biology. The NIS *Crepidula convexa*, for example, was misidentified as *Crepidula fornicata* for 20 years in Humboldt Bay (California). The former is a direct developer thus having a much lower dispersal potential than *C. fornicata*, a benthic-pelagic species (McGlashan, et al. 2008). These identification issues can often be overcome with the use of molecular tools, and especially DNA barcoding (Bucklin, et al. 2011).

DNA barcoding relies on the use of a DNA sequence for linking a specimen to a species name (Hebert et al. 2003). Some parts of the genome are, indeed, conserved enough throughout the tree of life to be present in all organisms but variable enough to be different between taxa, and sometimes species. The use of portions of these particular genes as markers allows to carry out taxonomic identification (ideally at the species level) on the sole amplification and sequencing of a small DNA fragment, and the comparison of the sequence obtained with those already present in databases compiling sequences to which taxa/species names are associated. The choice of the marker to be used depends on the taxa of interest and the possibility to assign a species name to a specimen depends, among other parameters, on the presence of a "barcoding gap" (i.e. a greater genetic variation between species than within species; Fig. 7) for the taxa considered.

A wide range of markers has been adopted for barcoding, such as the large chain of the RuBisCo gene (*rbcl*) for plants (Newmaster, et al. 2006), the nuclear ribosomal internal transcribed spacer (ITS) region for fungi (Schoch, et al. 2012), or the mitochondrial gene coding for the first subunit of the cytochrome oxidase (COI) for metazoans (Andújar, et al. 2018; Hebert, et al. 2003). Other markers are also used

such as genes coding for ribosomal subunits (e.g. nuclear 18S, mitochondrial 16S or 12S). Even if the efficiency of these markers has been validated for a wide range of taxa, their capacity of discriminating species can vary greatly depending on the taxonomic groups considered, and scientists have often designed specific primers targeting an optimal portion of the gene for the small number of species included in their study. Several primers, characterized as “universal”, have been designed to amplify a sensible barcoding region for a large group of organisms (e.g. for metazoans invertebrates; Folmer, et al. 1994) but they usually suffer from differential amplification success for some taxa.



**Figure 7** A key issue for a reliable barcoding approach is the presence of a barcoding gap. Sequence databases are made of sequences produced with a given marker for one or several specimens that have been identified (usually from morphological criteria). The possibility to use this sequence as a barcode depends on the fact that the molecular variability is lower within species (i.e. between individuals of this species) than among species (i.e. between individuals belonging to different close species, such as within a genus). Figure modified from Viard and Comtet (2015).



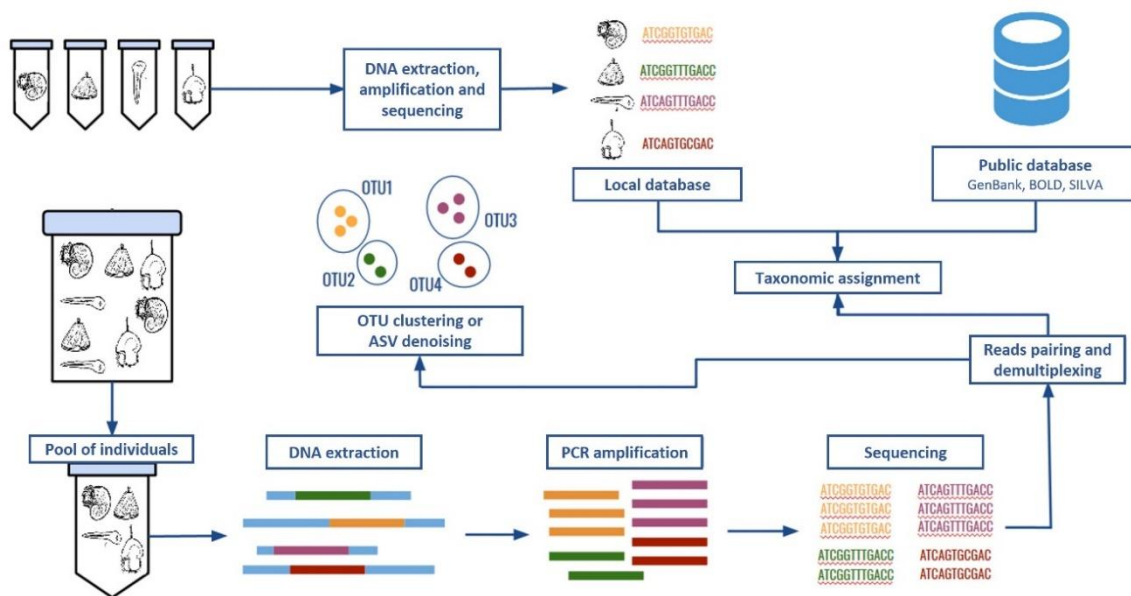
DNA barcoding does not intend to delineate or classify species based on molecular data. Its sole purpose is to name specimens, based on already described species and molecular criteria alone. In that sense, DNA barcoding is a valuable complement for species, and more specifically in our case NIS, identification (Bucklin, et al. 2011). In order to be able to give a name to an unknown specimen, its DNA sequence need to be compared to a database composed of reliable references, ideally produced from voucher specimens that have been identified by taxonomic experts. Most public databases, however, are filled with sequences which have been produced without expert identification made for barcoding purposes, and thus errors may occur challenging the reliability of the barcoding approach (e.g. Harris 2003). This is particularly true for NIS detection for which errors might lead to the misidentification of an introduced species as a native one, as shown recently for *Botrylloides* (Viard, et al. 2019). This issue needs to be taken into account (for instance by being cautious about the species list obtained, or by using dedicated or custom databases) but does not question the usefulness of such an approach in biological introduction studies.

### c. Producing and processing metabarcoding data

DNA metabarcoding is the taxonomic identification, *via* a combination of High-Throughput Sequencing (HTS) and barcoding, of multiple specimens, at once, without prior sorting (Fig. 8). The development of HTS thus overcomes some of the main issues of barcoding such as the time and cost required for naming a high number of individuals (Borrell, et al. 2017). Once eDNA or bulk DNA has been obtained (Box 1), it is amplified with PCR, then amplicons are sequenced, and reads obtained from HTS are processed using bioinformatics. Finally, sequences are compared to a reference database to be assigned to a taxon and to be used for diversity analyses (Fig. 8). Alternatively, the produced sequences can also be used as is (i.e. without assignment) for diversity analyses.

The use of DNA metabarcoding is appealing because of its ability to identify a high number of species/taxa without prior taxonomic knowledge. It is, however, still subject to a certain number of issues (e.g. Corse, et al. 2017; Alberdi, et al. 2018; Lamb, et al. 2019) which still need investigation to determine if they can or cannot be resolved by the advances in laboratory and informatics technologies.

DNA metabarcoding is used for a diversity of organisms, and multiple types of samples. In any case, all the DNA molecules present in a sample are jointly extracted. When environmental samples are used, such as water or sediments, the so-called 'environmental DNA' (eDNA, Box 1) is analysed. We will refer to bulkDNA when DNA is obtained from a homogenized pool of organisms (e.g. plankton samples or organisms scraped on settlement plates). Finally, DNA can also be extracted from solutions in which specimens were preserved (e.g. ethanol; Hajibabaei, et al. 2012).



**Figure 8** Chart of the different steps of the metabarcoding protocol, from sample collection to data analyses.

Collecting DNA from various types of samples is not straightforward and DNA extraction protocols must be adapted to the type of sample used (water, soil, homogenized organisms). For example, DNA extraction from soil samples must account for the high amount of humic substances that might act as PCR inhibitors. Moreover, organisms from an unsorted pool might not all have the same body composition which might lead to a differential success in DNA recovery (Eichmiller, et al. 2016; Deiner, et al. 2018). Several commercial kits have been recently developed for eDNA or bulk DNA extraction demonstrating the growing enthusiasm for molecular meta-omics analyses.

**Box 1: The use of environmental DNA (eDNA) and bulk DNA for studying and monitoring non-indigenous species (NIS)**

Environmental DNA (eDNA) is described as the complex mixture of genomic DNA from many different organisms found in an environmental sample, such as water, soil or even faecal samples (Taberlet, et al. 2012). It includes extracellular DNA resulting from cell death and subsequent release in the environment, or intracellular DNA present in living cells or multicellular microorganisms. In the case of marine organisms, and especially for benthic invertebrates, sampling techniques can be costly and are usually very damaging for the ecosystem. Moreover, rare species can be hard to detect and it can be challenging to get a comprehensive view of the whole community. In that regard, the use of eDNA allows the simplification of sampling procedures and is non-destructive of the organisms studied. The perspective of collecting DNA from every species leaving in a particular habitat by sampling water (Fig. 9) is very appealing and many studies have evaluated its potential for studying biological introductions (e.g. Ardura, et al. 2015; Borrell, et al. 2017). In this thesis we will not describe DNA extracted from a pool of organisms (e.g. zooplankton samples) as eDNA but we will prefer it the term of bulk DNA. This type of samples might be particularly interesting to use for NIS surveys, in particular when analysing planktonic larval stages of marine invertebrates because larvae play a major role at different stages of the introduction process. For instance, they can be transported in ballast water (Carlton and Geller 1993), and can thus be responsible for a species primary introduction. They are also the main vector for natural dispersal, thus playing a major role in expansion in the novel introduction area. Another category of bulk DNA, interesting to examine, is DNA obtained from homogenization of the organisms that have settled on experimental panels thus targeting the organisms able to settle on hard substrates.

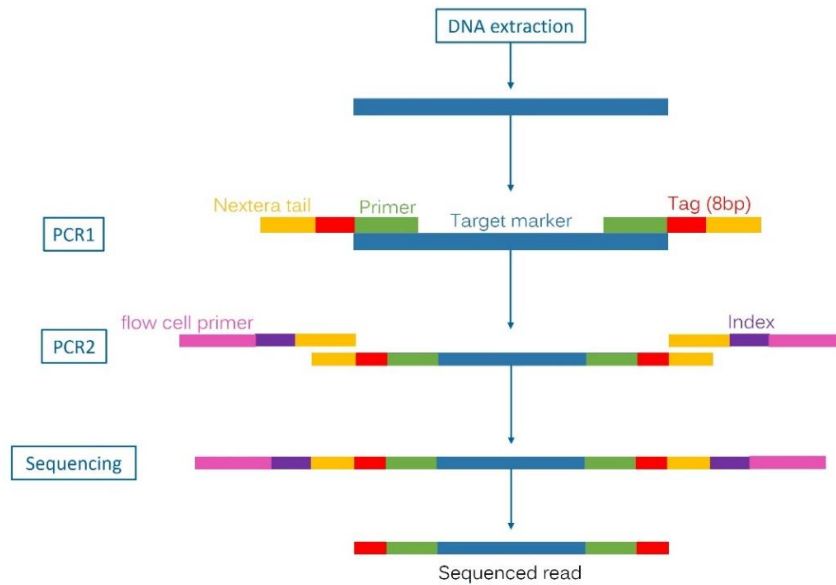
Several molecular techniques can subsequently be applied to eDNA and bulk DNA depending on the objectives. One of the major purposes of eDNA is the detection of rare or elusive species in aquatic environments. In the case of NIS, early detection, when they are still in low abundance in the environment, is a crucial point for future management and control, and thus a primary objective of eDNA-based studies. When targeting one or a few taxa, such as NIS listed on a "watch list" because of potential ecological and socio-economic risks if introduced, the use of PCR-based (end-point PCR, qPCR, ddPCR) methods associated with species-specific primers have been favored. They are more sensitive than HTS approaches and can give a more reliable quantitative assessment of the species distribution (Harper, et al. 2018; Wood, et al. 2019). The objective to report novel NIS over a given area, however, requires identifying species that were not specifically expected over a wide range of taxa, and thus needs to use a broader technique such as DNA metabarcoding (see main text). With this approach, data can be collected for many taxonomic groups at once, allowing both native and NIS detection, and providing an assessment of the overall community. Some issues inherent to eDNA could, however, introduce a number of biases into the analyses and should be taken into account when analyzing results from this type of samples. These include DNA dispersal (exclusively for aquatic samples), DNA degradation over time (10h to several days in marine water), the differential release of DNA according to the organisms' physiology, developmental stage or sex, or the presence of gametes in the water that might bias species relative abundances (Taberlet, et al. 2018).



**Figure 9** Picture of marine water sampling in a marina for later extraction of environmental DNA. Photo credit: Yann Fontana – Station Biologique de Roscoff.

Extracted DNA is then amplified by PCR to target a specific marker (Fig. 10). Primers should be carefully chosen to avoid amplification bias (i.e. preferential amplification of certain species over others in a mixture of DNAs from multiple taxa). A multi marker approach is more and more favoured in order to overcome this issue (e.g. Stefanni, et al. 2018; Cordier, et al. 2019). Moreover, rare species, representing a low proportion of the DNA in a mixture, might be stochastically not amplified during the PCR step. It is thus recommended to multiply PCR replicates in order to improve detection capacity (Ficetola, et al. 2015; Alberdi, et al. 2018).

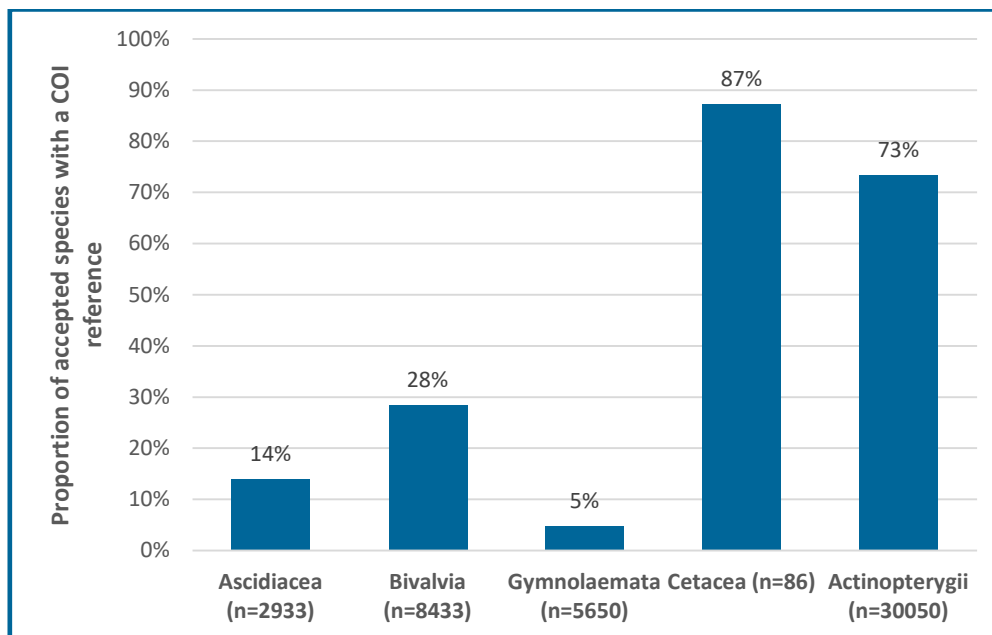
Finally, amplicons produced are sequenced using a HTS technique. Several sequencing platforms are available with different strengths and weaknesses. The most used, to date, for metabarcoding studies is the Illumina® platform which is based on sequencing by synthesis methodology. One of the major drawbacks of this approach is the length of the fragment that can be sequenced (so far, max 300 bp). Even if the technology offers the possibility to sequence both ends of the fragment, this limitation has compelled scientists to use markers of less than 500 bp which reduces their discriminating power for species assignment. Moreover, all HTS techniques have relatively high error rates and all erroneous sequences produced need to be sorted out of the dataset before analysing the results.



**Figure 10** Detailed steps of a dual-barcoded dual-indexed two-PCR library preparation for HTS sequencing used for most analyses done in this thesis. From the DNA sample, a first PCR is done with primers specific to the targeted gene region associated to one tag combination (specific to the PCR replicate) and one tail used in the following PCR. The second PCR is made to elongate the fragment with an index combination (specific to the sample) and primers to be used for the sequencing step.

Data collected after sequencing need to be processed using a bioinformatics pipeline in order to discriminate between “true” sequences and errors. Several bioinformatics tools are available implementing different algorithms which can be separated into two categories. The denoising tools try to identify “true” sequences called Amplicon Sequence Variants (ASVs) based on their abundance, their divergence, and their sequencing quality scores, in order to remove potential errors (e.g. DADA2; Callahan, et al. 2016, the obiclean command of the OBITOOLS pipeline; Boyer, et al. 2016). The second group, the clustering tools, produce Operational Taxonomic Units (OTUs) grouping together the “true” sequences and the errors deriving from it. They can either cluster reads depending on a fixed threshold for sequence similarity (e.g. the clustering tool implemented in VSEARCH; Rognes, et al. 2016) or be based on a network algorithm with a dynamic threshold (SWARM; Mahé, et al. 2015). All these tools have their advantages and limitations and their use is contingent on the type of analyses that will be further performed. For example, denoising tools produce ASVs that include the genetic diversity within each species whereas clustering tools produce OTUs supposed to illustrate the species diversity. Additional to clustering or denoising, errors produced during the amplification steps can be removed by tagging PCR replicates in order to identify them, allowing to

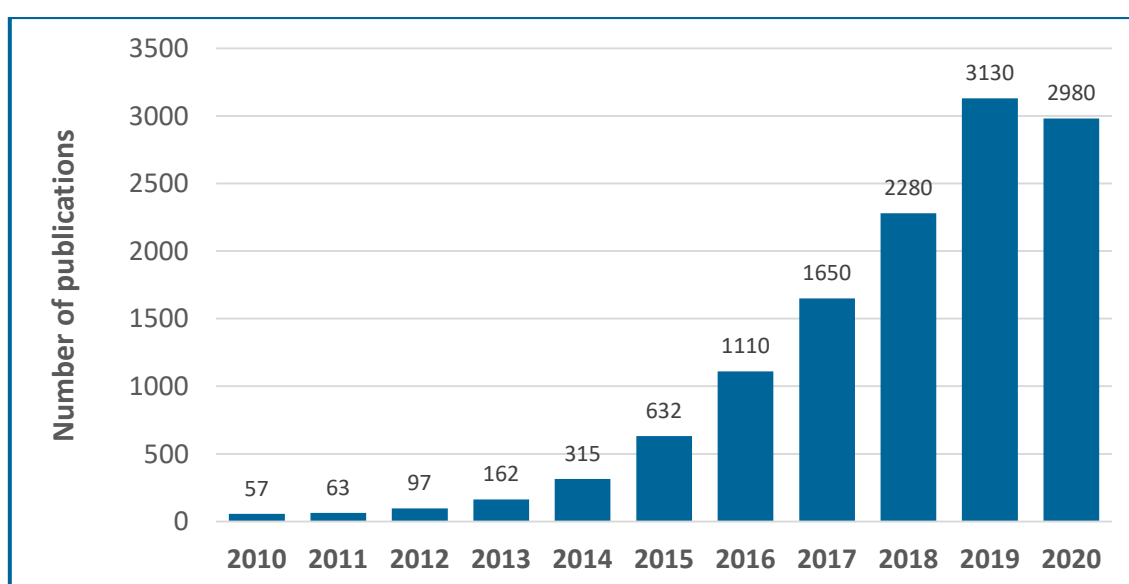
discard sequences present in one replicate only. When sequencing several samples at a time, amplicons produced during the amplification step need to be identified according to the sample they belong to. This is done by adding specific short tag sequences after the primers designed to target the marker of interest or by adding index sequences which will be recognized by the sequencing device (Fig. 10). Reads will further be attributed to a particular sample by demultiplexing and this step can be the source of errors called index-jump. In this case sequences can be wrongly assigned to a sample and lead to the false detection of a species in a location where it is actually not present. It is thus essential to account for this issue when preparing sequencing libraries and index-jump controls (unused index combinations) must be included to evaluate this phenomenon (Taberlet, et al. 2018).



**Figure 11** Illustration of the variation in percentage of accepted species (as listed in WORMS - World Register of Marine Species database; <http://marinespecies.org/>) with reference in public databases (here in BOLD - Barcoding of Life Database; <https://boldsystems.org/>) across different taxonomic groups. The number of accepted species for each class is given in parentheses. Data obtained on September 1, 2020.

When a set of ASVs or OTUs is obtained, diversity analyses can be performed directly on these data to get insights on the alpha-diversity (e.g. species richness) or dissimilarity patterns between localities for example (e.g. Cahill, et al. 2018; Bakker, et al. 2019). When the identification of species is important, however, like in the case of

NIS early detection, a taxonomic assignment must be performed. This step consists in the comparison of ASVs or representative sequences of OTUs against a reference database. In an ideal setting, every haplotype of every species potentially present in the samples analysed would have a corresponding reference in the database. This is, however, usually not the case and many species do not possess a reference, for every marker used in metabarcoding, in public databases. This lack of references, however, can vary greatly according to taxonomic groups (Fig. 11). Moreover, some species cannot be discriminated with some markers, because of poor taxonomic resolution, and share the same reference sequence. In this context, choosing the right parameters to get an accurate assignment can be a difficult task and compromises need to be made. Many tools are available to automatize this process, either based on global alignment strategies (e.g. *ecotag* from the *OBITOOLS*) or using local alignments (e.g. *RDP CLASSIFIER* in *DADA2*). Some of them use phylogenetic placement algorithms (e.g. *EPA-NG*; Barbera, et al. 2018), whereas others use only identity thresholds as a cut-off (e.g. *BLAST*-based methods; Altschul, et al. 1990). In every case, false positives can arise because of a lack of references or because of remaining PCR or sequencing errors. In the case of NIS detection, the erroneous detection of introduced species can be very problematic and any detection *via* a metabarcoding method should be carefully checked with other molecular or morphology-based approaches (Darling, et al. 2020; Sepulveda, et al. 2020).



**Figure 12** Number of publications using the term “metabarcoding” in Google Scholar by year of publication for the last decade. The number for 2020 only accounts papers published before September 2020.

The number of DNA metabarcoding studies has grown exponentially in the last decade (Fig. 12). Firstly used by microbiologists (Sogin, et al. 2006), metabarcoding is now extensively used in all kinds of habitats and for all types of organisms (e.g. Andersen, et al. 2012; Bakker, Klymus, et al. 2017; Marquina, et al. 2019). The diversity of applications has prompted the development of many tools, all producing different results (Pauvert, et al. 2019). Some tools are proposed wrapped in guided pipelines such as MOTHUR (Schloss, et al. 2009) or QIIME (Caporaso, et al. 2010) but others are developed as stand-alone and their combination can vary from one study to another. If we also take into account the various parameters of each tool which can be adapted to the desired application, the number of combinations that can be applied to a particular sample is colossal. There are almost as many metabarcoding pipelines as scientists using these tools. This multiplicity of methodologies is so extravagant that some authors have called their approach “Just Another Metabarcoding Pipeline” (JAMP; Elbrecht, et al. 2018). This situation renders the comparison between results very difficult, which calls for standardized methods, both for laboratory work and bioinformatics analyses.

In the case of marine biological introductions, DNA metabarcoding have been used in almost 50 original experiments in the last seven years and was applied to various types of samples using different markers (Table 3). Most of them used both 18S and COI as markers expected to be particularly appropriated for studying metazoan taxa, and they were mainly directed towards ballast water, fouling organisms or plankton samples. Despite this high number of studies, this technique is still in its infancy and evaluations are still needed to grasp all its potential and weaknesses.



**Table 3** List of papers (sorted by publication year) reporting empirical data based on a metabarcoding approach to study non-indigenous species in the marine environment. The study region is provided with the sample type and marker(s) used. Data extracted and completed from Duarte, et al. (2020).

Reference	Geographic region	Sample type	Marker
Pochon et al 2013	New Zealand	eDNA (sediment) + bulkDNA (plankton)	18S
Ardura et al 2015	International ships	eDNA (ballast water)	COI
Pochon et al 2015	New Zealand	bulkDNA (fouling organisms)	18S
Zaiko et al 2015a	Baltic Sea	bulkDNA (plankton)	COI
Zaiko et al 2015b	International ships	bulkDNA (plankton samples from ballast)	COI + rbcL
Zaiko et al 2015c	International ships	bulkDNA (plankton samples from ballast)	COI
Abad et al 2016	Bay of Biscay (Spain)	bulkDNA (plankton)	18S
Brown et al 2016	Canada	bulkDNA (plankton)	18S
Chain et al 2016	Canada	bulkDNA (plankton)	18S
Ghabooli et al 2016	International ships	bulkDNA (plankton samples from ballast)	18S
Pagenkopp Lohan et al 2016	International ships	eDNA (ballast water)	18S
Zaiko et al 2016	New Zealand	bulkDNA (fouling organisms)	18S
Ardura et al 2017	Baltic sea	bulkDNA (plankton)	COI
Borrell et al 2017	Bay of Biscay (Spain)	eDNA (water samples)	18S + COI
Fletcher et al 2017	New Zealand	bulkDNA (plankton samples from bilge water)	18S
Pagenkopp Lohan et al 2017	International ships	eDNA (ballast water)	18S
Pochon et al 2017	New Zealand	eDNA and eRNA (bilge water)	18S
Borrel et al 2018	Bay of Biscay (Spain)	eDNA (water samples)	COI
Darling et al 2018	International ships	bulkDNA (plankton)	18S
Deiner et al 2018	Southampton (UK)	eDNA (water samples)	18S
Grey et al 2018	Canada, USA, Australia, Singapore	eDNA (water samples)	18S + COI
Gunther et al 2018	Germany	eDNA (water samples)	18S + COI
Koziol et al 2018	Australia and Kazakhstan	eDNA (water and sediment samples) + bulkDNA (settlement plates and plankton)	18S + COI + 16S
Lacoursière-Roussel et al 2018	Canada	eDNA (water samples)	COI
Stefanni et al 2018	Adriatic sea (Italy)	bulkDNA (plankton)	18S + COI
von Ammon et al 2018a	Auckland (New Zealand)	bulkDNA (fouling organisms)	18S + COI
von Ammon et al 2018b	Auckland (New Zealand)	bulkDNA (fouling organisms)	18S
Wangensteen 2018	Spain	bulkDNA (fouling organisms)	18S + COI
Couton et al 2019	Brittany (France)	bulkDNA (plankton)	18S + COI
Holman et al 2019	United Kingdom	eDNA (water and sediment samples)	18S + COI
Leduc et al 2019	Canada	eDNA (water samples) + bulkDNA (plankton and benthic samples)	18S + COI

Pagenkopp Lohan et al 2019	California (USA)	bulkDNA (plankton)	COI
Petri et al 2019	International ships	eDNA (ballast water)	18S
Rey et al 2019	International ships	eDNA (ballast water)	18S + COI
Shang et al 2019	International ships	eDNA (sediment)	28S
Shaw et al 2019	International ships	eDNA (sediment)	18S
von Ammon et al 2019	Auckland (New Zealand)	eDNA (water samples) + bulkDNA (fouling organisms)	COI
Wood et al 2019	Auckland (New Zealand)	eDNA (water samples) + bulkDNA (fouling organisms)	18S + COI
Wright et al 2019	International ships	eDNA (ballast water)	18S
Ardura 2020	Bay of Biscay (Spain)	eDNA (ballast water + port water)	COI + rbcL
Azevedo 2020	Portugal	bulkDNA (fouling organisms)	16S + 18S + 23S + COI
Darling 2020a	International ships	bulkDNA (plankton)	18S
Huhn et al 2020	Indonesia	eDNA (water samples)	16S + ITS2 + 18S
Ibabe 2020	Asturias (Spain)	bulkDNA (fouling organisms)	COI
Lin et al 2020	International ships	bulkDNA (plankton samples from ballast)	18S
Rey et al 2020	Bilbao (Spain)	eDNA (water and sediment samples) + bulkDNA (settlement plates and plankton)	18S + COI
Suarez-Menendez et al 2020	France	eDNA (water samples)	COI
Westfall et al 2020	Departure Bay (Canada)	eDNA (water samples) + bulkDNA (plankton)	COI + 16S + 18S + ND4

## Thesis objectives

The deliberate or accidental translocation of organisms outside their native range by human activities is occurring across oceans and ecosystems, over wide spatial ranges. Biological invasions are a recognized human-driven stressor interacting with other drivers of global change on biodiversity. For studying non-indigenous species (NIS), and biological introduction processes, a pre-requisite is our ability to detect and accurately identify them. This seemingly simple task is however challenging, particularly when using traditional methods based on single specimen observation and morphological criteria. DNA-based tools, in particular barcoding, have been proposed to help achieving this goal but still requires individual handling of specimens. In that context, High-Throughput Sequencing techniques and metabarcoding have been proposed for a better efficiency. To date, most of the metabarcoding studies applied to marine organisms were dedicated to report

presence. They were applied to various types of samples in the hope of identifying newly introduced species, still in low abundance in the environment, or evaluating the distribution of previously reported NIS.

Previous metabarcoding studies showed interesting and very promising results, and these approaches might become used routinely in marine NIS surveillance and research. However, a certain number of concerns and biases have been raised, deserving further investigation (e.g. the sampling strategy, the rate of false positives *vs.* false negatives, the effects of the bioinformatics pipelines). Most of the studies, to date, are focused on NIS detection, but fewer experiments have targeted other aspects of the biological introduction process. Yet, metabarcoding might be of great interest to investigate other issues and research questions, such as the extent of NIS presence in wild habitats, NIS reproductive and dispersal abilities, NIS contribution to species assemblages, etc. In this thesis work I attempted to address some of these diverse issues.

As in previous studies, my first goal was to assess to which extent DNA metabarcoding correctly detects NIS. Nevertheless, I also aimed at gaining further insights into different aspects of marine introductions. Three specific objectives structured my thesis work: i) distinguishing cryptic species and lineages, ii) detecting NIS in high risks areas (here marinas, see below) with a broad taxonomic coverage; and iii) analyzing the potential escape of NIS into natural environments. Each of these objectives is addressing one issue in eco-evolutionary sciences, one challenge for NIS management, and one methodological challenge (Table 4). In each of the experiments that I carried out, the HTS/metabarcoding approach was conducted simultaneously with a traditional approach (based on *in situ* sampling, single specimen and morphological-based identification or barcoding-based species identification).

**Table 4** Research questions and challenges addressed in this thesis.

Research question	Application to surveillance and management	HTS-based approach		Methodological challenges
		Type of sample	Traditional method	
Co-occurrence and population diversity of cryptic and congeneric NIS	Ability of metabarcoding to detect cryptic NIS and document their distribution, and assess population diversity	Targeted (one genus) Preservative medium (ethanol) Sanger sequencing of individual organisms		Primer design Use of preservative medium (for non-destructive surveys) Choice of the bioinformatics pipelines to jointly analyze taxonomic and genetic diversity
NIS contribution to marina communities Extent of biotic homogenization in human-made habitats	Ability of metabarcoding to detect NIS and associated putative errors/uncertainties (ex. false negatives and false positives) Complementarity with traditional methods	Non-targeted/broad taxonomic spectrum (metazoans) Water samples Quadrats scrapped by diving followed by <i>in situ</i> species identification and count		Avoid technical biases (e.g. sampling protocol; control design, bioinformatics pipeline parameters) Completeness of reference databases
NIS' potential spread outside marinas: what is the most limiting life stage for colonizing new habitats (dispersal or recruitment success)?	Ability to detect NIS in natural habitats, including Marine Protected Areas (MPAs), and to assess the risk of spread in these habitats	Non-targeted/broad taxonomic spectrum (metazoans) Water samples / Settlement panel scraping/ Plankton samples Morphological identification on panels / Morphological identification of larvae (one species) / Sanger sequencing of individual larvae (one species)		Scrapping and homogenization procedure on large organisms Sampling accuracy Completeness of reference databases

I chose to address these questions targeting marinas, as study sites. A severe increase in introduction rates has been documented since the early 20th century, making invasive species a global issue. Such a trend is explained by the increased numbers of introduction vectors at a global scale (aquaculture, maritime traffic, etc.), and by the development of human-made infrastructures (e.g. dykes, ports, etc.) along the coasts (i.e. marine urbanization contributing to the ocean sprawl). Ports and marinas have been shown to be introduction hotspots that could favour NIS spread in surrounding environments. They should thus be prioritized in surveys and surveillance programmes. In addition, these artificial habitats are not surrogates of

neighbouring natural habitats, and display particular diversity and functioning, that deserve further investigation, such as the extent of the similarity between those habitats as compared to wild habitats. We worked at a regional scale, the Brittany region, which displays numerous marinas and is located in-between two biogeographic regions (i.e. the Lusitanian and Boreal provinces) very distinct from an environmental and biogeographic history point of view. As study species, I chose to work with benthic sessile species composing biofouling communities in marinas, as these assemblages are known to be composed of many NIS that are, moreover, easily dispersed with human activities such as leisure boating.

Regarding the three objectives and research questions listed above (Table 4), the structure of the thesis is as follows:

The first chapter focuses on a “targeted approach” (i.e. active surveillance) and evaluates the ability of High-Throughput-Sequencing (HTS) to identify both cryptic species and intraspecific variability. Haplotypic data could, for example, help figuring out the importance of founder events as well as the pathways of expansion. To that end, colonies of colonial ascidian species from the genus *Botrylloides* were collected in ten French marinas and were assembled into mock communities in ethanol. DNA extracted from the preservative ethanol was then amplified using specific COI primers designed to target this genus. Various metabarcoding pipelines were tested to evaluate their ability to detect species and haplotypes in each community.

In the second chapter, we tested the potential of water eDNA metabarcoding for NIS detection and community description in ten French marinas. In a first part, the detection efficiency of eDNA metabarcoding was tested by comparing the results with a traditional method based on scraping of biofoulers in quadrats and *in situ* morphology-based identification of taxa. We aimed at testing the limitations of metabarcoding, such as the non-detection of species observed within quadrats. Then, alpha- and beta diversity analyses were performed on various functional groups of marine organisms living in marinas. The impact of geographical (biogeographic region) and temporal factors (sampling was repeated twice, at different seasons) on the community diversity and structure was assessed, with focus on the response of NIS as compared to native species.

Finally, in the third chapter, we used DNA metabarcoding on various types of samples to assess the reproductive ability of NIS and their occurrence outside a marina (thus their potential for spread). As in the previous chapters, metabarcoding

was compared with results obtained with a traditional morphology-based approach. As we were interested in the dispersal of benthic sessile NIS fixed on hard substrates, we used settlement plates to target organisms for which propagules were present and able to metamorphose and settle. For a broader community assessment, this sampling was coupled with water sampling at the same location. Then, in a second part, we focused on the planktonic larval stage, with a dedicated plankton sampling carried out over 22 months with the aim to detect larval NIS, and get insights about the NIS reproductive patterns over time.



# CHAPTER I:

High-Throughput Sequencing on preservative ethanol is effective at jointly examining infra-specific and taxonomic diversity, although bioinformatics pipelines do not perform equally.







## Preamble

DNA barcoding has long been used to identify specimens at the species level when morphological criteria are lacking or difficult to use (i.e. (pseudo-)cryptic species). It can also provide additional information regarding genetic polymorphism at the species level, especially when using the COI marker, known to be variable in many marine invertebrates. This method, however, relies on Sanger sequencing of a single specimen which is both labor-intensive and costly. Metabarcoding has been developed more recently, with the rise of high-throughput sequencing (HTS) techniques, and offers the possibility to analyze hundreds of samples or specimens at once. Nevertheless, its potential for analyzing within-species polymorphism still needs to be investigated.

In this chapter, as a first 'metabarcoding' attempt, I examined the potential of HTS to jointly examine taxonomic and genetic diversity. I chose to target the colonial tunicates of the genus *Botrylloides*, particularly relevant here. Species identification is indeed difficult or ineffective based on morphological criteria, and yet this genus comprises both native and introduced species, the latter being particularly conspicuous in marinas and ports. This work also allowed me to get familiar with numerous bioinformatics pipelines, and understand their limitations or advantages. Similarly to the other tasks carried out during this PhD thesis, the results were examined as compared to traditional methods, here a regular barcoding analysis of the same colonies that were examined with HTS.



Collection of colonies from several *Botrylloides* species, all difficult to identify morphologically.

# High-Throughput Sequencing on preservative ethanol is effective at jointly examining infra-specific and taxonomic diversity, although bioinformatics pipelines do not perform equally.

**Running headline:** HTS to assess genetic and taxonomic diversity

Marjorie Couton<sup>1\*</sup>, Aurélien Baud<sup>1</sup>, Claire Daguin-Thiébaud<sup>1</sup>, Erwan Corre<sup>2</sup>, Thierry Comtet<sup>1</sup>, Frédérique Viard<sup>1</sup>

<sup>1</sup> Sorbonne université, CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

<sup>2</sup> Sorbonne université, CNRS, FR 2424, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

\* Correspondence author: [mcouton@sb-roscoff.fr](mailto:mcouton@sb-roscoff.fr)

**Keywords:** biodiversity, bioinformatics, bulkDNA, ethanol-based DNA, haplotype diversity, high-throughput sequencing, metabarcoding, tunicate

**This chapter has been submitted for publication in *Methods in Ecology and Evolution* and is currently in revision.**

## Abstract

High-throughput sequencing of amplicons (HTSA) has been proposed as an effective approach to evaluate taxonomic and genetic diversity at the same time. However there are still uncertainties as to how the results produced by different bioinformatics treatments impact the conclusions drawn on biodiversity and population genetics indices.

We evaluated the ability of six bioinformatics pipelines to recover taxonomic and genetic diversity from HTSA data obtained from controlled assemblages. To that end, 20 assemblages were produced using 354 colonies of *Botrylloides* spp., sampled in the wild in ten marinas around Brittany (France). We used DNA extracted from preservative ethanol (ebDNA) after various time of storage (3, 6, and 12 months), and from a bulk of preserved specimens (bulkDNA). DNA was amplified with specific primers targeting this ascidian genus. Results obtained from HTS data were compared with Sanger Sequencing on individual colonies (i.e. individual barcoding).

Species identification and relative abundance determined with HTSA data from either ebDNA or bulkDNA were similar to those obtained with traditional individual barcoding. However, after 12 months of storage the correlation between HTSA and individual-based data was lower than after shorter durations. The six bioinformatics pipelines were able to depict accurately the genetic diversity using standard population genetics indices ( $H_S$  and  $F_{ST}$ ), despite producing false positives and missing rare haplotypes. However, they did not perform equally and DADA2 was the only pipeline able to retrieve all expected haplotypes.

This study shows that ebDNA is a non-destructive alternative for both species identification and haplotype recovery, providing storage do not last more than six months before DNA extraction. Choosing the bioinformatics pipeline is a matter of compromise, aiming to retrieve all true haplotypes while avoiding false positives. We here recommend to process HTSA data using DADA2, including a chimera-removal step. Even if the possibility to use multiplexed primer sets deserve further investigation to expand the taxonomic coverage in future similar studies, we showed that specific primers allowed to reliably analyze the target genus within a complex community.

## Introduction

Although most biodiversity assessments rely on taxonomic diversity, many other components (functional, phylogenetic, genetic...) can provide complementary, and sometimes contrasting, information (Lindegren, Holt, MacKenzie, & Rahbek, 2018). In this context, next-generation biomonitoring (Makiola et al., 2020) based on high-throughput sequencing (HTS) of mixed DNAs offers the possibility to analyse simultaneously two biodiversity components (i.e. taxonomic and genetic), while solving problems related to morphology-based identification. It also allows to decrease handling time and costs as compared to individual-based methods.

The HTS of amplicons has already been tested for studying both taxonomic and genetic diversity, either by analysing DNA metabarcoding data obtained with universal primers (Elbrecht, Vamos, Steinke, & Leese, 2018; Pedro et al., 2017; Stat et al., 2017), or by targeting one or a few species using specific primers (Marshall & Stepien, 2019; Parsons, Everett, Dahlheim, & Park, 2018; Sigsgaard et al., 2016; Stepien, Snyder, & Elz, 2019; Tsuji et al., 2020a; Tsuji et al., 2020b). In metazoans, the COI mitochondrial gene has been preferentially used for such studies (e.g. Pedro et al., 2017), because of its high taxonomic resolution and ability to reveal within-species polymorphism (Andújar, Arribas, Yu, Vogler, & Emerson, 2018; Bucklin, Steinke, & Blanco-Bercial, 2011). Moreover, a considerable amount of sequences are available in public databases for this marker (Porter & Hajibabaei, 2018). Overall, HTS studies revealed that the most abundant haplotypes (i.e. unique sequences) are easily recovered, some rare ones can be missed, and some spurious sequences can be misidentified as haplotypes. Previous reports showed that different bioinformatics pipelines may produce divergent results when analysing HTS datasets for taxonomic diversity studies (Pauvert et al., 2019) but, to our knowledge, the consequences of the choice of divergent bioinformatics approaches (e.g. clustering vs. denoising) on haplotype recovery, as well as the impact of the resulting false positives and negatives on commonly used population genetics indices, have not been investigated.

When biodiversity assessments using HTS rely on the collection of a representative sample from the target community, they usually involve the homogenisation of all organisms to extract DNA from bulk. Processing each sample can be time-consuming and increases the risk of cross-contamination. Furthermore, this technique implies the destruction of the samples, rendering any further analyses impossible. Shokralla, Singer, and Hajibabaei (2010) first showed that preservative

ethanol could be used to recover and sequence invertebrate DNA without impacting the integrity of the samples. DNA extracted from preservative ethanol (ethanol-based DNA; ebDNA) was successfully used for HTS-based community analyses in terrestrial (Linard, Arribas, Andújar, Crampton-Platt, & Vogler, 2016; Marquina, Esparza-Salas, Roslin, & Ronquist, 2019; Zenker, Specht, & Fonseca, 2020), and freshwater organisms (Erdozain et al., 2019; Hajibabaei, Spall, Shokralla, & van Konynenburg, 2012; Martins et al., 2019; Zizka, Leese, Peinert, & Geiger, 2019). DNA was extracted after various storage durations (ranging from 12 hours to 15 months) and temperatures (from -25°C to ambient). Although Martins et al. (2019) showed that the yield and quality of ebDNA recovered increased in the first five to ten days of storage, to our knowledge, no experiment has investigated if community studies could be applied after several months of storage for marine organisms.

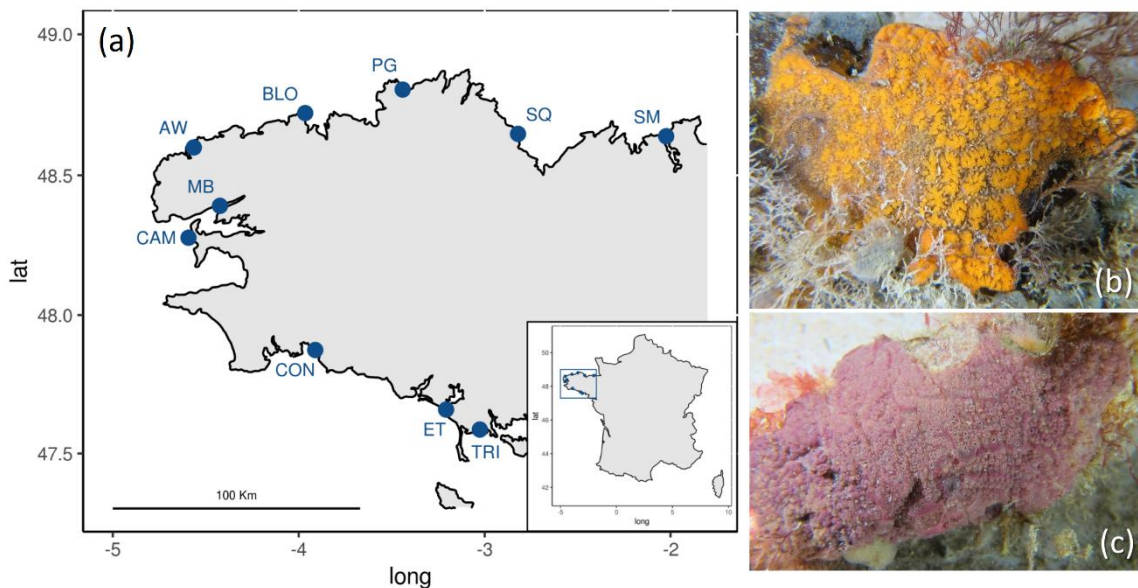
In this study we investigated the two knowledge gaps highlighted above. Our goal was to recommend an optimized methodology for jointly assessing taxonomic and genetic diversity *via* HTS on ebDNA. To this end, we evaluated the efficiency of six metabarcoding analysis pipelines, based on either a clustering or a denoising approach, to recover COI haplotypes and assess population genetic diversity indices. DNA was extracted from preservative ethanol of marine organisms stored at room temperature after up to twelve months. As a case study, we examined biofouling communities from marinas which are composed of many non-indigenous species, a major driver of biodiversity loss.

## Materials and methods

### a. Case study & sampling

We selected species of the genus *Botrylloides* as a case study. They are colonial ascidians composed of hundreds of individuals (zooids) embedded in a tunic (Figs. 1b and 1c). Among the 19 accepted species, two from our study area (English Channel), *Botrylloides violaceus* Oka, 1927, and *Botrylloides diegensis* Ritter & Forsyth, 1917, are recognized as globally invasive (Bock, Zhan, Lejeune, MacIsaac, & Cristescu, 2011; Viard, Roby, Turon, Bouchemousse, & Bishop, 2019), both originating from the North Pacific. They are a major component of biofouling communities and can have dramatic impacts on aquaculture facilities in their introduction range (Carman, Morris, Karney, & Grunden, 2010). The native *B. leachii* (Savigny, 1816) has

also been reported in our study area, in addition to a cryptic lineage, morphologically undistinguishable from *B. violaceus* (named BvX-H6 after Viard et al., 2019). *Botrylloides* species are notoriously difficult to identify based on morphology (Rocha et al., 2019; Viard et al., 2019). This issue can easily be solved by using the COI marker, effective in discriminating species from this genus (Rocha et al., 2019), and in detecting infra-specific diversity within the species present in the study region (Viard et al., 2019).



**Figure 1** (a) Collection sites of *Botrylloides* spp. colonies. SM = Saint-Malo, SQ = Saint-Quay-Portrieux, PG = Perros-Guirec, BLO = Bloiscon (Roscoff), AW = L'Aber Wrac'h, MB = Moulin Blanc (Brest), CAM = Camaret-sur-Mer, CON = Concarneau, ET = Étel, TRI = La Trinité-sur-Mer. (b) *Botrylloides diegensis*. (c) *Botrylloides violaceus*. Photo credit: Yann Fontana.

*Botrylloides* spp. colonies were sampled by scuba diving in 10 marinas around Brittany (English Channel and NE Atlantic, France; Fig. 1a). Between 32 and 36 colonies were randomly collected in each location along a 100-m transect below pontoons. A small piece of each colony was isolated in 100% ethanol for individual haplotype identification. The remaining parts of the colonies were stored together in 2-L plastic jars filled with 100% ethanol for further HTS-based analyses. We maximized the ethanol/tissue ratio by dividing the colonies into two jars (A and B) per marina. The samples were stored at room temperature.

## b. Sanger sequencing on individual zooid (SSIZ)

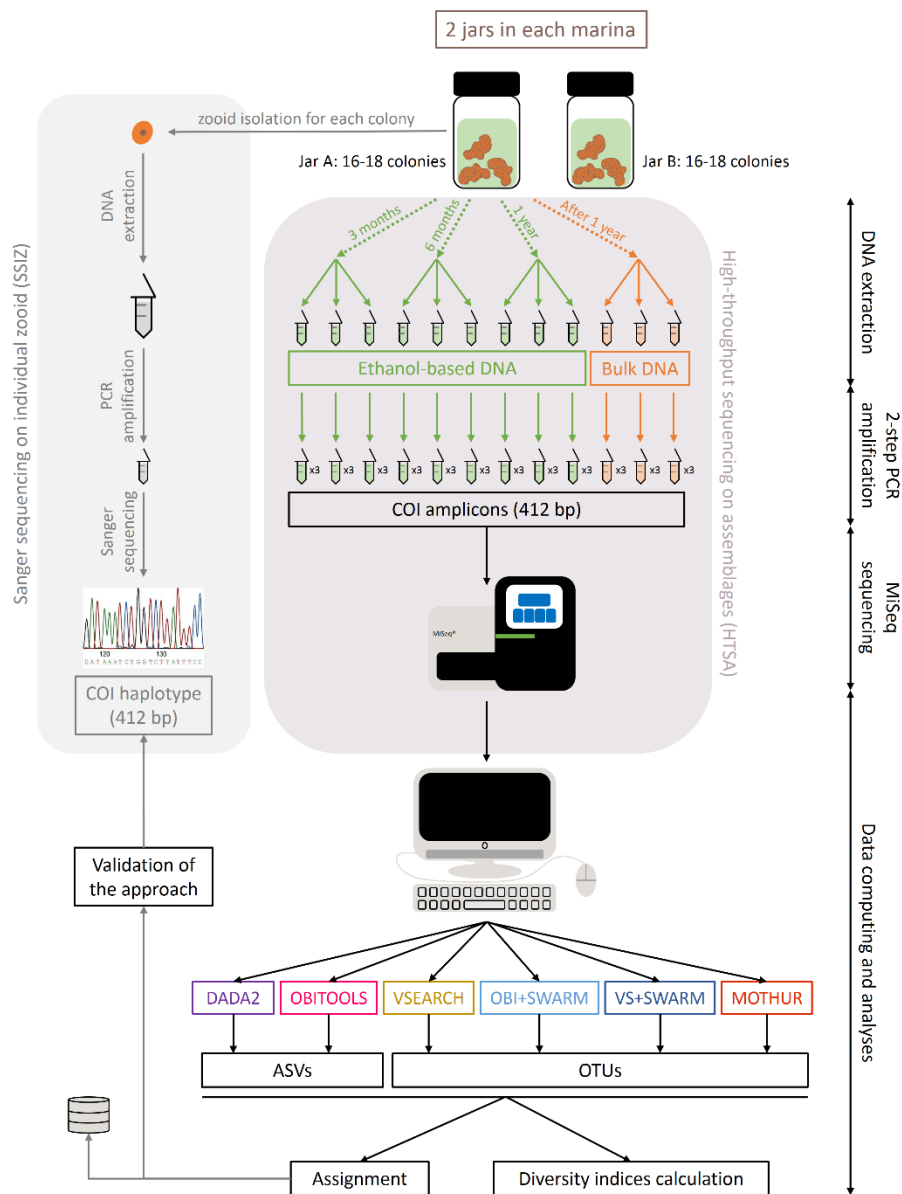
For each piece of colony preserved individually, DNA was extracted from a single zooid using the NucleoSpin® Tissue extraction kit (Macherey-Nagel) following the manufacturer's protocol with few modifications detailed in supporting information (protocol 1). A 709-bp portion of the COI gene was amplified using primers of Folmer, Black, Hoeh, Lutz, and Vrijenhoek (1994). After a first sequencing, and to improve sequence quality for 59 *B. diegensis* and 17 *B. violaceus* samples, a second amplification and sequencing was performed using primers designed by Callahan, Deibel, McKenzie, Hall, and Rise (2010) for *B. violaceus* (644-bp), and newly designed primers [Bdieg-COI-F: 5'-TGTCTACTAATCATAAAGATATTAG-3'; Bdieg-COI-R2: 5'-AATATACTTCAGGGTGTCCAA-3'] for *B. diegensis* (713-bp). Both target the Folmer region. Details are provided in protocol 1 (supporting information). Amplicons were sequenced in both directions by Eurofins Genomics (Germany GmbH) using Sanger technology. Sequences were checked and aligned using CodonCode Aligner v.5.0.1 (CodonCode Corporation, Dedham, MA). Species identification and haplotype names were provided according to Viard et al., (2019). Consecutive numbering was provided for newly discovered haplotypes.

## c. High-throughput sequencing on assemblages (HTSA)

### Sample processing

After 3, 6, and 12 months of storage, DNA was extracted from preservative ethanol (ebDNA), with three replicates of 1 mL per jar (Fig. 2). In addition, after 12 months, all colonies from a jar were blended, and DNA was extracted (bulkDNA) in three replicates (Fig. 2; supporting information, protocol 2).





**Figure 2** Overview of the experimental design from DNA extraction to data analyses. Dotted arrows represent the four different types of samples (3-, 6-, and 12-month ebDNA and bulkDNA). Data were processed with six bioinformatics pipelines. Extractions and amplifications protocols are detailed in the supporting information.

Because universal primers are commonly prone to amplification biases in metabarcoding approaches (Collins et al., 2019; Couton, Comtet, Le Cam, Corre, & Viard, 2019), *Botrylloides*-specific primers were designed to avoid any confounding factors in the recovery of target haplotypes. Since the fragment obtained with SSIZ is too long for Illumina sequencing, primers were designed to target a shorter 455-bp

portion (position 78 to 532 inside the Folmer fragment), sufficient to recover all known haplotypes (Viard et al., 2019; this study): COIBotrF2.2 – 5'-AGTGTTTYATTTCGTWTAGA-3', and COIBotrR7.1 – 5'-CAAACARAGAYATRGARAAAYAT-3'. The libraries were prepared using a dual-barcoded, dual-indexed two-step PCR procedure (Bourlat, Haenel, Finnman, & Leray, 2016) detailed in supporting information protocol 3. Briefly, each extraction replicate was amplified using three tagged-primer combinations. Three PCR products amplified with the same tagged-primer combination were pooled. Altogether, this resulted in a total of nine technical tagged replicates (i.e. three tagged-PCR replicates for each of the three extraction replicates) per sample. Then, all tagged PCR products for a given type of sample (dotted arrows; Fig. 2) were pooled and a second PCR was performed to add Nextera® indexed primers. Each sample was identified by a unique index combination. All amplicons were sequenced in-house using a MiSeq® Illumina instrument with a v3 Reagent Kit (600 cycles).

### Reads processing

The COI HTSA dataset was processed using six different pipelines (Fig. 2). DADA2 v-1.13.1 (Callahan et al., 2016) and OBITOOLS v-1.2.11 (Boyer et al., 2016) are based on denoising algorithms which remove PCR and sequencing errors and produce a set of amplicon sequence variants (ASVs). The four other pipelines use clustering algorithms producing operational taxonomic units (OTUs). VSEARCH v-2.14.1 (Rognes, Flouri, Nichols, Quince, & Mahé, 2016) and MOTHUR v-1.42.0 (Schloss et al., 2009), require an arbitrary threshold, set at 99.5% identity because of the high similarity between haplotypes. Contrarily, SWARM v-3.0.0 is free of threshold (Mahé, Rognes, Quince, de Vargas, & Dunthorn, 2015). Since SWARM only offers a clustering tool, reads preparation was performed with either the OBITOOLS (OBI+SWARM) or VSEARCH (VS+SWARM) processing tools.

False positives may arise from index-jump (Taberlet, Bonin, Zinger, & Coissac, 2018). To assess this phenomenon, 12 index combinations, not used in our PCR experiments, and chosen among the 96 available, were added to the MiSeq sequencing sample sheet in order to get the corresponding fastq files. The number of reads associated to these internal control index combinations was recorded (a maximum of 25 to 37 reads depending on the pipeline). Any ASV or OTU that did not account for more than twice the maximum number of reads in a control index

combination was discarded. Furthermore, we retained only ASVs/OTUs found in at least five out of the nine technical replicates per sample.

### d. Data analyses

#### Assignment

COI ASVs/OTUs retrieved from the HTSA dataset were compared to a database composed of 1107 reference sequences for 185 tunicate species collected from GenBank or produced locally (Couton et al., 2019). It included all known haplotypes from the three local *Botrylloides* species and BvX-H6 (Viard et al., 2019), as well as two new haplotypes found with SSIZ. Species assignment was performed using the Blast® command-line tool (Altschul, Gish, Miller, Myers, & Lipman, 1990). Only alignments covering 99% of the subject sequence were considered. If one ASV/OTU matched with several references, it was assigned to the one with the higher identity percentage. If two alignments with different references had the same identity, the ASV/OTU was classified as “unassigned”. For assignment at the haplotype level, only ASVs/OTUs which were 100% identical to one of the known haplotypes were assigned.

#### Haplotype comparison

For each pipeline and each type of sample, the proportion of reads assigned to a given haplotype in a jar was compared to the proportion of colonies associated to this haplotype by SSIZ in the same jar, using Pearson correlation with the basic stats package in R 3.4.4 (R Core Team, 2018). The effect of the pipeline and type of sample on the correlation coefficient ( $r$ ) was tested by a Friedman test, using the same package. For picturing the molecular distance between known and unassigned ASVs/OTUs, haplotype networks were built with the pegas v-0.10 R package (Paradis, 2010). Data were fourth-root transformed to reduce the impact of high abundance ASVs/OTUs on visualization.

## Diversity indices

To evaluate the reliability of ASV/OTU frequencies as infra-specific diversity descriptors, we calculated two common indices in population genetics: i) the average gene diversity per locus ( $H_S$ ) as described by Nei (1973), and ii) the population pairwise  $F_{ST}$  estimator (Weir & Cockerham, 1984), a measure of the genetic structure. Only ASVs/OTUs assigned to *B. diegensis*, the most conspicuous species, were used and data from both jars of a same marina were pooled. Computations were made using Arlequin 3.5.2.2 (Excoffier & Lischer, 2010) with either the haplotype frequencies from SSIZ or the ASV/OTU frequencies from each of the HTSA pipelines. Pearson correlation coefficients ( $r$ ) between indices computed from SSIZ and HTSA dataset were calculated using the stats package in R 3.4.4. The effect of the pipeline or the type of sample on correlation coefficients was tested by a Friedman test using the same package. Pairwise  $F_{ST}$  estimators from SSIZ and HTSA on ebDNA after 3 months of storage and processed with DADA2 were used to build a heatmap with the ggplot2 v-3.1.1 R package (Wickham, 2016) and dendrograms with the hclust function (method UPGMA) of the stats R package and the ggdendro v-0.1-20 R package (De Vries & Ripley, 2016).

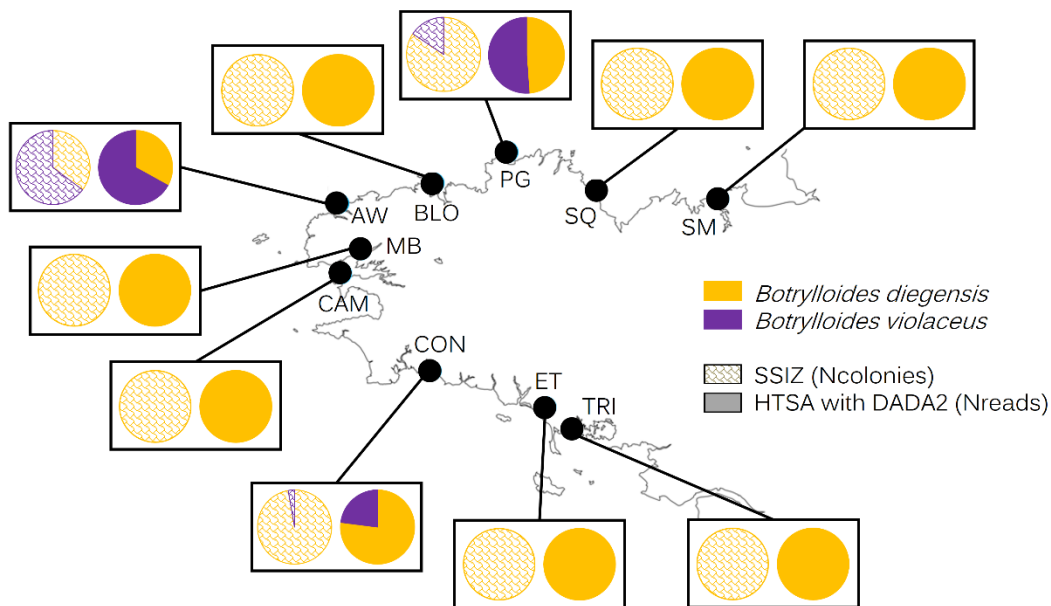
## Results

### a. Sanger sequencing on individual zooid (SSIZ)

Out of the 354 colonies, 353 were successfully amplified. The one that failed was later assigned to *Botrylloides violaceus* with cytochrome b (not shown). Only the two non-indigenous species *B. diegensis* and *B. violaceus* were present, *B. diegensis* being the most abundant (92% of the colonies; Fig. 3). Across the two species and all samples, nine haplotypes were found. Out of the seven haplotypes uncovered for *B. diegensis*, five (Bd-H1, Bd-H2, Bd-H3, Bd-H5, Bd-H6) were already reported in Viard et al. (2019), and two were new (Bd-H7 and Bd-H8). In *B. violaceus*, two haplotypes were detected, both of them already reported in Viard et al. (2019) (Bv-H1 and Bv-H4).

## b. Species assignment

None of the four negative controls of extraction and PCR contained any reads after the filtration steps. In total, the MiSeq run yielded 11,695,927 reads that globally resulted in 61 unique ASVs/OTUs, some being shared across methods. When compared to our tunicate COI database, all ASVs/OTUs were assigned to either *B. diegensis* or *B. violaceus* with more than 97% identity; 45 were assigned with more than 99% identity, the remaining 16 accounting for only 2% of the total amount of reads. In agreement with SSIZ, HTSA revealed the presence of *B. diegensis* in every location, whereas *B. violaceus* was detected in three marinas only (PG, AW, and CON; Fig. 3). The proportions of both species estimated from HTSA and SSIZ significantly differed in PG for ebDNA samples and CON for bulkDNA, but not in AW (Table S3). When different, HTSA always overestimated the abundance of *B. violaceus*.



**Figure 3** Distribution patterns of *Botrylloides diegensis* (yellow) and *Botrylloides violaceus* (purple) as uncovered by SSIZ (scale pattern) or HTSA (results from DADA2 3-month ebDNA; plain color). See Fig. 1 for location codes.

### c. Pipeline performance for HTSA-based haplotype detection

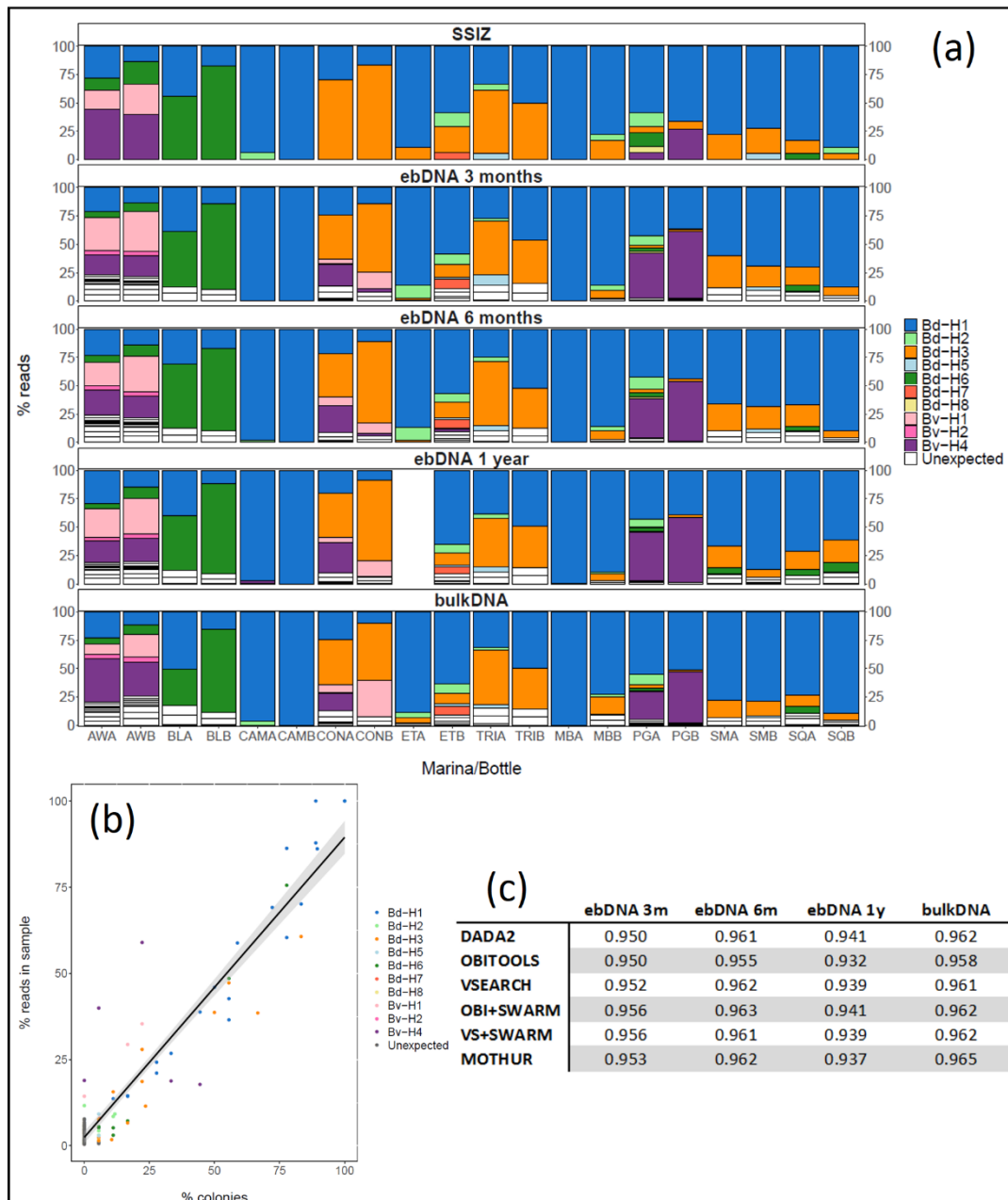
The six pipelines generated from 20 to 36 ASVs/OTUs (Table 1). This is two to four times the number of haplotypes expected from SSIZ (nine haplotypes). The five dominant haplotypes in SSIZ (Bd-H1, Bd-H3, Bd-H6, Bv-H1, Bv-H4; Fig. 4a) were retrieved by all pipelines. DADA2 retrieved all nine haplotypes but produced a high number of unexpected ASVs (20) whereas MOTHUR had the lowest number of unexpected sequences (14) but recovered only six expected haplotypes (Table 1). The proportion of reads associated with unexpected sequences was low (1.5-9%; Table 1), and most of them were not shared between pipelines (Fig. S3).

**Table 1** Number of ASVs/OTUs retained with the six pipelines, and after post-treatment corrections (index-jump, and selection on replicates). After comparison with SSIZ results, the number of expected haplotypes recovered, the names of missing haplotypes and the proportion of reads associated with unexpected sequences are indicated.

	ASVs/OTUs	Index-jumping correction	Present in at least five replicates	Expected haplotypes recovered	Missing haplotypes	% reads of unexpected sequences
<b>DADA2</b>	2115	58	29	9	-	9
<b>OBITOOLS</b>	4062	46	23	5	Bd-H2 Bd-H5 Bd-H7 Bd-H8	5
<b>VSEARCH</b>	3055	64	36	7	Bd-H2 Bd-H8	8
<b>OBI+SWARM</b>	896	46	23	5	Bd-H2 Bd-H5 Bd-H7 Bd-H8	3
<b>VS+SWARM</b>	1386	46	22	5	Bd-H2 Bd-H5 Bd-H7 Bd-H8	1.5
<b>MOTHUR</b>	3270	34	20	6	Bd-H2 Bd-H5 Bd-H8	2

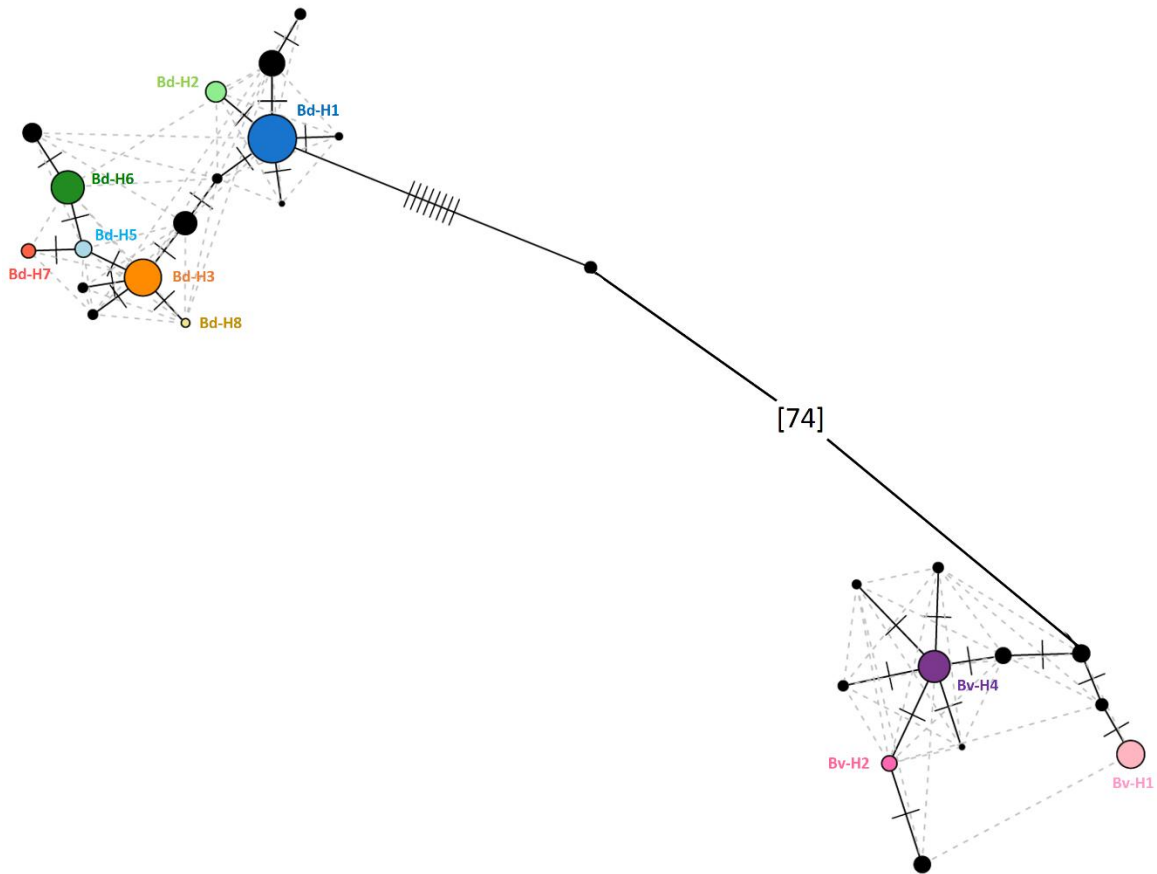
ASVs obtained with DADA2 from 3-month ebDNA (our recommended pipeline x type of sample combination; see discussion) were used to compute a haplotype network (Fig. 5). With one exception, all unexpected sequences differed by only one or two nucleotides from expected haplotypes. The ASV with an 8-bp difference from Bd-H1 was a chimera: the 381 first bases corresponded to Bd-H1 and the last 31 bases corresponded to Bv-H4 or Bv-H1. This sequence was recovered by all pipelines except MOTHUR (Figs. S4-S5).

In some cases, HTSA detected more known haplotypes (i.e. present in the database) than SSIZ. In particular, two haplotypes of *B. violaceus* (Bv-H1, Bv-H4) were detected by HTSA, with all pipelines and all sample types, in both jars from CON, where only one colony, and thus one haplotype, was associated to this species with SSIZ (jar A; unassigned colony, assigned later with *cyt b*; Fig. 4a). Additionally, one haplotype not identified anywhere by SSIZ (Bv-H2) was detected using HTSA (only with DADA2 and OBI+SWARM) in both jars from AW (Fig. 4a). Conversely, some rare haplotypes (e.g. Bd-H2 in SQB; Fig. 4a) were not always uncovered by HTSA.



**Figure 4 (a)** Proportion of colonies or reads per haplotype in each jar (A and B) for each location (see Fig. 1 for location codes), as revealed by SSIZ (top panel) or HTSA using DADA2 for the four types of samples (four lower panels). ebDNA for ETA could not be amplified after 1 year. **(b)** Correlation between the proportion of reads (DADA2, 3 months) and the proportion of colonies (SSIZ) of a given haplotype in the same jar, with 95% confidence interval in grey. **(c)** Pearson correlation coefficient for each pipeline and sample type, as shown in B. All values were significant ( $P < 2.2 \times 10^{-16}$ ).





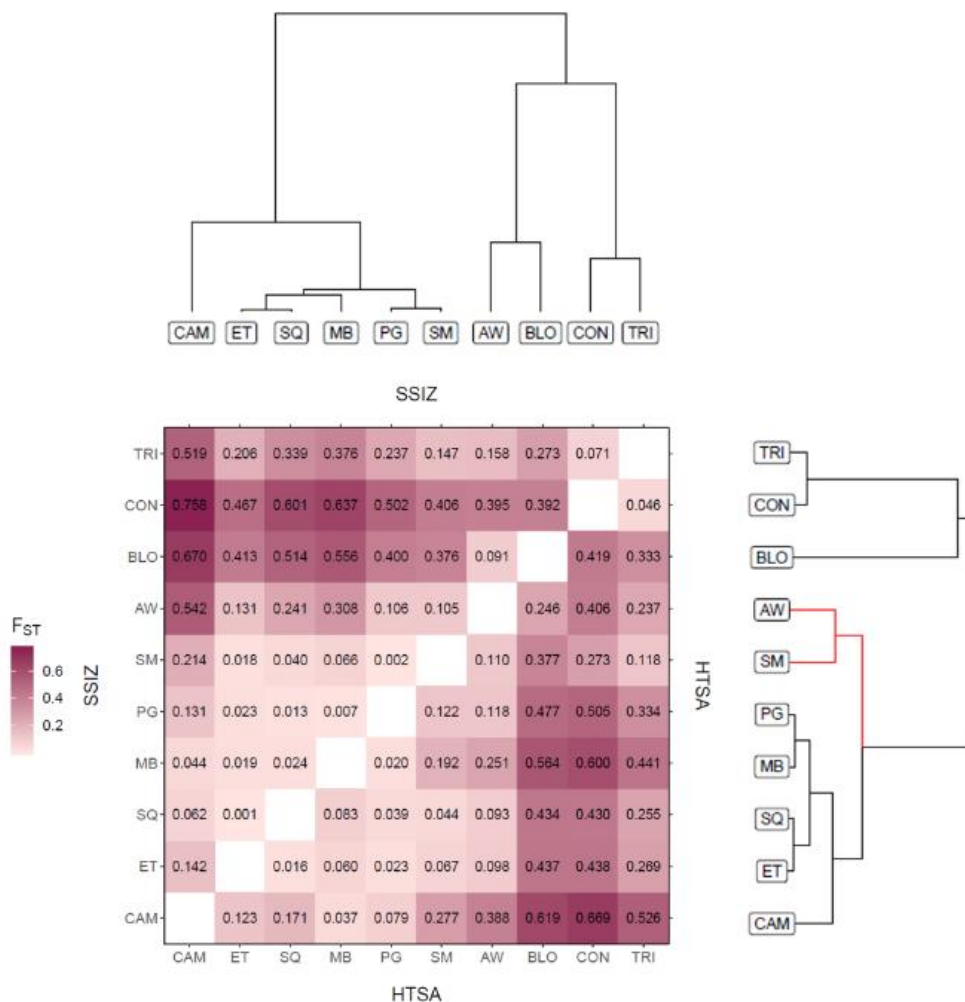
**Figure 5** Haplotype network built with ASVs produced by DADA2 on 3-month ebDNA data. Expected haplotypes are in colour, and unexpected sequences are in black. The size of the nodes represents the ASV abundance (fourth root of the number of reads) in the dataset. The number of crossing lines represents the number of mutations between two nodes. The dashed grey lines figure alternative links. The link between the two species has been shortened for visualization purposes and the 74-mutation step is written into brackets.

Haplotype distributions revealed by HTSA analyses were highly correlated to the one observed with SSIZ ( $r$  ranging from 0.932 to 0.965; Figs. 4b-c, S6-S10). Both an effect of the pipeline ( $\chi^2 = 11.462$ ;  $df = 5$ ;  $P = 0.0430$ ; Fig. S11) and the type of sample ( $\chi^2 = 16.4$ ;  $df = 3$ ;  $P = 9.387 \cdot 10^{-4}$ ; Fig. S12) were detected on correlations; the lowest correlation being observed for 1-year ebDNA processed with OBITOOLS (Fig. 4c).

## d. Population diversity indices

Haplotype diversity ( $H_S$ ) was computed for *B. diegensis* and for each marina using ASVs/OTUs abundance as individual counts (Table S4). All  $H_S$  values from HTSA were positively correlated to those obtained from SSIZ data, whatever the pipeline or the type of sample ( $r$  ranging from 0.668 to 0.935), with DADA2 showing the highest correlation values (Fig. S13). One-year ebDNA had consistently lower, although significant,  $r$  values (Fig. S13).

Pairwise  $F_{ST}$  values obtained with SSIZ and HTSA (DADA2, 3-month ebDNA) data were highly correlated ( $r = 0.941$ ;  $P < 2.2e^{-16}$ ). In both datasets, the most divergent marinas were BLO and CON (Fig. 6). Clustering locations based on their pairwise  $F_{ST}$  led to similar results with both datasets, except for AW and SM.



**Figure 6** Pairwise  $F_{ST}$  values computed from SSIZ (top left) or HTSA (DADA2 from 3-month ebDNA) (bottom right) data, and population clustering based on pairwise  $F_{ST}$ . The difference in clustering between the two datasets is highlighted in red. See Fig. 1 for location codes.

## Discussion

We here compared several bioinformatics pipelines to assess their ability to jointly analyze genetic (infra-specific) and taxonomic diversity, from High-Throughput-Sequencing (HTS) of DNA from preservative ethanol. Using samples collected in the wild, we evaluated the reliability of HTS results as compared to Sanger haplotype sequencing carried out simultaneously on the same assemblages. Below we highlight the important issues to be taken into consideration for further studies.

### **a. Ethanol-based DNA is a valid non-destructive alternative to bulkDNA, even after several months of storage**

DNA from preservative ethanol (ebDNA) has been used in a few metabarcoding studies on terrestrial or freshwater arthropods and fish (Zenker et al., 2020, and references therein). These studies showed that the amount of DNA released in ethanol differs depending on the taxa (Linard et al., 2016). Tunicates studied here might be particularly challenging in that regard: zooids are embedded in a non-cellular gelatinous tunic, composed of tunicin, which, like other polysaccharides, may decrease the amount of DNA released in ethanol and the quality of DNA extract (Aboul-Maaty & Oraby, 2019). Despite these particularities, we showed that ebDNA can be used to study marine invertebrates and is suitable for subsequent HTSA.

No major difference on haplotype distribution was observed between ebDNA and bulkDNA results (Fig. 4). However, the quality of ebDNA from one-year samples seemed lower. DNA quantification was indeed impossible after one year of storage and PCR amplifications were less efficient (several attempts have been made for every sample and no amplicon was obtained from ETA). The correlations between HTSA and SSIZ results were also lower for 1-year ebDNA, although being still significant (Figs. 4, S11-S12). These findings are congruent with those of Zenker et al. (2020) who had difficulties amplifying insect community DNA from preservative 98% ethanol after seven to fifteen months. Shokralla et al. (2010) successfully amplified plant and insect DNA from preservative 95% ethanol after storage at room temperature for seven to ten years, but specimens were preserved and sequenced individually. Because bulkDNA did not produce better data than ebDNA, and because ebDNA

allows to reuse the samples for other purposes (e.g. abundance estimation, morphological analyses), we recommend this approach for marine community analyses, preferably within six months after preservation. A longer storage might render ebDNA less reliable, a result that would agree with Barbato, Kovacs, Coleman, Broadhurst, and de Bruyn (2019), who found that bulkDNA performed better than ebDNA in recovering taxonomic diversity of rays' stomach content. However, in that case rays were first frozen (-20°C), then defrosted, and the stomach content was then preserved in ethanol. This would probably have led to the release of already-degraded DNA.

## **b. Specific primers can improve the quantitative use of HTSA data**

Population diversity indices are usually calculated from the frequency of individuals associated to each haplotype. With HTSA data, the number of reads of a particular haplotype is used as a proxy of its abundance, provided that quantification is not biased. However, several biases can occur during processing steps (DNA release in ethanol, extraction, amplification and sequencing) that can decrease the correlation between the abundance of haplotypes based on individual and read counts (Lamb et al., 2019). Bias may first arise in a mixture of organisms with various biomass and body composition. In particular, the amount of DNA released in ethanol can be highly variable depending on these two aspects (Marquina et al., 2019). Another major source of bias is the primer annealing efficiency, which can lead to differential amplification success, especially with universal primers targeting highly variable mitochondrial markers such as COI or 16S (Piñol, Senar, & Symondson, 2018). We circumvented this issue by designing a new set of primers specific to the genus *Botrylloides*. In addition to avoid any amplification failure, their size (455 bp) and position inside the Folmer fragment allowed us to encompass the same diversity than with the primers used for SSIZ. This would not have been the case with more traditional universal primers such as the ones designed by Leray et al. (2013). This 313-bp fragment would only have revealed four different haplotypes in our dataset (two for each species), thus decreasing the polymorphism that could be examined at infra-specific level. However, the use of specific primers obviously reduces the amount of information that can be collected from a complex community, and one may alternatively use multiplexes of several primer sets targeting a reduced number of taxa (Corse et al., 2019). The primer bias is well-exemplified with the data from the HTS of 16S conducted on 6-month ebDNA and bulkDNA, using the universal primers

of Kelly et al. (2016) (supporting information 4, Figs. S1-S2). Although tunicates by far dominated our assemblages, none were identified in this dataset. Instead, members of six phyla (mainly Bryozoa and Porifera) were identified, which most likely were epibionts (e.g. bryozoans) or species embedded in the *Botrylloides*' tunic (e.g. bivalves). These accompanying data on the metazoan diversity of our assemblages showed that they were more complex than simple one-genus mock communities (as built), which supports the generalisation of our approach to complex communities.

### c. Careful choice of bioinformatics pipeline is needed to examine genetic diversity

All tested pipelines described successfully the species composition (i.e. presence and relative distribution of *Botrylloides* species), as well as the overall genetic diversity of each community. However, they produced a high number of unexpected sequences, as reported in similar studies using other pipelines (Elbrecht et al., 2018; Stat et al., 2017). As a consequence, diversity indices based on haplotype counts (such as haplotype number or richness) are unreliable. Nonetheless, population genetic diversity and structure were correctly recovered when using indices based on frequency data ( $H_S$ ,  $F_{ST}$ ), because most spurious ASVs/OTUs accounted for only a small proportion of the total number of reads (1.5-9%).

Our results showed a significant effect of the pipeline on the correlation values. In pipelines that failed to retrieve some haplotypes, the missed ones were always removed at the denoising/clustering steps, except Bd-H2, which was accurately clustered with *vSEARCH* and *SWARM*-based pipelines but discarded at the index-jump and replicate filtering steps. In such cases, the threshold chosen for post-treatment filtering could be loosen in order to keep this particular haplotype but this would be at the expense of specificity since this would lead to the preservation of additional false positives. For example with *vSEARCH*, the OTU corresponding to Bd-H2 is only represented by 1 to 6 reads per sample. Keeping it would thus require not to apply an index-jump correction, which would lead to a total of 1149 OTUs at the end of the data processing steps.

Unexpected ASVs/OTUs that are slightly divergent from a haplotype might be either PCR or sequencing errors. PCR-born unexpected sequences were however unlikely as we retained only ASVs/OTUs present in at least 5 technical replicates

(combining extraction and PCR replicates). This points to the necessity of using tagged-PCR replicates to detect false positives, as also suggested by Turon, Antich, Palacín, Præbel, and Wangensteen (2020). The unexpected ASVs/OTUs might also be true haplotypes not identified by SSIZ because of colony heteroplasmy due to chimerism (induced by colony fusion). Although not reported in the studied species, chimerism is documented in *Botrylloides niger* Herdman, 1886 colonies with a prevalence of 1.9% (Sheets, Cohen, Ruiz, & da Rocha, 2016). Finally, they might also come from small fragments of other colonies that could have been put accidentally into the jar, and identified by HTSA. This could be the case for Bv-H2 in AW or Bv-H1 and Bv-H4 in CON. For instance, Bv-H1 had been observed in most samples collected in 2011 in CON (FV, unpublished data).

Sixteen ASVs/OTUs were highly divergent (< 99% identity) from the known haplotypes, and were easily manually identified as technical chimeras. The two pipelines including a chimera-removal step successfully removed most of them (all for MOTHUR, all but one for DADA2), the others retained between 11 and 14 chimeras. Contrary to Tsuji et al. (2020b) who chose not to include a chimera-removal step in their pipelines because of the high similarity between haplotypes, our results suggested that this step is crucial for limiting the number of unexpected ASVs/OTUs and does not necessarily impair the detection of true haplotypes.

#### d. Improving haplotype detection – a matter of compromise

Choosing an appropriate approach for read processing is a trade-off between removing all technical errors and keeping all true sequences. The most sensitive pipelines, able to retrieve the highest number of haplotypes (DADA2, VSEARCH), were also the ones producing the highest number of unexpected sequences. Results might be improved by fine-tuning some of the criteria used, especially for clustering or by adjusting index-jump correction and replicate-filtering thresholds. Looking at DADA2, the only method that identified all nine expected haplotypes, 18 out of the 20 unexpected ASVs were more abundant than the rarest expected haplotype, and any decrease of the thresholds would not improve our results. Other approaches have been proposed to discriminate between errors and true sequences, such as LULU, which is based on sequence co-occurrence in samples, or the protocol described in Turon et al., (2020), which is based on changes in the entropy (*sensu* Shannon entropy) ratio between the second and third codon positions. As a further denoising

step, we thus processed the ASVs produced with DADA2 with the LULU R package v-0.1.0 (Frøslev et al., 2017). This lowered the number of false positives by 35%, but removed two rare true haplotypes (Bd-H7 and Bd-H8, often not recovered by the other pipelines; Table 1). Index-jump correction and replicate filtering thus appeared efficient enough to remove most PCR and sequencing errors, as suggested by Taberlet, Coissac, Hajibabaei, and Rieseberg (2012) or Tsuji et al. (2020b), removing 98.6% of unexpected sequences produced by DADA2.

Overall, we showed that, when using community samples, ebDNA is a non-destructive alternative for a joint assessment of taxonomic and genetic diversity. For this purpose, we also recommend: 1) using specific primer sets designed to target a genus or a family, if possible multiplexed to overcome limitations in taxonomic coverage, 2) using DADA2 which includes a chimera-removal step, and 3) using post-treatment filters based on index-jump correction (by means of control index combinations) and on replicates filtering, which requires several PCR replicates.

## Acknowledgments

We thank the Diving and Marine core service from Roscoff Biological Station for sampling. We thank Gwenn Tanguy from the Biogenouest Genomer Platform for advice and access to the sequencing facilities, and the Biogenouest ABIMS Platform for access to the calculation resources. This project was supported by TOTAL foundation (project Aquanis2.0). MC acknowledges a PhD grant by Région Bretagne (ENIGME ARED project) and Sorbonne Université (ED 227 "Science de la nature et de l'homme").

## Authors' contributions

MC, EC, TC, and FV designed the study and/or the analyses; TC and FV supervised field work; MC, AB, and CDT collected the molecular data; MC, AB and FV analysed the data; MC, TC, and FV led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

## References

- Aboul-Maaty, N. A.-F. & Oraby, H. A.-S. (2019). Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. *Bulletin of the National Research Centre*, 43(1), 25 doi: 10.1186/s42269-019-0066-1.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410 doi: 10.1016/S0022-2836(05)80360-2.
- Andújar, C., Arribas, P., Yu, D. W., Vogler, A. P. & Emerson, B. C. (2018). Why the COI barcode should be the community DNA metabarcode for the metazoa. *Molecular Ecology*, 27(20), 3968-3975 doi: 10.1111/mec.14844.
- Barbato, M., Kovacs, T., Coleman, M. A., Broadhurst, M. K. & de Bruyn, M. (2019). Metabarcoding for stomach-content analyses of Pygmy devil ray (*Mobula kuhlii* cf. *eregoodootenkee*): Comparing tissue and ethanol preservative-derived DNA. *Ecology and Evolution*, 9(5), 2678-2687 doi: 10.1002/ece3.4934.
- Bock, D. G., Zhan, A., Lejeune, C., MacIsaac, H. J. & Cristescu, M. E. (2011). Looking at both sides of the invasion: patterns of colonization in the violet tunicate *Botrylloides violaceus*. *Molecular Ecology*, 20(3), 503-516 doi: 10.1111/j.1365-294X.2010.04971.x.
- Bourlat, S. J., Haenel, Q., Finnman, J. & Leray, M. (2016). Preparation of amplicon libraries for metabarcoding of marine eukaryotes using illumina miseq: The dual-pcr method. In S. J. Bourlat Ed., *Marine Genomics: Methods and Protocols* (pp. 197-207). New York, NY: Springer New York.
- 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 doi: 10.1111/1755-0998.12428.
- Bucklin, A., Steinke, D. & Blanco-Bercial, L. (2011). DNA barcoding of marine metazoa. *Annual Review of Marine Science*, 3(1), 471-508 doi: 10.1146/annurev-marine-120308-080950.
- Callahan, A. G., Deibel, D., McKenzie, C. H., Hall, J. R. & Rise, M. L. (2010). Survey of harbours in Newfoundland for indigenous and non-indigenous ascidians and an analysis of their cytochrome c oxidase I gene sequences. *Aquatic Invasions*, 5(1), 31-39 doi: 10.3391/ai.2010.5.1.5.
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A. & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13, 581 doi: 10.1038/nmeth.3869.
- Carman, M. R., Morris, J. A., Karney, R. C. & Grunden, D. W. (2010). An initial assessment of native and invasive tunicates in shellfish aquaculture of the North American east coast. *Journal of Applied Ichthyology*, 26(s2), 8-11 doi: 10.1111/j.1439-0426.2010.01495.x.
- Collins, R. A., Bakker, J., Wangensteen, O. S., Soto, A. Z., Corrigan, L., Sims, D. W., ... Mariani, S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution*, 10(11), 1985-2001 doi: 10.1111/2041-210x.13276.
- Corse, E., Tougard, C., Archambaud-Suard, G., Agnès, J.-F., Messu Mandeng, F. D., Bilong Bilong, C. F., ... Dubut, V. (2019). One-locus-several-primers: A strategy to improve the taxonomic and haplotypic coverage in diet metabarcoding studies. *Ecology and Evolution*, 9(8), 4603-4620 doi: 10.1002/ece3.5063.
- Couton, M., Comtet, T., Le Cam, S., Corre, E. & Viard, F. (2019). Metabarcoding on planktonic larval stages: an efficient approach for detecting and investigating life cycle dynamics of benthic aliens. *Management of Biological Invasions*, 10(4), 657-689 doi: 10.3391/mbi.2019.10.4.06
- De Vries, A. & Ripley, B. (2016). "Package 'ggdendro'. Create dendrograms and tree diagrams using 'ggplot2'." accessed 2020, from <https://github.com/andrie/ggdendro>.
- Elbrecht, V., Vamos, E. E., Steinke, D. & Leese, F. (2018). Estimating intraspecific genetic diversity from community DNA metabarcoding data. *PeerJ*, 6, e4644 doi: 10.7717/peerj.4644.
- Erdozain, M., Thompson, D. G., Porter, T. M., Kidd, K. A., Kreutzweiser, D. P., Sibley, P. K., ... Hajibabaei, M. (2019). Metabarcoding of storage ethanol vs. conventional morphometric identification in relation to the use of stream macroinvertebrates as ecological indicators in forest management. *Ecological Indicators*, 101, 173-184 doi: 10.1016/j.ecolind.2019.01.014.



- Excoffier, L. & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564-567 doi: 10.1111/j.1755-0998.2010.02847.x.
- Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology*, 3(5), 294-299.
- Hajibabaei, M., Spall, J. L., Shokralla, S. & van Konynenburg, S. (2012). Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecology*, 12(1), 28 doi: 10.1186/1472-6785-12-28.
- Kelly, R. P., O'Donnell, J. L., Lowell, N. C., Shelton, A. O., Samhoury, J. F., Hennessey, S. M., ... Williams, G. D. (2016). Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ*, 4, e2444 doi: 10.7717/peerj.2444.
- Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G. & Taylor, M. I. (2019). How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology*, 28(2), 420-430 doi: 10.1111/mec.14920.
- Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., ... Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, 10(1), 34 doi: 10.1186/1742-9994-10-34.
- Linard, B., Arribas, P., Andújar, C., Crampton-Platt, A. & Vogler, A. P. (2016). Lessons from genome skimming of arthropod-preserving ethanol. *Molecular Ecology Resources*, 16(6), 1365-1377 doi: 10.1111/1755-0998.12539.
- Lindgren, M., Holt, B. G., MacKenzie, B. R. & Rahbek, C. (2018). A global mismatch in the protection of multiple marine biodiversity components and ecosystem services. *Scientific Reports*, 8(1), 4099 doi: 10.1038/s41598-018-22419-1.
- Mahé, F., Rognes, T., Quince, C., de Vargas, C. & Dunthorn, M. (2015). Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ*, 3, e1420 doi: 10.7717/peerj.1420.
- Makiola, A., Compson, Z. G., Baird, D. J., Barnes, M. A., Boerlijst, S. P., Bouchez, A., ... Bohan, D. A. (2020). Key questions for next-generation biomonitoring. *Frontiers in Environmental Science*, 7(197), doi: 10.3389/fenvs.2019.00197.
- Marquina, D., Esparza-Salas, R., Roslin, T. & Ronquist, F. (2019). Establishing arthropod community composition using metabarcoding: Surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. *Molecular Ecology Resources*, 19(6), 1516-1530 doi: 10.1111/1755-0998.13071.
- Marshall, N. T. & Stepien, C. A. (2019). Invasion genetics from eDNA and thousands of larvae: A targeted metabarcoding assay that distinguishes species and population variation of zebra and quagga mussels. *Ecology and Evolution*, 9(6), 3515-3538 doi: 10.1002/ece3.4985.
- Martins, F. M. S., Galhardo, M., Filipe, A. F., Teixeira, A., Pinheiro, P., Paupério, J., ... Beja, P. (2019). Have the cake and eat it: Optimizing nondestructive DNA metabarcoding of macroinvertebrate samples for freshwater biomonitoring. *Molecular Ecology Resources*, 19(4), 863-876 doi: 10.1111/1755-0998.13012.
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70(12), 3321-3323 doi: 10.1073/pnas.70.12.3321.
- Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26(3), 419-420 doi: 10.1093/bioinformatics/btp696.
- Parsons, K. M., Everett, M., Dahlheim, M. & Park, L. (2018). Water, water everywhere: environmental DNA can unlock population structure in elusive marine species. *Royal Society Open Science*, 5(8), 180537 doi: doi:10.1098/rsos.180537.
- Pauvert, C., Buée, M., Laval, V., Edel-Hermann, V., Fauchery, L., Gautier, A., ... Vacher, C. (2019). Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal ecology*, 41, 23-33 doi: 10.1016/j.funeco.2019.03.005.
- Pedro, P. M., Piper, R., Bazilli Neto, P., Cullen, J. L., Dropa, M., Lorencao, R., ... Turati, D. T. (2017). Metabarcoding analyses enable differentiation of both interspecific assemblages and intraspecific

- divergence in habitats with differing management practices. *Environmental Entomology*, 46(6), 1381-1389 doi: 10.1093/ee/nvx166.
- Piñol, J., Senar, M. A. & Symondson, W. O. C. (2018). The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology*, 1-13 doi: doi:10.1111/mec.14776.
- Porter, T. M. & Hajibabaei, M. (2018). Over 2.5 million COI sequences in GenBank and growing. *PLOS ONE*, 13(9), e0200177 doi: 10.1371/journal.pone.0200177.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
- Rocha, R. M., Salonna, M., Griggio, F., Ekins, M., Lambert, G., Mastrototaro, F., ... Gissi, C. (2019). The power of combined molecular and morphological analyses for the genus *Botrylloides*: identification of a potentially global invasive ascidian and description of a new species. *Systematics and Biodiversity*, 17(5), 509-526 doi: 10.1080/14772000.2019.1649738.
- Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, 4, e2584 doi: 10.7717/peerj.2584.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., ... Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied Environmental Microbiology*, 75(23), 7537-7541 doi: 10.1128/aem.01541-09.
- Sheets, E. A., Cohen, C. S., Ruiz, G. M. & da Rocha, R. M. (2016). Investigating the widespread introduction of a tropical marine fouling species. *Ecology and Evolution*, 6(8), 2453-2471 doi: 10.1002/ece3.2065.
- Shokralla, S., Singer, G. A. C. & Hajibabaei, M. (2010). Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. *BioTechniques*, 48(3), 305-306 doi: 10.2144/000113362.
- Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E. D., Robinson, D. P., Knudsen, S. W., ... Thomsen, P. F. (2016). Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution*, 1, 0004 doi: 10.1038/s41559-016-0004.
- Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., ... Bunce, M. (2017). Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, 7(1), 12240 doi: 10.1038/s41598-017-12501-5.
- Stepien, C. A., Snyder, M. R. & Elz, A. E. (2019). Invasion genetics of the silver carp *Hypophthalmichthys molitrix* across North America: Differentiation of fronts, introgression, and eDNA metabarcode detection. *PloS one*, 14(3), e0203012-e0203012 doi: 10.1371/journal.pone.0203012.
- Taberlet, P., Coissac, E., Hajibabaei, M. & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, 21(8), 1789-1793 doi: 10.1111/j.1365-294X.2012.05542.x.
- Taberlet, P., Bonin, A., Zinger, L. & Coissac, E. (2018). *Environmental DNA: For Biodiversity Research and Monitoring*. Oxford: Oxford University Press.
- Tsuji, S., Maruyama, A., Miya, M., Ushio, M., Sato, H., Minamoto, T. & Yamanaka, H. (2020a). Environmental DNA analysis shows high potential as a tool for estimating intraspecific genetic diversity in a wild fish population. *Molecular Ecology Resources*, 00, 1-11 doi: 10.1111/1755-0998.13165.
- Tsuji, S., Miya, M., Ushio, M., Sato, H., Minamoto, T. & Yamanaka, H. (2020b). Evaluating intraspecific genetic diversity using environmental DNA and denoising approach: A case study using tank water. *Environmental DNA*, 2(1), 42-52 doi: 10.1002/edn3.44.
- Turon, X., Antich, A., Palacín, C., Præbel, K. & Wangensteen, O. S. (2020). From metabarcoding to metaphylogeography: separating the wheat from the chaff. *Ecological Applications*, 30(2), e02036 doi: 10.1002/eap.2036.
- Viard, F., Roby, C., Turon, X., Bouchemousse, S. & Bishop, J. D. D. (2019). Cryptic diversity and database errors challenge non-indigenous species surveys: an illustration with *Botrylloides* spp. in the English Channel and Mediterranean Sea. *Frontiers in Marine Science*, 6, 615 doi: 10.3389/fmars.2019.00615.
- Weir, B. S. & Cockerham, C. C. (1984). Estimating f-statistics for the analysis of population structure. *Evolution*, 38(6), 1358-1370 doi: 10.2307/2408641.
- Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag.

Zenker, M. M., Specht, A. & Fonseca, V. G. (2020). Assessing insect biodiversity with automatic light traps in Brazil: Pearls and pitfalls of metabarcoding samples in preservative ethanol. *Ecology and Evolution*, 10(5), 2352-2366 doi: 10.1002/ece3.6042.

Zizka, V. M. A., Leese, F., Peinert, B. & Geiger, M. F. (2019). DNA metabarcoding from sample fixative as a quick and voucher-preserving biodiversity assessment method. *Genome*, 62(3), 122-136 doi: 10.1139/gen-2018-0048.

# CHAPTER II:

How effective is metabarcoding for studying communities living in marinas?



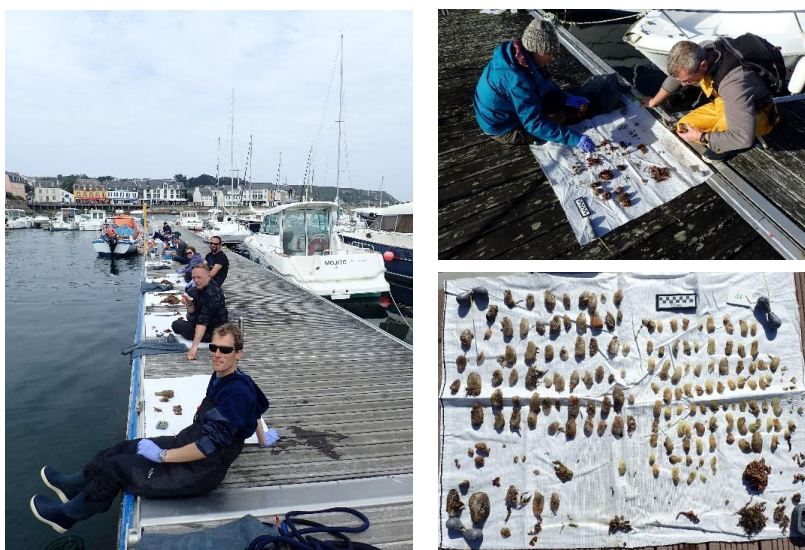
© Wilfried Thomas - SBR



## Preamble

In the first part, I showed that High-Throughput Sequencing techniques can be effective for jointly examining taxonomic and genetic diversity in invasive colonial tunicates, providing results similar to those obtained with regular barcoding. I showed the presence of two introduced species of *Botrylloides*, one of them being particularly widespread at a regional scale, with polymorphic populations. For this work, targeting a particular genus, I developed specific primers amplifying a fragment longer than what is commonly used in metabarcoding. I could thus not take advantage of the full-benefit of metabarcoding by analyzing the whole community or by using environmental samples. Metabarcoding on environmental DNA could be particularly interesting in studying marinas, which are singular and novel habitats (as compared to natural ones). In particular, they may help assessing the regional distribution of NIS and their contribution to the structure of marina communities.

In this second part, I thus examined the potential of metabarcoding, as compared to one traditional method (quadrat scraping examined in the field), to 1) identify non-indigenous species in marinas over space and time ([Chapter II.1](#)), and 2) monitor marina communities, including NIS contribution ([Chapter II.2](#)). I made a focus on sessile species found in fouling communities under floating pontoons, as many NIS are contributing to these communities. I aimed to examine to which extent the two methods were complementary, and what is the “added-value” of metabarcoding, if any.



The sampling team, in a marina, sorting scraped organisms from under the pontoon

## Non-indigenous and native species from biofouling communities in marinas: the (un)detected fraction from environmental DNA metabarcoding

**Running title:** eDNA for marine NIS detection in marinas

Marjorie Couton<sup>1</sup>, Laurent Lévêque<sup>2</sup>, Claire Daguin-Thiébaud<sup>1</sup>, Thierry Comtet<sup>1</sup>,  
Frédérique Viard<sup>1\*</sup>

<sup>1</sup> Sorbonne université, CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

<sup>2</sup> Sorbonne université, CNRS, FR 2424, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

\* Correspondence author: viard@sb-roscoff.fr; +33 2 98 29 23 12

**Keywords:** biofouling, biological invasions, detection, environmental DNA, metabarcoding, morphology, non-indigenous species, quadrats

**This part is in preparation to be submitted to the journal Environmental DNA**

## Abstract

Marinas are points of entry for non-indigenous species (NIS), and may promote the human-driven spread of NIS at a regional scale, especially through leisure boating. They are, consequently, high-priority targets for the detection of novel NIS and the monitoring of the spread of already reported NIS. To circumvent limitations of traditional methods, such as the difficulty to carry out fast and reliable morphological identification in the field, and the time needed to process many samples, metabarcoding has been proposed as an effective approach to detect NIS. Here, a joint assessment of taxa detected by a traditional method (i.e. quadrat sampling and *in situ* morphological identification) and by metabarcoding from eDNA water samples, using three distinct markers to increase its detection capacity, was carried out in ten marinas around Brittany (France). We aimed at reporting and comparing native and non-indigenous species identified by the two methods, and providing a baseline for future surveys. We focused on the fauna composing biofouling communities (i.e. sessile taxa), known to host numerous NIS. With the traditional approach, 48 taxa were identified; 39 of them at the species level. Metabarcoding revealed, as expected, a much higher number of taxa (207 taxa, incl. 114 at the species level). Among the 18 NIS and cryptogenic species recovered in the quadrats, 12 were detected in eDNA with at least one marker, whereas one-third was not recovered. These false negatives were likely due to both the sampling strategy and the bioinformatics processing steps. Furthermore, 18 species were identified with metabarcoding that were not previously reported in the study area. With one exception, all of them were most likely false positives, due to errors in taxonomic assignment of various causes. Overall, many hard-bottom invertebrates observed with traditional methods, including notorious NIS, were successfully identified with eDNA metabarcoding, highlighting the interest of this approach for NIS surveillance programs. Our results however showed that particular caution should be taken, especially when reporting novel NIS.

## Introduction

Human-mediated transportation of species outside of their native range is one of the major causes for biodiversity change and ecosystem alteration (Katsanevakis et al., 2014; Simberloff et al., 2013). Since marine species are mainly transported *via* shipping (ballast water, bilge water, hull fouling), ports act as points of entry for non-



indigenous species (NIS) (Molnar, Gamboa, Revenga, & Spalding, 2008; Ojaveer et al., 2018). At regional scales, marinas play a key role in NIS dispersal through recreational boating, which is now considered as a main spreading vector (Ferrario, Caronni, Occhipinti-Ambrogi, & Marchini, 2017; Ojaveer et al., 2018; Peters, Sink, & Robinson, 2019; Ulman et al., 2019). They also offer a wide range of artificial hard substrates that might facilitate NIS settlement and further establishment in nearby environments (Airoldi, Turon, Perkol-Finkel, & Rius, 2015; Glasby, Connell, Holloway, & Hewitt, 2007).

As the number of introductions is growing worldwide (Seebens et al., 2017), much effort has been made to avoid new species displacement (e.g. Ballast Water Management Convention) or to limit the impact of NIS on ecosystems and human activities (Giakoumi et al., 2019). Nevertheless, vectors of introductions are not all efficiently regulated (Clarke Murray, Pakhomov, & Therriault, 2011), and measures of containment or eradication are more difficult to apply when introduced species have become too abundant or widespread (Ojaveer et al., 2015). Monitoring tools are thus required to detect NIS when they are still in low abundances in the environment. In addition, as marinas suffer from high anthropization, they may not be surrogates of the neighbouring natural habitats, and especially of natural rocky reefs. They display novel assemblages, in which NIS coexist with cryptogenic and native species (e.g. Albano & Obenat, 2019). They also exhibit particular species interactions (e.g. consumptive biotic resistance; Leclerc, Viard, & Brante, 2020) that are not quite fully understood and still require investigations. In that respect, monitoring tools also need to be able to document NIS spatial and temporal distributions within biofouling communities, and to inventory coexisting native species.

Diverse protocols and methods exist to monitor marinas, from Rapid Assessment Surveys to full species inventories (Lehtiniemi et al., 2015). Traditional approaches usually depend on the identification of organisms based on their morphology and might include the collection of specimens (e.g. Bishop, Wood, Lévêque, Yunnice, & Viard, 2015; Rogers, Byrnes, & Stachowicz, 2016). They thus require taxonomic expertise, and the time constraint of field sessions can lead to erroneous identification in complex taxa. Furthermore, these approaches do not allow the detection of introductions by cryptic species (Morais & Reichard, 2018) which are very common among NIS (Viard, David, & Darling, 2016), and especially in those contributing to biofouling assemblages, such as sponges (e.g. *Haliclona* sp.; Knapp, Forsman, Williams, Toonen, & Bell, 2015), bryozoans (e.g. *Bugula neritina*; Fehlauer-Ale et al., 2014), or ascidians (e.g. *Botrylloides* sp.; Viard, Roby, Turon, Bouchemousse,

& Bishop, 2019). Finally, morphology-based assessments are not always sufficient to discriminate early settlement stages of certain groups of organisms (e.g. bivalves; Martel, Auffrey, Robles, & Honda, 2000; Meistertzheim, Héritier, & Lejart, 2017), that might be key targets for NIS early detection. To circumvent these issues, molecular barcoding has been largely adopted and is now routinely applied to confirm introduced species identification (Comtet, Sandionigi, Viard, & Casiraghi, 2015). This technique is, however, time-consuming because it still requires extensive sampling and processing of each individual separately.

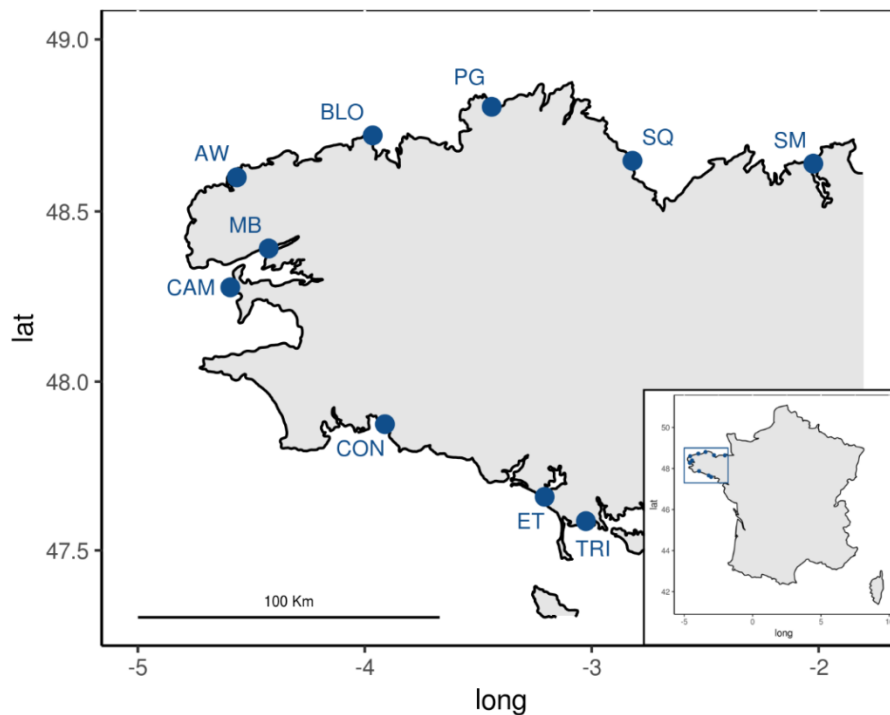
With the development of high-throughput sequencing (HTS) techniques, metabarcoding has been proposed as a new survey method for NIS detection and management in ports (Pochon, Bott, Smith, & Wood, 2013; Rey, Basurko, & Rodriguez-Ezpeleta, 2020; Zaiko, Samuiloviene, Ardura, & Garcia-Vazquez, 2015). It has been successfully applied to various sample types in this context (e.g. plankton samples; Brown, Chain, Zhan, MacIsaac, & Cristescu, 2016), sediment samples; Shaw, Weyrich, Hallegraeff, & Cooper, 2019), and water samples; Borrell, Miralles, Do Huu, Mohammed-Geba, & Garcia-Vazquez, 2017). Its use on environmental DNA (eDNA) extracted from water or sediment samples has been intensively tested (e.g. Deiner et al., 2018; Grey et al., 2018) because of the potential sensitivity and sampling simplicity of this approach. In addition to being able to detect NIS and being less costly than traditional methods (Borrell et al., 2017), it offers the possibility to grasp an image of the whole community in one sample (Stat et al., 2017). The methods, however, still require full optimization, and their limitations, especially with respect to detection failures as compared to traditional methods, still need to be investigated before including eDNA metabarcoding in routine biosurveillance protocols (Darling, Pochon, Abbott, Inglis, & Zaiko, 2020; Makiola et al., 2020).

In this study, we jointly assessed taxa in ten marinas around Brittany (France) by both a traditional method (i.e. quadrat sampling and morphological identification) and eDNA metabarcoding from water samples. We focused on fauna composing biofouling communities (i.e. sessile taxa), with a particular interest for NIS. DNA amplification was performed using three markers (18S, COI, and 16S) to increase our detection capacity. We aimed at i) reporting native and non-indigenous species detected with both methods, ii) assessing the capacity of eDNA metabarcoding to recover species observed with quadrat sampling, iii) evaluating the potential of eDNA metabarcoding to detect additional species, and iv) providing a baseline and recommendations to improve this approach for future surveys.

## Materials and methods

### a. Sampling

Biological communities fouling the immersed part of floating pontoons and seawater samples were collected in ten marinas located around Brittany (Western English Channel and NE Atlantic, France; Fig. 1). Sampling was performed in fall (October) 2017 and spring (May-June) 2018. Two pontoons were selected in each marina, in order to include potential differences in environmental conditions.



**Figure 1** Location of the sampling sites. Abbreviations are as follows: SM = Saint-Malo, SQ = Saint-Quay-Portrieux, PG = Perros-Guirec, BLO = Bloscon (Roscoff), AW = L'Aber Wrac'h, MB = Moulin Blanc (Brest), CAM = Camaret-sur-Mer, CON = Concarneau, ET = Étrel, TRI = La Trinité-sur-Mer.

Organisms were collected below pontoons every 5 m along a 50-m transect by diving. At each of the ten sampling points, divers scraped all organisms within a 0.25 m<sup>2</sup> quadrat, and placed them in a closing bag. The content of each bag was immediately sorted, identified and counted on site. Specimens (> 5 mm) were

morphologically identified to the lowest possible taxonomic level (family, genus or species-level). Solitary animals were counted whereas the abundance of colonial organisms was measured using a semi-quantitative scale with four levels (0 to 3). In order to confirm field identification, particularly for the class Gymnolaemata, a subsample of collected specimens was isolated at each site for further laboratory identification. In that case only presence-absence data were recorded.

Seawater was collected for eDNA analyses at the same exact points where organisms were scraped under pontoons. The sampling was performed before divers entered the water to avoid any contamination between sites. At each sampling point, 200 mL of marine water were collected in a plastic jar using a sterile 100 mL syringe fixed to a sampling rod and immersed at 1 m below the surface. A total volume of 2 L was thus collected per pontoon. To investigate the effect of sampled volume, in spring 2018, an additional 2-L replicate was collected, resulting in a total volume of 4 L per pontoon. Each jar was immediately filtered on site on a Millipore® Sterivex™ filter unit (0.22 µm) using a Masterflex® L/S® economy drive peristaltic pump. After filtration, 2 mL of lysis buffer (sucrose 0.75 M, Tris 0.05 M pH 8, EDTA 0.04 M) were added before storing the Sterivex unit at -20°C in a portable freezer, until transfer to the laboratory. Before all field work, all consumables not sold as DNA-free (including tubing for the filtration step) were immersed in 12.5 % commercial bleach (ca. 0.65 % hypochlorite) for at least 30 min, rinsed with ultrapure water and placed under UV light for at least 15 min. The sampling rod was immersed in 12.5 % commercial bleach for at least 30 min before sampling and rinsed with tap water between each pontoon and each site. Six sampling controls (i.e. one in three different marinas for each season) were performed by sampling 10 replicates of 200 mL of ultrapure water along a 50-m transect on a pontoon, with the same equipment, and after applying our decontamination protocol, thus mimicking the seawater sampling protocol.

## **b. DNA extraction**

At every step, caution was taken to avoid any contamination of DNA extracts with external DNA, as recommended by Goldberg et al. (2016). All equipment and bench surfaces were DNA decontaminated before use and consumables not sold as DNA-free were processed as detailed above.

Sterivex units were placed on a clean bench to thaw for 30 min before adding 100 µL of SDS (20%) and 100 µL of proteinase K (40 mg mL<sup>-1</sup>). They were then stored

in a hybridization oven at 56 °C for four hours. After this lysis step, the content of each Sterivex was pushed out in a 15 mL centrifuge tube using a sterile 10 mL syringe and 1 mL of ultrapure water was added to rinse the filter and was pushed out in the same tube using the same syringe. DNA was then extracted using the NucleoSnap® Finisher Midi extraction kit (Macherey-Nagel) following the manufacturer's protocol, except that an additional volume of 550 µL of PL3 buffer was added to each sample before adjusting the binding conditions. Elution was performed twice to increase the yield with 2x100 µL elution buffer at room temperature. Two extraction controls were performed by adding SDS and proteinase K in a Sterivex unit filled with ultrapure water. DNA was quantified by absorbance in a Spark TECAN reader using a NanoQuant Plate™. Samples were stored at -20°C for no longer than a year, until further processing.

### c. Library preparation and sequencing

All library preparation steps were performed in the respect of strict rules for preventing contamination of samples, such as UV irradiation of all tips, plates and tubes and use of filter tips. Library preparation was performed using a dual-barcoded, dual-indexed two-step PCR procedure detailed in Couton, Lévêque, Daguin-Thiébaud, Comtet, & Viard, (submitted; see [chapter I](#)). Each DNA sample was amplified with three primer pairs each targeting a different marker: i) a 365-bp COI portion amplified using primers designed by Leray et al. (2013), ii) a 389-bp to 489-bp portion of the 18S rDNA V1-V2 region amplified using primers designed by Fonseca et al. (2010), and iii) a 159-bp portion of the 16S rDNA amplified using primers designed by Kelly et al. (2016). We chose to combine these markers to increase our detection ability, as they differ in their taxonomic coverage and resolution, and conservation of primer binding sites (Couton, Comtet, Le Cam, Corre, & Viard, 2019; Makiola et al., 2020).

For COI, PCRs were performed in a total volume of 10 µL, composed of 5 µL Master Mix from the Qiagen® Multiplex PCR Plus kit, 1 µM of forward and reverse tagged primers, and 2 ng of stock DNA template. Amplification involved an initial denaturation step at 95 °C for 5 min, followed by 35 cycles at 94 °C for 50 s, 57 °C for 90 s and 72 °C for 30 s, and a final extension step at 68 °C for 10 min. For 18S, PCR was performed in a total volume of 10 µL, composed of 0.3 U of Q5® High-Fidelity DNA polymerase (New England Biolabs®, Inc.), 1X reaction buffer, 170 µM dNTPs,

0.42  $\mu\text{M}$  of both tagged primers, and 2 ng of stock DNA template. Amplification involved an initial denaturation step at 98 °C for 4 min, followed by 30 cycles at 94 °C for 1 min, 57 °C for 45 s and 72 °C for 1 min, and a final extension step at 72 °C for 10 min. For 16S, PCR reaction was performed in a total volume of 10  $\mu\text{L}$ , composed of 0.4 U of Q5® High-Fidelity DNA polymerase (New England Biolabs®, Inc.), 1X reaction buffer, 200  $\mu\text{M}$  dNTPs, 0.5  $\mu\text{M}$  of both tagged primers and 2 ng of stock DNA template. Amplification involved an initial denaturation step at 98 °C for 4 min, followed by 40 cycles at 94 °C for 50 s, 61 °C for 45 s and 72 °C for 50 s, and a final extension step at 72 °C for 10 min. PCR products were visualized on a 1.5 % agarose gel, under UV light, after ethidium bromide staining.

For each DNA sample and marker, a total of 15 PCRs were made and pooled by groups of three to constitute five replicates. Each of the five replicates was identified using a unique combination of 8-bp tags coupled with a Nextera tail added at the 5' end of each primer (Table S1). To reduce stochastic biases that could appear during amplification, the three PCRs with a given tagged-primer combination (i.e. specific to one replicate) were made using three different thermocyclers. After this first PCR step, the five tagged replicates from the same DNA extraction sample were pooled according to their intensity on the agarose gel visualisation. All pools were purified with paramagnetic beads in order to remove excess primers and putative primer dimers using the NucleoMag® NGS Clean up and Size Select kit following the manufacturer's protocol (ratio of 1:1 PCR products *vs.* beads). Three PCR negative controls were performed for each marker.

A second PCR was performed to complete Illumina® adapters and insert an index allowing sample identification. We used the list of indexed adapters described in the Illumina® Nextera XT library preparation protocol (8 indexes i5 and 12 i7 allowing 96 combinations; Table S2). PCR reactions were performed in a total volume of 10  $\mu\text{L}$  composed of 0.4 U of Q5® Hotstart High-Fidelity DNA polymerase (New England Biolabs®, Inc.), 1X reaction buffer, 200  $\mu\text{M}$  dNTPs, 0.13  $\mu\text{M}$  of each primer and 1  $\mu\text{L}$  DNA template. Amplification involved an initial denaturation step at 98 °C for 4 min, followed by 12 cycles at 98 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s, and a final extension step at 72 °C for 5 min.

After the second PCR, all products were checked on a 1.5% agarose gel and pooled according to their intensity on the agarose gel. The pool was purified by paramagnetic beads using the NucleoMag® NGS Clean up and Size Select kit following manufacturer's protocol (ratio 1:1 PCR products *vs.* beads). Quantification

of the library was performed with qPCR using the NEBNext®Library Quant Kit for Illumina® (New England Biolabs®, Inc.) and a DNA profile was performed using a DNA 1000 chip in a 1200 Bioanalyzer equipment (Agilent technologies, Inc.). Sequencing was performed for each marker separately on an Illumina® MiSeq sequencer. The COI and 18S markers were sequenced using a 600 cycles v3 protocol and the 16S marker was sequenced using a 300 cycles v2 protocol, both with two index reads.

### d. Reads processing and filtering

The three datasets produced were processed similarly. After the first demultiplexing step performed by the sequencing machine, reads were associated to their sample according to their index reading. Then, reads were assigned to one of the five PCR replicates according to their tag combination, and primers and tags were removed using CUTADAPT v-2.8. Then a set of amplicon sequence variants (ASVs) was produced using DADA2 v-1.13.1 (Callahan et al., 2016) (see Table S3 for detailed parameter values). When demultiplexing, reads can be falsely assigned to a sample when an index combination is not correctly recognized, which is referred to as “index-jump” (Taberlet, Bonin, Zinger, & Coissac, 2018). To assess this phenomenon, 12 index combinations, not used in our PCR experiments, and chosen among the 96 available, were added to the MiSeq sequencing sample sheet in order to get the corresponding fastq files. For each ASV, the number of reads assigned to one of these internal control index combinations was divided by the total number of reads in the complete dataset for this same ASV. The maximal proportion occurring in an index control was recorded for each marker (Table S3). Any ASV that did not account for more than this maximal proportion in a sample was discarded from this sample. Furthermore, we retained only ASVs found in at least two out of the five PCR replicates per sample.

### e. Taxonomic assignment

A first taxonomic assignment was performed to assess the proportion of reads assigned to a metazoan taxon. For that purpose, each dataset was aligned against references retrieved from the GenBank nt database using the ecotag command from

the OBITOOLS v-1.2.11 package (Boyer et al., 2016) with no minimum identity threshold. To perform taxonomic assignment at lower taxonomic levels, assignment was later performed by aligning all ASVs against a restricted reference database using the BLAST® command line tool (Altschul, Gish, Miller, Myers, & Lipman, 1990). This database was composed of sequences retrieved from GenBank and sequences produced locally (Couton et al. 2019; Table S4), targeting ten metazoan phyla: Annelida, Arthropoda (only crustaceans), Bryozoa, Chordata (only fish and tunicates), Cnidaria, Echinodermata, Mollusca, Nemertea, Platyhelminthes, and Porifera. These phyla were selected because most of them were observed in the scraped samples, and all are commonly found on or near floating pontoons. Only alignments covering 99% of the subject sequence and displaying a minimum identity threshold (18S: 99%; COI: 92%; 16S: 97%) based on barcoding gaps computed from reference sequences (Figs. S1-S3), were considered. If one ASV matched with several references, it was assigned to the one with the higher identity percentage. If two alignments with different references had the same identity, the ASV was assigned to the lowest common taxonomic rank. All assignments assigned to a rank higher than the family were classified as “unassigned”.

#### f. Comparison between quadrats and metabarcoding dataset

In order to evaluate the ability of eDNA metabarcoding to detect species, and particularly non-indigenous species (NIS), that were identified in quadrats, we chose to restrict the comparison to four classes of organisms (Ascidiacea, Bivalvia, Gastropoda, and Gymnolaemata). We previously established a list of putative non-indigenous species in the study area, and obtained local references for native species and NIS in these four groups (Couton et al., 2019), thus minimizing issues due to the incompleteness of public reference databases. These classes are also representative of diverse phylogenetic, ecological and functional diversity, and include a substantial number of NIS in the study region, as shown in previous surveys (Bishop et al., 2015). To look for putative NIS not identified during the quadrats analysis, all species detected *via* metabarcoding were attributed a status “expected” if they were listed in the Roscoff biological species inventory (<http://abims.sb-roscoff.fr/inventaires/>), reported before by our team members and/or in the literature. A first comparison of the detection ability of both methods was performed on presence/absence data. Then, for six solitary taxa (*Ascidiella* spp., *Asterocarpa humilis* (Heller, 1878), *Ciona intestinalis* (Linnaeus, 1767), *Ciona robusta* Hoshino & Tokioka, 1967, *Corella eumyota*



Traustedt, 1882, and *Styela clava* Herdman, 1881), abundances were compared. These taxa were chosen because of their wide distribution in our study area, their cosmopolitan distribution, and the fact that they were identified in the morphology-based and the metabarcoding-based datasets (with the 18S marker). Pearson correlation coefficients were calculated for each species by comparing the proportion of 18S reads and the number of individuals observed within quadrats for each sample.

## Results

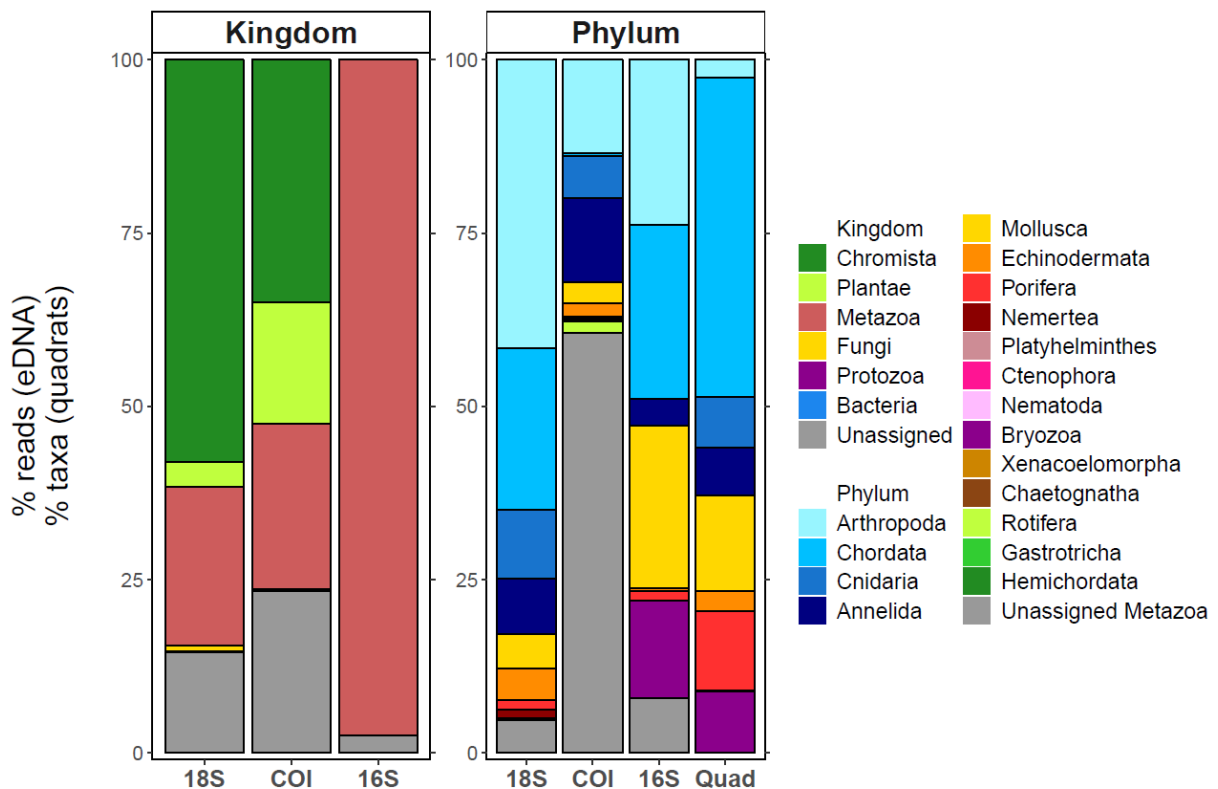
### a. Reads processing and taxonomic assignment

The MiSeq runs yielded 15,754,233 reads, 13,108,322 reads, and 13,396,162 reads for 18S, COI, and 16S respectively, resulting in 3972, 4258, and 347 unique ASVs. After the filtering steps, none of the eleven negative controls of sampling (6), extraction (2) and PCR (3) contained any reads for 18S. For COI, three sampling negative controls contained a total of 183 reads corresponding to four ASVs. The best hit for three of them were fungi species (*Verticillium dahliae* Kleb., 1913 or *Sydowia polyspora* (Bref. & Tavel) E. Müll., (1953) which are both Ascomycota) but their identity percentage was lower than 90%, and the fourth one was 100% identical to a sequence from an insect genus (*Bryophaenocladus* Thienemann, 1934). Only one sampling negative control exhibited reads (n=7467) for 16S. All were associated with a single ASV assigned to a human sequence with 100% identity.

**Table 1** Number and proportion (in parentheses) of ASVs and reads assigned to one of the reference sequence from the restricted database targeting ten metazoan phyla, for each marker (18S, COI, and 16S). The number indicated for "Taxa" includes assignments to species-, genus- and family-level. The last three columns indicate the number of species, the number of genera and the number of family detected for each marker.

	ASVs	Reads	Taxa	Species	Genus	Family
<b>18S</b>	389 (9.8%)	1,881,029 (15.8%)	291	203	233	159
<b>COI</b>	943 (22.1%)	706,023 (7.0%)	286	235	221	177
<b>16S</b>	173 (49.9%)	7,311,390 (54.6%)	104	65	78	70

When aligned against the GenBank nt database using the ecotag tool, only a small proportion of reads were assigned to Metazoa for 18S (23%) and COI (24%), which contrasted with the results for 16S, for which 98% of reads were assigned to Metazoa (Fig. 2). Whatever the assignment level (i.e. kingdom or metazoan phylum), a high proportion of reads remained unassigned with COI (23% and 61% respectively). The most abundant metazoan phyla identified across markers were Arthropoda, Chordata, Cnidaria, Annelida, and Mollusca, for a total of 88%, 35% and 76% of the reads for 18S, COI and 16S, respectively. The most abundant amplified phyla, however, were not the same among markers. For instance, bryozoans accounted for 14% of 16S reads but were assigned to less than 1% of reads for the two other markers. These discrepancies were also observed when assigning ASVs to the ten targeted metazoan phyla. For instance, 18S identified almost three times more taxa, and COI almost four times more species than 16S (Table 1).

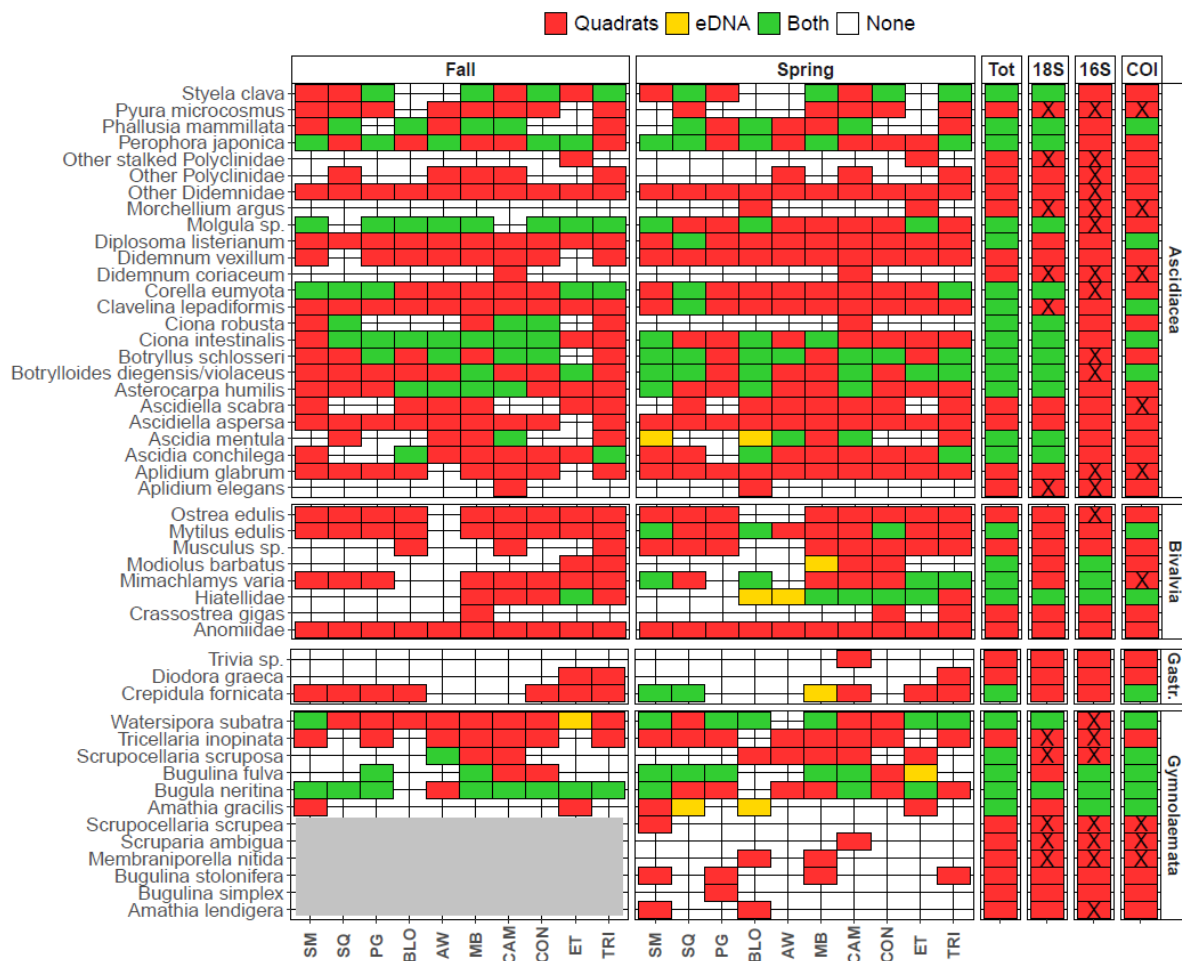


**Figure 2** Percentage of reads assigned to different kingdoms (left) or metazoan phyla (right), over all samples, for each of the three markers used in this study, as well as for Quadrats data (Quad).

## b. Comparison of quadrats (morphology-based) and eDNA metabarcoding datasets

Following scraping and morphological identification in the field, 48 taxa were reported across all study sites in the four targeted classes. Most (n=39) were identified at the species level, four at the genus level, and five at the family level (three in Ascidiacea, and two in Bivalvia). More than half (n=25) were ascidians. With eDNA metabarcoding, 207 taxa from the same four classes were identified across samples and markers, 11.6% of them being observed within quadrats. Interestingly, half of the taxa observed in quadrats were recovered in the eDNA metabarcoding dataset (column Tot in Fig. 3). Six taxa could not be recovered because they did not have any reference sequences in our database. A higher number of taxa observed within quadrats were identified with 18S (15) as compared to COI (13) and 16S (6), with 18S being more efficient for Ascidiacea whereas the two latter were more suited for Bivalvia and Gymnolaemata (last three columns in Fig. 3). Among all taxa observed with eDNA metabarcoding, 114 were identified at the species-level across all three markers (Table S4).

The ability of the two methods to identify species varied according to taxa. For example, the morphology-based approach allowed the discrimination of the two tunicate species *Ascidella aspersa* (Müller, 1776) and *Ascidella scabra* (Müller, 1776) when the metabarcoding could only identify them at the genus level, because the only marker without amplification failure, 18S, exhibited identical references for the two species. A similar situation occurred with *Bugulina stolonifera* (Ryland, 1960) and *Bugulina simplex* (Hincks, 1886) since they have the same sequence for the only marker with an available reference (16S). For other taxa, metabarcoding efficiently identified species when only the genus could be recovered in the field. This was the case, for example, for the three *Molgula* species: *Molgula socialis* Alder, 1863, *Molgula bleizi* Lacaze-Duthiers, 1877, and *Molgula complanata* Alder & Hancock, 1870.



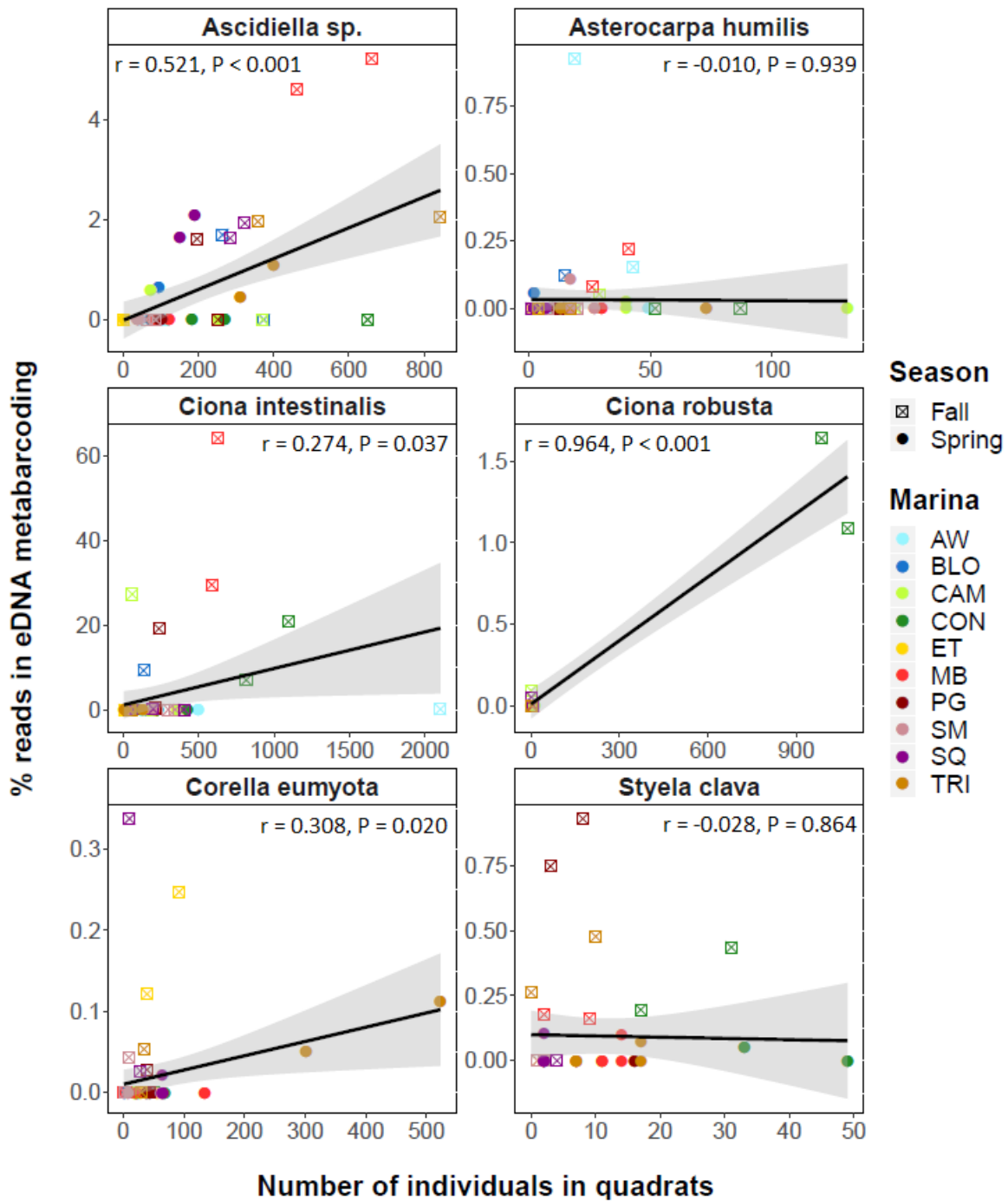
**Figure 3** Comparison of eDNA metabarcoding and morphology-based detection. Only the taxa observed within quadrats and belonging to the four targeted classes of organisms are considered here. Taxa found with both methods within the same marina are displayed in green, those found only in quadrats are in red, and those only found in eDNA are in yellow. The first two panels correspond to samples from fall 2017 and from spring 2018, respectively. Metabarcoding detection is recorded across all markers. The column Tot represents the presence of each taxon in the whole dataset. The last three columns represent the presence of a taxon in the whole dataset (all marinas and dates combined) but for each marker separately. Black crosses indicate that no reference sequence was available for a particular taxon. The grey area represents species that were not isolated during morphological identification in the fall 2017. They were all pooled in a unique category called “erected bryozoans”. See figure 1 for location codes.

Within quadrats, 15 NIS and 3 cryptogenic species were identified. Twelve (66.7 %) of them were detected in eDNA samples with at least one marker (Table 2). Four (*Botrylloides violaceus* Oka, 1927, *Crassostrea gigas* (Thunberg, 1793), *Bugulina stolonifera*, and *Bugulina simplex*) could not be identified with metabarcoding because their reference sequences were identical to sequences of closely related

species, resulting in the assignment at the genus level. The other two NIS (*Didemnum vexillum* Kott, 2002, and *Tricellaria inopinata* d'Hondt & Occhipinti Ambrogi, 1985) remained undetected despite the availability of species-specific references, their large abundance in some locations, and their presence in almost all localities at both seasons. In the metabarcoding dataset, 18 species were not expected (i.e. never reported in the study area, and not seen in the quadrats). Sixteen of these 18 unexpected species were assigned with more than 99% identity with seven being assigned with 100 % identity. All were however identified with only one marker, except *Mytilus trossulus* Gould, 1850, identified with COI and 16S.

**Table 2** List of non-indigenous (NIS) and cryptogenic (Crypto) species within the four target classes, known from the study area, and either observed in quadrats or identified in eDNA samples. Stars indicate that sequences were assigned to the genus or to another species of the same genus. The absence of reference sequence for each marker is indicated.

Class	Species	Status	Observed in quadrats	Detected in eDNA samples		
				18S	COI	16S
Ascidacea	<i>Asterocarpa humilis</i>	NIS	yes	yes	no	no
Ascidacea	<i>Botrylloides diegensis</i>	NIS	yes	yes	yes	no
Ascidacea	<i>Botrylloides violaceus</i>	NIS	yes	no*	no*	no
Ascidacea	<i>Botryllus schlosseri</i>	Crypto	yes	yes	no	no
Ascidacea	<i>Ciona robusta</i>	NIS	yes	yes	yes	no
Ascidacea	<i>Corella eumyota</i>	NIS	yes	yes	no	no
Ascidacea	<i>Didemnum vexillum</i>	NIS	yes	no	no	no
Ascidacea	<i>Diplosoma listerianum</i>	Crypto	yes	no	yes	no
Ascidacea	<i>Perophora japonica</i>	NIS	yes	yes	no	no
Ascidacea	<i>Styela clava</i>	NIS	yes	yes	no	no
Bivalvia	<i>Crassostrea gigas</i>	NIS	yes	no*	no	no
Bivalvia	<i>Mya arenaria</i>	NIS	no	yes	no	no*
Bivalvia	<i>Ruditapes philippinarum</i>	NIS	no	yes	yes	yes
Gastropoda	<i>Crepidula fornicata</i>	NIS	yes	no	yes	no
Gymnolaemata	<i>Bugula neritina</i>	NIS	yes	yes	yes	yes
Gymnolaemata	<i>Bugulina fulva</i>	Crypto	yes	no	yes	yes
Gymnolaemata	<i>Bugulina simplex</i>	NIS	yes	no	no*	no*
Gymnolaemata	<i>Bugulina stolonifera</i>	NIS	yes	no	no*	no*
Gymnolaemata	<i>Tricellaria inopinata</i>	NIS	yes	no reference	no	no reference
Gymnolaemata	<i>Watersipora subatra</i>	NIS	yes	yes	yes	no reference *



**Figure 4** Correlation between the number of individuals counted in all quadrats from a pontoon and the proportion of 18S reads within the sample of the same pontoon, for six tunicate species. Results of Pearson correlation tests are indicated within each plot. See figure 1 for location codes.

Even if half of the taxa observed within quadrats were recovered by eDNA metabarcoding over all samples, discrepancies between the two methods were observed in the distribution of given taxa across sampling localities and seasons

(Fig. 3). Successful recovery in eDNA samples of taxa observed in the field occurred only in 130 out of 556 cases (23.4 %). Conversely, eDNA metabarcoding allowed the identification of taxa in localities where they were not reported based on morphology in 10 additional cases (e.g. *Amathia gracilis* (Leidy, 1855) in SQ or BLO during spring; Fig. 3). Focusing on NIS or cryptogenic species, out of the 240 occurrences in quadrats, only 81 (33.8 %) were successfully recovered via eDNA metabarcoding (Fig. 3). Conversely, there were only three occurrences of a NIS or cryptogenic species solely recovered by eDNA (*Crepidula fornicata* in MB, *Watersipora subatra* in ET, and *Bugulina fulva* in ET).

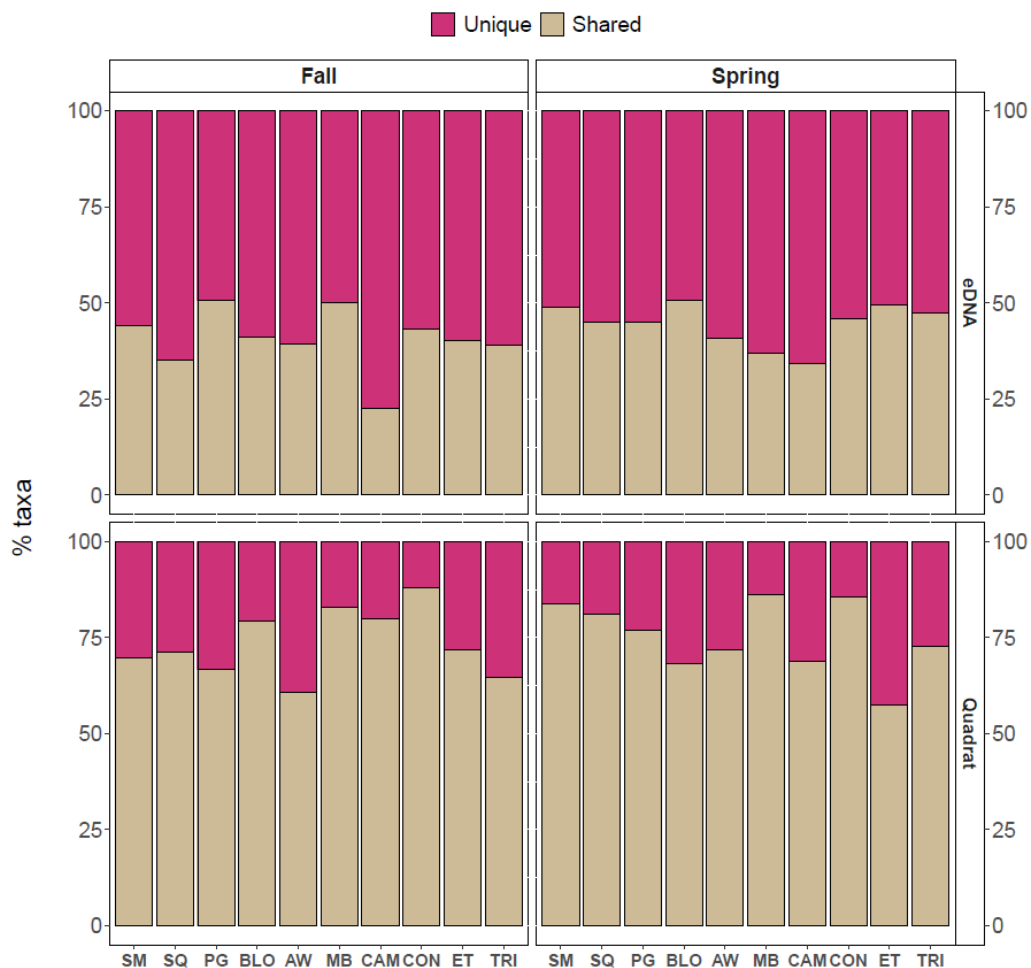
Focusing on six ascidian taxa, contrasting results were obtained on the correlation between individual counts and relative abundance of reads (Fig. 4). Significant correlations ( $P < 0.001$ ) were observed for *Ascidiella* sp. (Pearson correlation coefficient  $r = 0.521$ ) and *Ciona robusta* ( $r = 0.964$ ). For the latter, the correlation reflected two extreme situations with sites where the species is rare and sites where it is very abundant. For *Ciona intestinalis* and *Corella eumyota*, although significantly different from 0, the Pearson correlation coefficient was low. For *Asterocarpa humilis* and *Styela clava* no significant correlation was observed.

### c. Effect of the sampling strategy on taxa detection

In order to take into account putative variations in environmental conditions, and thus species presence, two pontoons were sampled in each marina. On average, less than half of the taxa (42.5 %) were identified simultaneously at the two pontoons of a given locality with eDNA metabarcoding (shared; Fig. 5). The smallest proportion was found in CAM during fall where only 12 of the 53 identified taxa (22.6 %) were shared between pontoons. Conversely, the highest proportion (50.8 %) was found in PG in fall, where 33 taxa out of 65 were shared between pontoons. The proportion of shared taxa between pontoons was much higher within quadrats ranging from 57.4 % in ET in spring to 87.8 % in CON in fall, with an average of 74.3 % (Fig. 5).

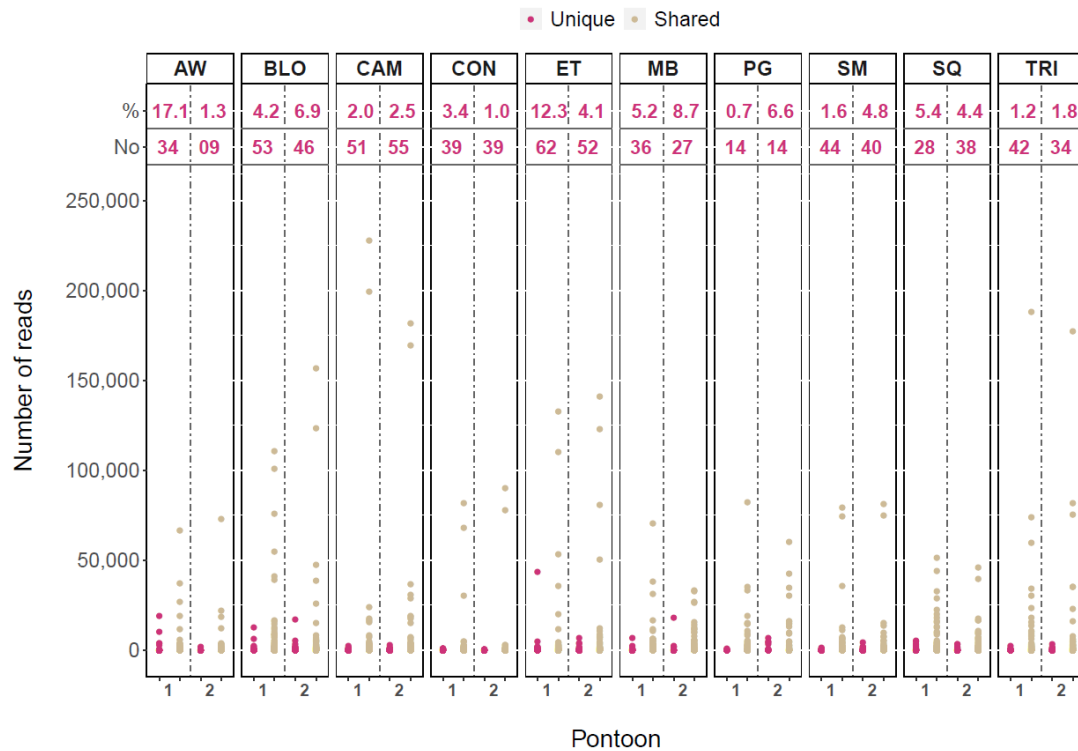
Regarding sampling strategy, in spring 2018, two 2-L replicates were sampled to test for the effect of the volume of water filtered on taxa detection. The total number of assigned taxa was higher when merging the results obtained with the two replicates (482 taxa) than when using only one of them (436 and 428, for the two series analysed separately), over all marinas. This result is consistent at every

pontoon, with a significantly higher number of taxa identified when using 4 L of water than when using one of the 2 2-L samples (Wilcoxon tests for paired data:  $V = 0$ ,  $P < 0.001$ , and  $V = 15$ ,  $P < 0.001$ , for the two series). The number of taxa unique to one 2L-jar ranged from 9 for pontoon 2 at AW to 62 for pontoon 1 at ET (Fig. 6). However, the proportion of reads assigned to these taxa, unique to a given 2L-jar, is most often very low (mean value: 1270 reads) as compared to taxa shared between the two jars for a given pontoon (mean values: 7268 reads), ranging from 1% for pontoon 2 at CON to 17.1% for pontoon 1 at AW (Fig. 6). Based on read numbers per taxon, all abundant species were successfully recovered in each of the two 2-L-jars. One noticeable exception is the copepod *Temora longicornis* (Müller O.F., 1785), which is found in only one replicate from the first pontoon of Étél, with 43407 reads. Rare species, however, were found in both or only one 2-L replicate.



**Figure 5** Proportion of taxa recovered with eDNA metabarcoding across all markers or within quadrats, in samples from the two pontoons of a marina for a given season (shared) or in samples from only one pontoon (unique). See figure 1 for location codes.





**Figure 6** Distributions of read abundances (sum of all markers) for taxa found either in both 2-L replicates (Shared) or in only one replicate (Unique) for each pontoon. The number of taxa found in only one of the two 2L-replicates is indicated at the top of each panel (No) as well as the proportion of reads assigned to these taxa (%), for each pontoon in each marina. See figure 1 for location codes.

## Discussion

The growing “hardening” of marine coastal areas offers a wide range of new artificial substrates for fouling organisms, many of them being non-indigenous species (NIS). These new communities are thus particularly targeted by biodiversity monitoring programs, especially those focusing on introduced taxa. Here, we investigated the benefits and limitations of eDNA metabarcoding from seawater as compared to traditional methods (i.e. data collected by scraping organisms) in marinas. Over all locations, many sessile metazoan species observed with traditional methods, including notorious NIS, were successfully detected with eDNA metabarcoding but substantial discrepancies were observed locally. Although it allowed the detection of additional taxa, including NIS, eDNA metabarcoding did not recover most of the taxa observed in each location. Possible approaches for increasing eDNA metabarcoding accuracy and its potential use for biodiversity surveys and NIS surveillance will be discussed in the following sections.

### a. Taxon detection with eDNA metabarcoding is effective but major flaws challenge NIS survey

The use of eDNA metabarcoding allowed us to recover 63% of the 41 species, and 67% of the 18 NIS or cryptogenic species, observed within quadrats (Fig. 3, Table 2), for the four selected classes. These substantial numbers, as well as those retrieved in previous studies, confirm that metabarcoding on environmental samples is able to detect species from fouling communities, including NIS (e.g. Borrell et al., 2017; von Ammon et al., 2018). Nevertheless, more than one third of the species observed in quadrats were missing in the eDNA dataset, including abundant ones. This is well exemplified with the solitary tunicate *Ascidella aspersa*, native in our area and known as a notorious invasive species in other coastal areas in the world (e.g. USA, South-Africa, Australia; <http://invasions.si.edu/nemesis/browseDB/SpeciesSummary.jsp?TSN=159213>). It was found in every sample over the two sampling periods, reaching up to 96 individuals in one 0.25m<sup>2</sup> quadrat in MB, but was lacking in the eDNA-metabarcoding dataset. Similarly, among the six NIS not recovered with eDNA metabarcoding (*Botrylloides violaceus*, *Didemnum vexillum*, *Crassostrea gigas*, *Bugulina simplex*, *Bugulina stolonifera*, and *Tricellaria inopinata*), the invasive bryozan *T. inopinata* was abundant and observed within quadrats from every location (Fig. 3).

The high proportion of false negatives emphasizes the value of traditional methods and highlights the occurrence of several flaws that may impede the extensive use of eDNA-metabarcoding. They include the species-specific DNA shedding rate (Barnes & Turner, 2016), the differential primer binding efficiency between taxa (Piñol, Senar, & Symondson, 2018), and issues related to the taxonomic assignment of recovered ASVs. The accuracy of the latter, for example, can be hindered by the lack of resolution of the markers used. Among the six missing NIS or cryptogenic species, four (*B. violaceus*, *C. gigas*, *B. simplex* and *B. stolonifera*) could not be reliably assigned with 18S or 16S because of identical sequences with congeneric species. A similar situation occurred for *Ascidella aspersa* and *A. scabra* that shared the same sequence for 18S, a marker known to have a low species resolution (Andújar, Arribas, Yu, Vogler, & Emerson, 2018; Clarke, Beard, Swadling, & Deagle, 2017). This issue, however, was addressed in this study by combining several markers as suggested by Zhang, Chain, Abbott, and Cristescu (2018). The benefits of this approach were exemplified by the detection of *Crepidula fornicata* (Linnaeus, 1758), which was not recovered with 18S because one reference sequence for this

species was identical to those of ten other gastropod species. It was, however, identified with COI with 100% identity.

Since the presence of false negatives in a metabarcoding dataset can also be explained by the unavailability of reference data, an effort was made, in this study, to produce reference sequences for at least one marker for each NIS observed within quadrats. None of the six undetected NIS or cryptogenic species could thus be attributed to this limitation in our results. Six of the native species, however, lacked references for all markers, and were thus not detected. Interestingly, gaps in reference databases were not solely responsible for the production of false negatives but might also have generated false positives. The absence of reference for a species present in our dataset might have caused the erroneous assignment of its sequence to a closely-related species for which a reference is available. For example, one 18S ASV was assigned with 99.16% identity to *Clavelina meridionalis* (Herdman, 1891), a species known from the southern hemisphere and unreported in Europe. This ASV, however, was most likely erroneously assigned because no 18S reference sequence for *C. lepadiformis* (Müller, 1776), a native species common in our quadrats, was available. This hypothesis is further supported by the detection of *C. lepadiformis* with COI (100 % identity), a marker able to reliably distinguish these two species (Pérez-Portela & Turon, 2008). The availability of reference sequences can vary greatly depending on the marker, with COI being the most populated in public databases (Andújar et al., 2018; Porter & Hajibabaei, 2018) or depending on the taxonomic group. For instance, a barcode search in BOLD revealed that only 14 % and 5 % of the 2925 and 5651 accepted species, for the classes Ascidiacea and Gymnolaemata respectively, were associated to a COI reference sequence (WoRMS and BOLD search in August 2020). The proportion of represented species in public databases seem, however, higher for invasive species. Briski, Ghabooli, Bailey, and MacIsaac (2016) reported proportions of 59 % and 72 % of 45 and 36 invasive species, for the same two classes and marker respectively (list of invasive species based on 55 papers). These higher values might be explained by the long-lasting and still increasing use of DNA barcoding in the study of marine NIS (Duarte, Vieira, & Costa, 2020). Such an imbalance in the proportion of sequences available for native and non-indigenous species might, in turn, increase the risk of false detections of new introductions, as in the example above for *Clavelina* species.

The reliability of the taxonomic affiliation of reference sequences is as important as their availability to avoid both false positives and false negatives (Harris, 2003; Viard et al., 2019). Another interesting case of a false positive may rely on

evolutionary processes that are usually not considered in metabarcoding approaches, and is exemplified by the assignment of COI ASVs to the bivalve *Mytilus trossulus*. The ASVs were assigned to *M. trossulus* although the sequences were also very similar to *M. edulis* references (only one missing base on the query cover). The reference sequences for *M. trossulus* corresponded to individuals from populations originating from the Baltic Sea, a well-known hybrid zone between *M. trossulus* and *M. edulis*. One particular feature of this hybrid zone is that the mitochondrial genome of *M. trossulus* from the eastern Baltic Sea has been fully introgressed by the mitochondrial genome of *M. edulis*, so that the COI of the Baltic *M. trossulus* is in fact the COI of *M. edulis* (e.g. Kijewski, Zbawicka, Väinölä, & Wenne, 2006). Aligning our ASVs with sequences of *M. trossulus* from Scotland, which have retained their ancestral mitochondrial genome (Zbawicka, Burzyński, Skibinski, & Wenne, 2010), led to 83.4% identity only, confirming that these ASVs were not *M. trossulus*, but *M. edulis*. Finally, examining carefully the references available that led to the assignment of the 18 unexpected species of our dataset (Table S4), most of them were not retained as putative new introductions (see details below).

## **b. The added value of environmental DNA metabarcoding.**

Using eDNA from water samples allows the detection of species over a broader range of taxa and environments than what could be expected from targeted sampling of specimens. Marinas are indeed composed of diverse micro-habitats, such as pillars, seawalls, pontoons, and soft-sediment bottoms, which display contrasting faunal assemblages. For instance, the NIS contribution to fouling communities can be larger on floating compared to fixed habitats (Leclerc et al., 2020), whereas seawalls display communities more similar to those of natural rocky habitats (Megina, González-Duarte, & López-González, 2016). In this context, traditional standardized surveillance protocols (e.g. HELCOM, 2013; Hewitt & Martin, 2001) need to target water, sediment, and various hard substrates to encompass the whole diversity of marinas, but with important time and labor costs. This study showed that eDNA metabarcoding of water samples can provide data that span a broader spectrum of habitats than what could be uncovered during the same time by traditional methods. Considering the 96 species identified with eDNA metabarcoding (Table S4, excluding 18 unexpected species), only 51 were sessile organisms living fixed on hard substrates, 3 might eventually live fixed (occasionally producing a byssus; *Aequipecten opercularis* (Linnaeus, 1758), *Limaria hians* (Gmelin, 1791), and

*Venerupis corrugata* (Gmelin, 1791)), 16 were species living on soft bottoms, and 26 were motile gastropods. When focusing on introduced species, two non-indigenous bivalves, already known from our study area, were recovered by metabarcoding (Table 2). These two species (*Mya arenaria* Linnaeus, 1758 and *Ruditapes philippinarum* (A. Adams & Reeve, 1850)) live buried in soft sediments, and were thus neither targeted nor expected in the quadrat sampling.

One expectation of the use of eDNA metabarcoding is its potential to detect species new to a given area, either introduced or expanding their current distribution range, and that are not actively looked for. Over the four targeted classes, 18 species yet unreported in our study area were recovered. Most were likely misidentified, with sequences that should have been assigned to local species. This could be the case for *Scrupocaberea maderensis* Busk, 1860 which was identified only with 16S with 97.37% identity, when no reference sequence was available for the closely-related native species *Scrupocellaria scruposa* Linnaeus, 1758 identified with COI (see also the above example of *Clavelina meridionalis*). Similarly, eight species were assigned with 18S only, among which six with identity percentage below 100%. Considering the low taxonomic resolution of 18S, and that most of them have native congeners in the study area, they are unlikely to be new introductions. Three species were however assigned with 100% identity with COI, and might thus be true new arrivals or species never detected before in traditional surveys. The tunicate *Polycarpa tenera* Lacaze-Duthiers & Delage, 1892 could in fact be present in the English Channel where it might have been reported as *P. gracilis*, a Mediterranean species (Monniot, 1974; Vazquez, Ramos-Espla, & Turon, 1995). A similar situation could explain the presence of the gastropod *Cerithiopsis petanii*, described from the Croatian coast (Prkić & Mariottini, 2009), and belonging to the *C. tubercularis* species complex (Modica, Mariottini, Prkić, & Oliverio, 2013), justifying why *C. tubercularis* reported in our area might, in fact, be *C. petanii*. The only likely new record would thus be *Haminoea ortea* Talavera, Murillo & Templado, 1987, reported in the Mediterranean Sea and the southern part of the North East Atlantic (Garabedian et al., 2017). These cases demonstrate that eDNA-metabarcoding identification of every unreported species should be carefully checked to avoid false positives (Darling et al., 2020). Metabarcoding should thus be used with extreme caution for early detection of new records, and results need to be confirmed using alternative methods, either traditional and/or more sensitive and targeted molecular approaches such as qPCR or ddPCR (Harper et al., 2018; Wood et al., 2019).

### c. Improving sampling strategy and data processing to overcome limitations

Besides the challenges of taxonomic assignment, sampling effort and processing of raw sequencing data might also contribute to the discrepancies observed between traditional and metabarcoding approaches. Sampling two points (i.e. two pontoons) in each marina increased (sometimes doubled) the number of metazoan taxa recovered (Fig. 5). These findings are in agreement with those of Grey et al. (2018), and suggest that our spatial sampling effort might not be sufficient to detect every sessile species within each marina. Furthermore, the volume of water filtered also influenced the results. Filtering 4L of water per pontoon increased the number of metazoan taxa recovered by 20.7 % across all sites. Increasing the volume filtered mainly improved the detection of rare species (i.e. low number of reads), which were either found in both 2L samples or in only one, whereas all the most abundant species were recovered with only 2L (Fig. 6). Increasing the volume might thus be particularly useful for NIS early detection. For example, in the three marinas where it has been detected in spring with eDNA-metabarcoding, *Asterocarpa humilis* was identified from one pontoon only (two replicates) in BLO and one replicate only (out of four) in CAM and SM. Similarly, sampling only one pontoon or only one replicate per pontoon may have led to the undetection of *Corella eumyota* in SQ (found in all replicates in TR). Conversely, *Botrylloides diegensis* was found in all four replicates in SM, SQ and TRI, and in three replicates in BLO. Our results further underlined the importance of seasonal sampling. Because many sessile invertebrates composing the biofouling communities are short-lived, such as tunicates or erected bryozoans, their abundance can vary greatly depending on the season. For example, the tunicate *Ciona robusta* would not have been identified if the samples had only been collected in spring (Fig. 3). Considering the use of eDNA metabarcoding as a surveillance tool for NIS, sampling efforts should be maximized, with as many sampling points in space and time as possible according to time and cost constraints.

Despite the substantial number of species observed in quadrats that were recovered by eDNA metabarcoding over the whole study, our detection capacity was much lower locally, i.e. when considering presence/absence data per site and date (a "species x site x date" combination will further be referred to as "unit", which corresponds to one cell of the panels "Fall" and "Spring" in Fig. 3). Before any correction (index-jump and replicate filtering) both quadrat observations and eDNA metabarcoding detected the presence of a given taxon in 46.2 % of all "presence" units. This number fell down to 23.4 % after corrections, with 127 units of eDNA

metabarcoding detection disappearing. Considering NIS or cryptogenic species, this proportion decreased from 61.4 % to 33.8 % (Figs. S4-S5). With three exceptions (*Musculus* sp., *Bugulina stolonifera*, and *B. simplex*), all taxa identified before correction were kept after correction in the whole dataset (columns TB and TA; Figs. S4-S5). Most changes (61.4 %) were lost after correction for index-jump. Index- and tag-jump is a technical problem that needs to be addressed in order to avoid the false detection of species in some samples (Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012). We discarded reads from a particular ASV in a given sample if the proportion relative to the total number of reads from this ASV was lower than a given threshold. The thresholds were chosen for each marker according to the maximum proportion observed in index-jump controls, so that using lower thresholds would have left index-jump errors in our dataset. This approach is more sensible than simply selecting a fixed number of reads arbitrarily as a threshold. Without index-jump correction, many species would have been identified from ASVs with read abundances lower than what would result from index-jump, making it impossible to determine if they were “true” detections. PCR replicate filtering also resulted in a loss of taxa (corresponding to 49 units, Figs. S4-S5). Although the use of PCR replicates has proven useful to limit false positives due to PCR errors (Alberdi, Aizpurua, Gilbert, & Bohmann, 2018; Tsuji et al., 2020), choosing the right threshold is a matter of compromise. We used the lowest possible value (i.e. an ASV is kept if present in at least 2 PCR replicates) in order to increase the chances of detecting rare species, while maintaining some level of stringency. Despite this low threshold, this filtering step might have increased the rate of false negatives. Finally, another aspect of the methodology that could be improved is the sequencing depth. The vast majority of reads produced with either 18S (77%) or COI (76%) were not assigned to metazoans (Fig. 2), a common issue when using universal primers (e.g. Andújar et al., 2018 for COI). None of the species detected within the four targeted classes represented more than 1% of read abundances, with the notable exception of *Ciona intestinalis* with 18S (2.9%). The use of universal primers for NIS detection is problematic in this regard, as was also demonstrated by Cowart et al. (2015) or Collins et al. (2019), and more specific primers might be preferred. Since it can be challenging to design primers targeting several taxonomic groups without amplification biases, multiplexing several primer sets would minimize unwanted amplifications while targeting a wide range of taxa (Corse et al., 2019; Zhang et al., 2018).

#### d. Some issues are still unresolved

One of the main remaining issue when considering eDNA metabarcoding as a biosurveillance tool is the quantitative aspect of the method. Several studies have addressed this question for metabarcoding on various types of samples (Elbrecht & Leese, 2015; Lamb et al., 2019; Sun et al., 2015) and did not find any clear relationship between metabarcoding results and individual counts or biomass. This can be explained by many factors such as the differential primers binding efficiency between taxa (Piñol et al., 2018), the differing levels of eDNA release depending on the species, individual age and biomass, or various other factors (Barnes & Turner, 2016; Maruyama, Nakamura, Yamanaka, Kondoh, & Minamoto, 2014; Sassoubre, Yamahara, Gardner, Block, & Boehm, 2016). In our study, the comparisons between read proportions and individual counts in quadrats for six common ascidian taxa showed a positive correlation for only two taxa, *Ascidiella* sp. and *Ciona robusta* (Fig. 4). Our data, however, were mainly influenced by our low rate of detection with eDNA metabarcoding as described above, with many points having a null read proportion although having been observed in various abundances within quadrats. Furthermore, quantitative data collected within quadrats have their own bias and do not necessarily depict accurately the species abundances in the whole marina. Further testing is direly needed to evaluate the quantitative power of eDNA metabarcoding in the context of marine species monitoring and this method should, for now, either be limited to presence/absence or semi-quantitative data, or be coupled with more sensitive approaches such as qPCR or ddPCR assays (Wood et al., 2019).

As a conclusion, our results suggest that eDNA metabarcoding can be of great value for marine NIS detection in marinas and should be included in biosurveillance programs. Some major issues, however, can lead to both false negatives and false positives with major implications for NIS monitoring (Sepulveda, Nelson, Jerde, & Luikart, 2020) which cannot allow its exclusive use in this context. False negatives prevent the detection of early novel introductions or from monitoring correctly the spread of reported NIS, whereas false positives may lead to an overestimation of introduced species and in turn of their distribution range. Most of these issues can be overcome with more efforts directed towards the generation of reliable curated databases of reference sequences, including databases specific to NIS (Darling et al., 2017). Some local attempts have been made in that direction (e.g. Dias et al., 2017) but a more worldwide approach would be beneficial. Improvements in the



metabarcoding process itself are further needed (e.g. additional markers or new sets of primers). Before reaching optimal conditions for taxonomic assignment, caution should be taken, and in-depth examination undertaken, in case of detection of previously unreported (non-indigenous) species. The quantitative aspect of the approach, however, still cannot be resolved and needs further testing to evaluate its impact on species surveillance. Increasing the sampling effort might help improve the detection of rare taxa. Furthermore, eDNA from water samples, despite being easier to collect and process, has been shown to not always be the substrate of choice for certain taxonomic groups (Koziol et al., 2019; Rey et al., 2020). The combination of various types of samples might thus be of great value to increase NIS detection capacity.

## Acknowledgments

We are thankful to the Diving and Marine core service from the Roscoff Biological Station, and particularly Yann Fontana, Wilfried Thomas and Mathieu Camusat for the quadrat sampling, as well as Jérôme Coudret and Stéphane Loisel for their help for the field work. We thank Gwenn Tanguy from the Biogenouest Genomer Platform for advice and access to the sequencing facilities, and the Biogenouest ABIMS Platform for access to the calculation resources. This project was supported by the TOTAL Foundation (project AquaNIS2.0). MC acknowledges a PhD grant by Région Bretagne (ENIGME ARED project) and Sorbonne Université (ED 227 “Sciences de la Nature et de l’Homme”).

## Authors' contributions

MC, LL, TC, and FV designed the study and/or the analyses; TC, LL, and FV supervised field work; MC, and CDT collected the molecular data; FV and LL supervised morphological identifications in the field and LL carry out further analyses in the lab; MC analysed the data; MC, TC, and FV led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

## References

- Airoidi, L., Turon, X., Perkol-Finkel, S., & Rius, M. (2015). Corridors for aliens but not for natives: effects of marine urban sprawl at a regional scale. *Diversity and Distributions*, *21*(7), 755-768. doi:doi:10.1111/ddi.12301
- Albano, M. J., & Obenat, S. M. (2019). Fouling assemblages of native, non-indigenous and cryptogenic species on artificial structures, depths and temporal variation. *Journal of Sea Research*, *144*, 1-15. doi:https://doi.org/10.1016/j.seares.2018.10.002
- Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*(1), 134-147. doi:10.1111/2041-210x.12849
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410. doi:10.1016/S0022-2836(05)80360-2
- Andújar, C., Arribas, P., Yu, D. W., Vogler, A. P., & Emerson, B. C. (2018). Why the COI barcode should be the community DNA metabarcode for the metazoa. *Molecular Ecology*, *27*(20), 3968-3975. doi:10.1111/mec.14844
- Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, *17*(1), 1-17. doi:10.1007/s10592-015-0775-4
- Bishop, J. D. D., Wood, C. A., Lévêque, L., Yunnice, A. L. E., & Viard, F. (2015). Repeated rapid assessment surveys reveal contrasting trends in occupancy of marinas by non-indigenous species on opposite sides of the western English Channel. *Marine Pollution Bulletin*, *95*(2), 699-706. doi:10.1016/j.marpolbul.2014.11.043
- Borrell, Y. J., Miralles, L., Do Huu, H., Mohammed-Geba, K., & Garcia-Vazquez, E. (2017). DNA in a bottle—Rapid metabarcoding survey for early alerts of invasive species in ports. *PLOS ONE*, *12*(9), e0183347. doi:10.1371/journal.pone.0183347
- 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. doi:10.1111/1755-0998.12428
- Briski, E., Ghabooli, S., Bailey, S. A., & MacIsaac, H. J. (2016). Are genetic databases sufficiently populated to detect non-indigenous species? *Biological Invasions*, *18*(7), 1911-1922. doi:10.1007/s10530-016-1134-1
- Brown, E. A., Chain, F. J. J., Zhan, A., MacIsaac, H. J., & Cristescu, M. E. (2016). Early detection of aquatic invaders using metabarcoding reveals a high number of non-indigenous species in Canadian ports. *Diversity and Distributions*, *22*(10), 1045-1059. doi:10.1111/ddi.12465
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*, *13*, 581. doi:10.1038/nmeth.3869
- Clarke, L. J., Beard, J. M., Swadling, K. M., & Deagle, B. E. (2017). Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. *Ecology and Evolution*, *7*(3), 873-883. doi:10.1002/ece3.2667
- Clarke Murray, C., Pakhomov, E. A., & Therriault, T. W. (2011). Recreational boating: a large unregulated vector transporting marine invasive species. *Diversity and Distributions*, *17*(6), 1161-1172. doi:doi:10.1111/j.1472-4642.2011.00798.x
- Collins, R. A., Bakker, J., Wangensteen, O. S., Soto, A. Z., Corrigan, L., Sims, D. W., . . . Mariani, S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution*, *10*(11), 1985-2001. doi:10.1111/2041-210x.13276
- Comtet, T., Sandionigi, A., Viard, F., & Casiraghi, M. (2015). DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions*, *17*(3), 905-922. doi:10.1007/s10530-015-0854-y
- Corse, E., Tougard, C., Archambaud-Suard, G., Agnès, J.-F., Messu Mandeng, F. D., Bilong Bilong, C. F., . . . Dubut, V. (2019). One-locus-several-primers: A strategy to improve the taxonomic and haplotypic coverage in diet metabarcoding studies. *Ecology and Evolution*, *9*(8), 4603-4620. doi:10.1002/ece3.5063
- Couton, M., Comtet, T., Le Cam, S., Corre, E., & Viard, F. (2019). Metabarcoding on planktonic larval stages: an efficient approach for detecting and investigating life cycle dynamics of benthic aliens. *Management of Biological Invasions*, *10*(4), 657-689. doi:10.3391/mbi.2019.10.4.06

- Couton, M., Lévêque, L., Daguin-Thiébaud, C., Comtet, T., & Viard, F. (submitted). High-Throughput Sequencing on preservative ethanol is effective at jointly examining infra-specific and taxonomic diversity, although bioinformatics pipelines do not perform equally. *Methods in Ecology and Evolution*.
- Cowart, D. A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., & Arnaud-Haond, S. (2015). Metabarcoding Is Powerful yet Still Blind: A Comparative Analysis of Morphological and Molecular Surveys of Seagrass Communities. *PLOS ONE*, *10*(2), e0117562. doi:10.1371/journal.pone.0117562
- Darling, J. A., Galil, B. S., Carvalho, G. R., Rius, M., Viard, F., & Piraino, S. (2017). Recommendations for developing and applying genetic tools to assess and manage biological invasions in marine ecosystems. *Marine Policy*, *85*, 54-64. doi:10.1016/j.marpol.2017.08.014
- Darling, J. A., Pochon, X., Abbott, C. L., Inglis, G. J., & Zaiko, A. (2020). The risks of using molecular biodiversity data for incidental detection of species of concern. *Diversity and Distributions*, *26*(9), 1116-1121. doi:10.1111/ddi.13108
- Deiner, K., Lopez, J., Bourne, S., Holman, L., Seymour, M., Grey, E. K., . . . Pfrender, M. E. (2018). Optimising the detection of marine taxonomic richness using environmental DNA metabarcoding: the effects of filter material, pore size and extraction method. *Metabarcoding and Metagenomics*, *2*, e28963.
- Dias, J., Fotedar, S., Munoz, J., Hewitt, M., Lukehurst, S., Hourston, M., . . . snow, M. (2017). Establishment of a taxonomic and molecular reference collection to support the identification of species regulated by the Western Australian Prevention List for Introduced Marine Pests. *Management of Biological Invasions*, *8*(2), 215-225. doi:10.3391/mbi.2017.8.2.09
- Duarte, S., Vieira, P., & Costa, F. (2020). Assessment of species gaps in DNA barcode libraries of non-indigenous species (NIS) occurring in European coastal regions. *Metabarcoding and Metagenomics*, *4*. doi:10.3897/mbmg.4.55162
- Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PLOS ONE*, *10*(7), e0130324. doi:10.1371/journal.pone.0130324
- Fehlauer-Ale, K. H., Mackie, J. A., Lim-Fong, G. E., Ale, E., Pie, M. R., & Waeschenbach, A. (2014). Cryptic species in the cosmopolitan *Bugula neritina* complex (Bryozoa, Cheilostomata). *Zoologica Scripta*, *43*(2), 193-205. doi:10.1111/zsc.12042
- Ferrario, J., Caronni, S., Occhipinti-Ambrogi, A., & Marchini, A. (2017). Role of commercial harbours and recreational marinas in the spread of non-indigenous fouling species. *Biofouling*, *33*(8), 651-660. doi:10.1080/08927014.2017.1351958
- Fonseca, V., Carvalho, G., Sung, W., Johnson, H. F., Power, D., Neill, S., . . . Creer, S. (2010). Second-generation environmental sequencing unmasking marine metazoan biodiversity. *Nature Communications*, *1*, 98.
- Garabedian, K., Malaquias, M., Crocetta, F., Zenetos, A., Kavadas, S., & Valdés, Á. (2017). *Haminoea orteai* Talavera, Murillo and Templado, 1987 (Mollusca: Gastropoda: Heterobranchia: Cephalaspidea), a widespread species in the Mediterranean and northeastern Atlantic. *Cahiers de Biologie Marine*, *58*, 107-113.
- Giakoumi, S., Katsanevakis, S., Albano, P. G., Azzurro, E., Cardoso, A. C., Cebrian, E., . . . Sghaier, Y. R. (2019). Management priorities for marine invasive species. *Science of The Total Environment*, *688*, 976-982. doi:10.1016/j.scitotenv.2019.06.282
- Glasby, T. M., Connell, S. D., Holloway, M. G., & Hewitt, C. L. (2007). Nonindigenous biota on artificial structures: could habitat creation facilitate biological invasions? *Marine Biology*, *151*(3), 887-895. doi:10.1007/s00227-006-0552-5
- 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. doi:10.1111/2041-210x.12595
- Grey, E. K., Bernatchez, L., Cassey, P., Deiner, K., Deveney, M., Howland, K. L., . . . Lodge, D. M. (2018). Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Sci Rep*, *8*(1), 8843. doi:10.1038/s41598-018-27048-2
- Harper, L. R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H. C., Gough, K. C., . . . Haenfling, B. (2018). Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*). *bioRxiv*. doi:10.1101/215897
- Harris, D. J. (2003). Can you bank on GenBank? *Trends in Ecology & Evolution*, *18*(7), 317-319. doi:http://dx.doi.org/10.1016/S0169-5347(03)00150-2

- HELCOM, O. (2013). Joint Harmonised Procedure for the Contracting Parties of OSPAR and HELCOM on the Granting of Exemptions Under International Convention for the Control and Management of Ships' Ballast Water and Sediments. *Regulation A-4*.
- Hewitt, C. L., & Martin, R. B. (2001). Revised protocols for baseline port surveys for introduced marine species: survey design sampling protocols and specimen handling (Technical report no. 22). Centre for research on introduced marine pests Hobart, Australia
- Katsanevakis, S., Wallentinus, I., Zenetos, A., Leppäkoski, E., Çinar, M. E., Oztürk, B., . . . Cardoso, A. C. (2014). Impacts of invasive alien marine species on ecosystem services and biodiversity: a pan-European review. *Aquatic Invasions*, 9(4), 391-423.
- Kelly, R. P., O'Donnell, J. L., Lowell, N. C., Shelton, A. O., Samhour, J. F., Hennessey, S. M., . . . Williams, G. D. (2016). Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ*, 4, e2444. doi:10.7717/peerj.2444
- Kijewski, T. K., Zbawicka, M., Väinölä, R., & Wenne, R. (2006). Introgression and mitochondrial DNA heteroplasmy in the Baltic populations of mussels *Mytilus trossulus* and *M. edulis*. *Marine Biology*, 149(6), 1371-1385. doi:10.1007/s00227-006-0316-2
- Knapp, I. S., Forsman, Z. H., Williams, G. J., Toonen, R. J., & Bell, J. J. (2015). Cryptic species obscure introduction pathway of the blue Caribbean sponge (*Haliclona* (*Soestella*) *caerulea*), (order: Haplosclerida) to Palmyra Atoll, Central Pacific. *PeerJ*, 3, e1170-e1170. doi:10.7717/peerj.1170
- Kozioł, A., Stat, M., Simpson, T., Jarman, S., DiBattista, J. D., Harvey, E. S., . . . Bunce, M. (2019). Environmental DNA metabarcoding studies are critically affected by substrate selection. *Molecular Ecology Resources*, 19(2), 366-376. doi:10.1111/1755-0998.12971
- Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G., & Taylor, M. I. (2019). How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology*, 28(2), 420-430. doi:10.1111/mec.14920
- Leclerc, J.-C., Viard, F., & Brante, A. (2020). Experimental and survey-based evidences for effective biotic resistance by predators in ports. *Biological Invasions*, 22(2), 339-352. doi:10.1007/s10530-019-02092-9
- Lehtiniemi, M., Ojaveer, H., David, M., Galil, B., Gollasch, S., McKenzie, C., . . . Pederson, J. (2015). Dose of truth—Monitoring marine non-indigenous species to serve legislative requirements. *Marine Policy*, 54, 26-35. doi:https://doi.org/10.1016/j.marpol.2014.12.015
- Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., . . . Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, 10(1), 34. doi:10.1186/1742-9994-10-34
- Makiola, A., Compson, Z. G., Baird, D. J., Barnes, M. A., Boerlijst, S. P., Bouchez, A., . . . Bohan, D. A. (2020). Key questions for next-generation biomonitoring. *Frontiers in Environmental Science*, 7(197). doi:10.3389/fenvs.2019.00197
- Martel, A. L., Auffrey, L. M., Robles, C. D., & Honda, B. M. (2000). Identification of settling and early postlarval stages of mussels (*Mytilus* spp.) from the Pacific coast of North America, using prodissoconch morphology and genomic DNA. *Marine Biology*, 137(5), 811-818. doi:10.1007/s002270000442
- 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, e114639. doi:10.1371/journal.pone.0114639
- Megina, C., González-Duarte, M. M., & López-González, P. J. (2016). Benthic assemblages, biodiversity and invasiveness in marinas and commercial harbours: an investigation using a bioindicator group. *Biofouling*, 32(4), 465-475. doi:10.1080/08927014.2016.1151500
- Meistertzheim, A.-L., Héritier, L., & Lejart, M. (2017). High-Resolution Melting of 18S rDNA sequences (18S-HRM) for discrimination of bivalve's species at early juvenile stage: application to a spat survey. *Marine Biology*, 164(6), 133. doi:10.1007/s00227-017-3162-5
- Modica, M. V., Mariottini, P., Prkić, J., & Oliverio, M. (2013). DNA-barcoding of sympatric species of ectoparasitic gastropods of the genus *Cerithiopsis* (Mollusca: Gastropoda: Cerithiopsidae) from Croatia. *Journal of the Marine Biological Association of the United Kingdom*, 93(4), 1059-1065. doi:10.1017/S0025315412000926
- Molnar, J. L., Gamboa, R. L., Revenga, C., & Spalding, M. D. (2008). Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment*, 6(9), 485-492. doi:10.1890/070064
- Monniot, C. (1974). Ascidiés littorales et bathyales récoltées au cours de la campagne Biaçores: Phlébobranches et Stolidobranches. *Bulletin du Museum National d'Histoire Naturelle 3ème série*, 251(173), 1327-1352.

- Morais, P., & Reichard, M. (2018). Cryptic invasions: A review. *Science of The Total Environment*, 613-614, 1438-1448. doi:<https://doi.org/10.1016/j.scitotenv.2017.06.133>
- Ojaveer, H., Galil, B. S., Campbell, M. L., Carlton, J. T., Canning-Clode, J., Cook, E. J., . . . Ruiz, G. (2015). Classification of non-indigenous species based on their impacts: considerations for application in marine management. *PLOS Biology*, 13(4), e1002130. doi:10.1371/journal.pbio.1002130
- Ojaveer, H., Galil, B. S., Carlton, J. T., Alleway, H., Gouletquer, P., Lehtiniemi, M., . . . Zaiko, A. (2018). Historical baselines in marine bioinvasions: Implications for policy and management. *PLOS ONE*, 13(8), e0202383. doi:10.1371/journal.pone.0202383
- Pérez-Portela, R., & Turon, X. (2008). Phylogenetic relationships of the Clavelinidae and Pycnoclavellidae (Asciacea) inferred from mtDNA data. *Invertebrate Biology*, 127(1), 108-120. doi:10.1111/j.1744-7410.2007.00112.x
- Peters, K., Sink, K. J., & Robinson, T. B. (2019). Sampling methods and approaches to inform standardized detection of marine alien fouling species on recreational vessels. *Journal of Environmental Management*, 230, 159-167. doi:<https://doi.org/10.1016/j.jenvman.2018.09.063>
- Piñol, J., Senar, M. A., & Symondson, W. O. C. (2018). The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology*, 1-13. doi:10.1111/mec.14776
- Pochon, X., Bott, N. J., Smith, K. F., & Wood, S. A. (2013). Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. *PLOS ONE*, 8(9), e73935. doi:10.1371/journal.pone.0073935
- Porter, T. M., & Hajibabaei, M. (2018). Over 2.5 million COI sequences in GenBank and growing. *PLOS ONE*, 13(9), e0200177. doi:10.1371/journal.pone.0200177
- Prkić, J., & Mariottini, P. (2009). Description of two new Cerithiopsis from the Croatian coast, with comments on the Cerithiopsis tubercularis complex (Gastropoda, Cerithiopsidae). 5, 3-27.
- Rey, A., Basurko, O. C., & Rodriguez-Ezpeleta, N. (2020). Considerations for metabarcoding-based port biological baseline surveys aimed at marine nonindigenous species monitoring and risk assessments. *Ecology and Evolution*, 10(5), 2452-2465. doi:10.1002/ece3.6071
- Rogers, T. L., Byrnes, J. E., & Stachowicz, J. J. (2016). Native predators limit invasion of benthic invertebrate communities in Bodega Harbor, California, USA. *Marine Ecology Progress Series*, 545, 161-173.
- 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. doi:10.1021/acs.est.6b03114
- Seebens, H., Blackburn, T. M., Dyer, E. E., Genovesi, P., Hulme, P. E., Jeschke, J. M., . . . Essl, F. (2017). No saturation in the accumulation of alien species worldwide. *Nature Communications*, 8, 14435. doi:10.1038/ncomms14435  
<https://www.nature.com/articles/ncomms14435#supplementary-information>
- Sepulveda, A. J., Nelson, N. M., Jerde, C. L., & Luikart, G. (2020). Are Environmental DNA Methods Ready for Aquatic Invasive Species Management? *Trends in Ecology & Evolution*, 35(8), 668-678. doi:<https://doi.org/10.1016/j.tree.2020.03.011>
- Shaw, J. L. A., Weyrich, L. S., Hallegraeff, G., & Cooper, A. (2019). Retrospective eDNA assessment of potentially harmful algae in historical ship ballast tank and marine port sediments. *Molecular Ecology*, 28(10), 2476-2485. doi:10.1111/mec.15055
- Simberloff, D., Martin, J.-L., Genovesi, P., Maris, V., Wardle, D. A., Aronson, J., . . . Vilà, M. (2013). Impacts of biological invasions: what's what and the way forward. *Trends in Ecology & Evolution*, 28(1), 58-66. doi:10.1016/j.tree.2012.07.013
- Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., . . . Bunce, M. (2017). Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Sci Rep*, 7(1), 12240. doi:10.1038/s41598-017-12501-5
- Sun, C., Zhao, Y., Li, H., Dong, Y., MacIsaac, H. J., & Zhan, A. (2015). Unreliable quantitation of species abundance based on high-throughput sequencing data of zooplankton communities. *Aquatic Biology*, 24(1), 9-15.
- Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For Biodiversity Research and Monitoring*. Oxford: Oxford University Press.
- Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, 21(8), 1789-1793. doi:10.1111/j.1365-294X.2012.05542.x

- Tsuji, S., Miya, M., Ushio, M., Sato, H., Minamoto, T., & Yamanaka, H. (2020). Evaluating intraspecific genetic diversity using environmental DNA and denoising approach: A case study using tank water. *Environmental DNA*, 2(1), 42-52. doi:10.1002/edn3.44
- Ulman, A., Ferrario, J., Forcada, A., Seebens, H., Arvanitidis, C., Occhipinti-Ambrogi, A., & Marchini, A. (2019). Alien species spreading via biofouling on recreational vessels in the Mediterranean Sea. *Journal of Applied Ecology*, 56(12), 2620-2629. doi:10.1111/1365-2664.13502
- Vazquez, E., Ramos-Espla, A. A., & Turon, X. (1995). The genus *Polycarpa* (Asciacea, Styelidae) on the Atlantic and Mediterranean coasts of the Iberian Peninsula. *Journal of Zoology*, 237(4), 593-614. doi:10.1111/j.1469-7998.1995.tb05017.x
- Viard, F., David, P., & Darling, J. A. (2016). Marine invasions enter the genomic era: three lessons from the past, and the way forward. *Current zoology*, 62(6), 629-642. doi:10.1093/cz/zow053
- Viard, F., Roby, C., Turon, X., Bouchemousse, S., & Bishop, J. D. D. (2019). Cryptic diversity and database errors challenge non-indigenous species surveys: an illustration with *Botrylloides* spp. in the English Channel and Mediterranean Sea. *Frontiers in Marine Science*, 6, 615. doi:10.3389/fmars.2019.00615
- von Ammon, U., Wood, S. A., Laroche, O., Zaiko, A., Tait, L., Lavery, S., . . . Pochon, X. (2018). Combining morpho-taxonomy and metabarcoding enhances the detection of non-indigenous marine pests in biofouling communities. *Sci Rep*, 8(1), 16290. doi:10.1038/s41598-018-34541-1
- Wood, S. A., Pochon, X., Laroche, O., von Ammon, U., Adamson, J., & Zaiko, A. (2019). A comparison of droplet digital polymerase chain reaction (PCR), quantitative PCR and metabarcoding for species-specific detection in environmental DNA. *Molecular Ecology Resources*, 0(0), 1-13. doi:10.1111/1755-0998.13055
- Zaiko, A., Samuiloviene, A., Ardura, A., & Garcia-Vazquez, E. (2015). Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. *Marine Pollution Bulletin*, 100(1), 53-59. doi:https://doi.org/10.1016/j.marpolbul.2015.09.030
- Zbawicka, M., Burzyński, A., Skibinski, D., & Wenne, R. (2010). Scottish *Mytilus trossulus* mussels retain ancestral mitochondrial DNA: Complete sequences of male and female mtDNA genomes. *Gene*, 456(1), 45-53. doi:https://doi.org/10.1016/j.gene.2010.02.009
- Zhang, G. K., Chain, F. J. J., Abbott, C. L., & Cristescu, M. E. (2018). Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. *Evolutionary Applications*, 11(10), 1901-1914. doi:doi:10.1111/eva.12694

# Marina communities are not homogenous at a regional scale nor over time, as jointly shown by eDNA metabarcoding and quadrat analyses

**Running headline:** eDNA metabarcoding for marine biodiversity surveys

Marjorie Couton<sup>1</sup>, Laurent Lévêque<sup>2</sup>, Claire Daguin-Thiébaud<sup>1</sup>, Thierry Comtet<sup>1</sup>,  
Frédérique Viard<sup>1\*</sup>

<sup>1</sup> Sorbonne université, CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

<sup>2</sup> Sorbonne université, CNRS, FR 2424, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

\* Correspondence author: viard@sb-roscoff.fr; +33 2 98 29 23 12

**Keywords:** biodiversity, biological invasions, biomonitoring, environmental DNA, metabarcoding, morphology, non-indigenous species, quadrats

**This part is the preliminary version of a future paper**

## Abstract

With the growing hardening of marine coasts, marinas are forming dense networks connected by recreational boating. This is now recognized as a major vector of non-indigenous species (NIS) expansion. In addition, because they are not surrogates of nearby hard-bottom natural habitats, and display particular abiotic features, marinas are likely to select for particular communities. The joint effects of human-mediated dispersal and environmental filtering should enforce the homogenization of biological communities (i.e. biotic homogenization), particularly for NIS. To test this expectation, we examined 10 marinas spread over two biogeographical provinces and a transition zone, using two methods: i) one approach with *in situ* morphological identification (quadrat survey), targeting biofouling communities found under floating pontoons, and ii) metabarcoding applied to eDNA water samples, which allowed a global biodiversity assessment. Alpha- and beta-diversity analyses showed similar results whatever the method, and, for metabarcoding, whatever the functional group considered (i.e. Chromista, mobile metazoans, and benthic metazoans). All datasets showed a significant temporal (fall vs. spring) differentiation of the communities, suggesting a high-turnover. In addition, all datasets revealed an unexpected significant spatial structure, explained by both marinas and regional effects, suggesting that natural processes are not fully offset by the properties of marinas and human-mediated dispersal, with no sign of biotic homogenization at the study scale. Focusing on NIS and cryptogenic species from biofouling communities, 14 and 32 species were detected in quadrat surveys and metabarcoding, respectively. NIS and cryptogenic species were found in every location, sometimes with high proportions (up to 50% and 47% of the species in quadrat and metabarcoding datasets, respectively). The spatial (regional) differentiation in hard-bottom communities was lower with NIS and cryptogenic species only, than with native species only. This suggested a greater influence of human-mediated dispersal for NIS. Altogether, this study showed that eDNA metabarcoding can be effective to monitor the diversity and structure of marina communities, and that biotic homogenization did not operate at the scale of our study (more than 400 km of coast line).



## Introduction

The artificialization of coastal areas is constantly growing, contributing to the phenomenon of 'ocean sprawl' as first coined by Duarte, et al. (2013). One type of artificial structure taking part in this phenomenon are marinas, which have seen their numbers increasing worldwide with the rising interest in recreational boating (Airoldi and Beck 2007). Marinas are highly anthropogenic environments with physical and chemical conditions distinct from surrounding natural habitats such as increased water retention, high turbidity, high levels of contaminants, or extensive shading due to moored boats and pontoons (Glasby 1999; Rivero, et al. 2013). These particular features influence the species assemblages that are found in marinas, especially (but not exclusively) on hard substrates, which are not surrogates for natural communities, (Bulleri and Chapman 2004; Bulleri and Chapman 2010). As such, they are expected to select some particular species and communities in a way comparable to what has been shown in human-altered terrestrial habitats, such as cultivated fields (Hufbauer, et al. 2012). Very diverse sessile organisms are found on artificial hard substrates such as pilings, pontoons or boat hulls (Bulleri and Chapman 2004; Ruiz, et al. 2009). The natural dispersal ability of many of these fouling species, including bryozoans, tunicates or sponges, is typically low because either they do not release free-swimming larvae (direct developers) or their larval stages only spend a few hours in the water column (Shanks, et al. 2003; Shanks 2009). Consequently, their dispersal at a regional scale will be tightly linked to human-mediated transport *via* hull fouling, and their distribution should not be related to natural barriers, such as currents. Even for species that release free-swimming larvae, and even more for those which brood their progeny (like calyptraeid gastropods or barnacles), hull-mediated transport may further increase their natural dispersal abilities. Moreover, boats travelling back and forth between marinas within a region might tend to homogenize marina communities by transporting sessile organisms from one place to another. The joint effects of human-mediated dispersal and environmental filtering are expected to enforce biotic homogenization for communities established in marinas (i.e. the homogenization of biological communities at functional, taxonomic and genetic levels; Olden 2006) well above the scale usually predicted from natural processes.

These characteristics are particularly noteworthy when considering biological introductions. The number of introduced species worldwide is constantly growing (Seebens, et al. 2017; Sardain, et al. 2019), with shipping being one of the main vectors for marine species displacement (Molnar, et al. 2008; Nuñez, et al. 2014). Organisms can either be transported inside ships *via* ballast water or bilge water

(Briski, et al. 2012; Fletcher, et al. 2017), or travel attached to the hull as part of fouling communities (Sylvester, et al. 2011). Marinas and ports thus serve as points of entry for introduced species, and are expected to sustain a high colonization and propagule pressure (Sylvester, et al. 2011). Consequently, marinas display a higher number of NIS than natural settings (Ruiz, et al. 2009; Soares, et al. 2018; O'Shaughnessy, et al. 2020). They also offer a wide range of artificial substrates that may facilitate the settlement of non-indigenous species (NIS). They can also serve as stepping stones for secondary introductions (Glasby, et al. 2007; Ruiz, et al. 2009; Bulleri and Chapman 2010).

Initially, more attention has been directed towards commercial ports because international shipping was expected to be the main vector for new species introductions. However, marinas have been shown to be also problematic (Clarke Murray, et al. 2011; Lacoursière-Roussel, et al. 2012). Leisure boats are less frequently cleaned, can accumulate more biofouling through extended mooring periods, and their lower speed can facilitate the settlement of new species (Mineur, et al. 2012). Moreover, they can transport newly established species to nearby marinas or surrounding natural environments, so that they are commonly viewed as vectors for secondary spread (Clarke Murray, et al. 2011). Dafforn, et al. (2009a) found no difference in the number of NIS between ports and marinas, further emphasizing the equal importance of these two types of infrastructures in marine introduction processes. Leclerc, et al. (2018) however found a larger number of NIS in international ports as compared to local ports in Chile, but they did not find any differences in NIS contribution to the structure of the biofouling communities between the two port categories.

As potential entry points for new introductions, marinas and ports are frequently studied to detect the arrival of new species and to monitor NIS already present. Traditional methods usually involve the use of settlement panels or the collection of individuals on site for subsequent morphological identification (e.g. Wasson, et al. 2001; Dafforn, et al. 2009a). These techniques being especially time consuming, some other studies have relied on rapid assessment surveys based on visual inspections of organisms fixed to artificial structures (Bishop, et al. 2015a; Bishop, et al. 2015b). All of these methods, however, depend on the taxonomic expertise of specialists, and on the presence of reliable diagnostic morphological criteria, which are not always available or difficult to use in the field (Turon, et al. 2020). In the last decade, metabarcoding approaches have been used on various sample types to monitor NIS in diverse habitats (Brown, et al. 2016; Zaiko, et al. 2016;

Borrell, et al. 2017; Borrell, et al. 2018; von Ammon, et al. 2018; Couton, et al. 2019). Beyond the interest for monitoring NIS, metabarcoding offers the possibility to provide a more global assessment of biodiversity (Taberlet, et al. 2012; Ji, et al. 2013; Cristescu 2014; Valentini, et al. 2016), allowing to better understand how NIS contribute to the local diversity and how they influence the structure of communities at a regional scale. Combined with the use of environmental DNA (eDNA) (e.g. Borrell et al. 2017), and despite a number of limitations (see [chapter II.1](#)), metabarcoding thus offers the appealing promise of providing a global picture of the communities established in marinas, across space and time, with limited sampling time and cost.

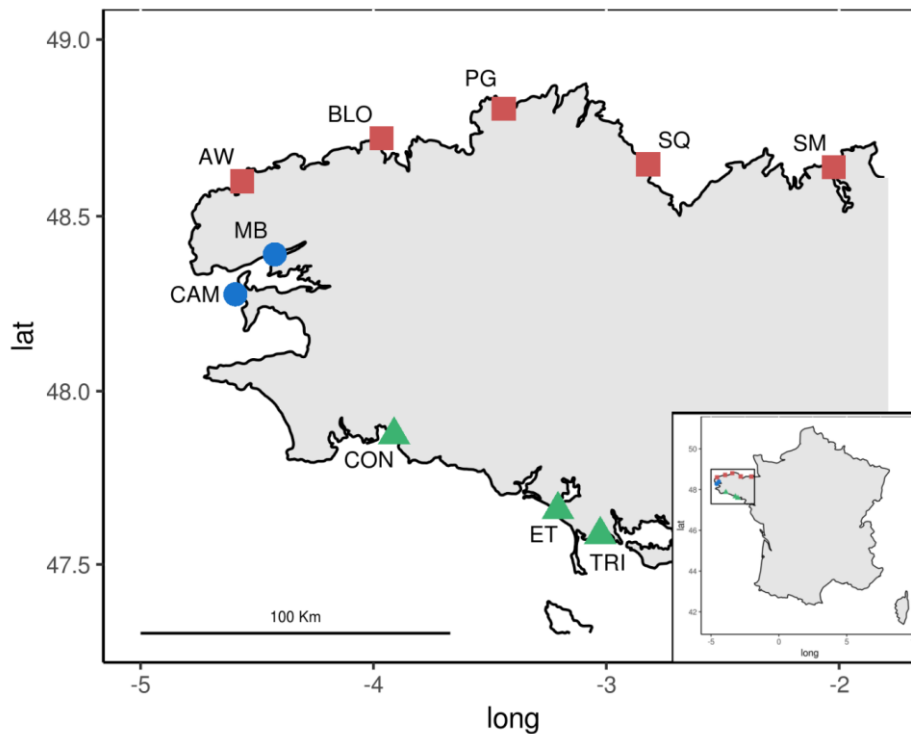
In marinas, where natural dispersal may be replaced by human-mediated dispersal for numerous sessile species, it is expected that dissimilarities between communities of different locations would be low, and not related to the geographical distance or natural biogeographic barriers. Moreover, the high proportion of short-lived organisms in fouling communities coupled with the frequent disturbances that they experience (frequent cleaning, flooding, and pollution) let expect a strong temporal variation in species assemblages. In order to test these hypotheses, and to evaluate the role of NIS in the observed patterns, we used datasets previously obtained in ten French marinas, with a morphology-based approach (quadrat survey) and a molecular-based approach (eDNA-metabarcoding) (see [chapter II.1](#)). The marinas were chosen to be distributed across two biogeographic regions and a transition zone, which have been delimited according to their relatively homogeneous species composition, clearly distinct from adjacent systems (Spalding, et al. 2007). The morphology-based dataset focuses solely on biofouling communities found under floating pontoons whereas the eDNA datasets will allow us to have a more integrative approach by targeting different taxonomic and functional groups.

## Materials and methods

### a. Quadrat and eDNA datasets

We processed and analysed four datasets, one 'quadrat' dataset and three eDNA datasets previously obtained as described in [Chapter II.1](#). Briefly, samples of biofouling communities (quadrat) and seawater (eDNA) were collected along two floating pontoons (10 sampling unit every 5 meters at each pontoon) in October 2017 (Fall) and May-June 2018 (Spring). As opposed to the analysis in [chapter II.1](#),

only one of the two 2-L eDNA replicate sampled in Spring was used for each pontoon. The 10 targeted marinas are distributed around Brittany (France; Fig. 1) and belong to two biogeographic regions (English Channel and North East Atlantic) and a transition zone (the Iroise Sea).



**Figure 1** Map of the ten collection sites. Red squares indicate marinas located in the Western English Channel (Boreal province), blue circles indicate marinas on the Iroise Sea (transition zone), and green triangles are for marinas located in southern Brittany (Lusitanian province). Marina codes are as follows: SM = Saint-Malo, SQ = Saint-Quay-Portrieux, PG = Perros-Guirec, BLO = Bloscon (Roscoff), AW = L'Aber Wrac'h, MB = Moulin Blanc (Brest), CAM = Camaret-sur-Mer, CON = Concarneau, ET = Étrel, TRI = La Trinité-sur-Mer.

The 'quadrat' dataset is composed of a list of all species collected within 0.25 m<sup>2</sup> quadrats (20 replicates per marina), scraped by divers. All specimens (> 5 mm) were morphologically identified on site, to the lowest possible taxonomic level (family, genus or species). Solitary animals were counted whereas the abundance of colonial organisms was measured using a semi-quantitative scale with four levels (0 to 3).

The three eDNA datasets were obtained through metabarcoding of DNA extracted from seawater samples (2L per pontoon filtered on site; two pontoons per marina, see [Chapter II.1](#)), using three different markers. First, a 389-bp to 489-bp portion of the 18S rDNA (V1-V2 regions) was amplified using primers designed by Fonseca, et al. (2010). Then a 365-bp COI portion was amplified using primers designed by Leray, et al. (2013). Finally a 159-bp portion of the 16S mitochondrial rDNA was amplified using primers designed by Kelly, et al. (2016). These three markers were chosen because all were designed to target metazoan taxa and complement each other in terms of resolution and amplification bias. Negative controls were included at each step (6 for sampling, 2 for DNA extraction and 3 for PCR). Library preparation was performed using a dual-barcoded, dual-indexed two-step PCR procedure detailed in Couton et al. (submitted; see [Chapter I](#)), with five tagged PCR replicates for each sample. Sequencing was performed in three different runs on an Illumina® MiSeq sequencer. The COI and 18S markers were sequenced using a 600 cycles v3 protocol with two index reads and the 16S marker was sequenced using a 300 cycles v2 protocol with two index reads.

The three datasets were then processed similarly, as detailed in [Chapter II.1](#), using CUTADAPT v-2.8 (Martin 2011) to remove primers and tags, and DADA2 v-1.13.1 (Callahan, et al. 2016) to produce a set of amplicon sequence variants (ASVs).

## **b. ASVs clustering and taxonomic assignment**

To limit the impact of intraspecific variability on later diversity analyses (Brandt, et al. 2020), ASVs were clustered into OTUs using SWARM-3.0.0 (Mahé, et al. 2015), with the *d* parameter set to 1 as advised by the authors. The *fastidious* option was then used to refine clustering with different values for the *b* parameter for each marker (see Table S1). The OTU table was then filtered for index-jump correction (see [chapter II.1](#)) and only OTUs found in at least two out of five PCR replicates per sample were retained.

A first taxonomic assignment was performed in order to evaluate the proportion of reads assigned to the different kingdoms and metazoan phyla for each marker. ASVs from each dataset were aligned against references retrieved from the GenBank nt database using the *ecotag* command from the OBITOOLS v-1.2.11 package (Boyer, et al. 2016) with no minimum identity threshold. Then an assignment at a

lower taxonomic level was performed by aligning all ASVs against a restricted reference database ([Chapter II.1](#)) using the BLAST® command line tool (Altschul, et al. 1990). This reference database focused on ten metazoan phyla, most of them commonly observed in the scraped samples, and/or found on or near floating pontoons: Annelida, Arthropoda (only crustaceans), Bryozoa, Chordata (only fish and tunicates), Cnidaria, Echinodermata, Mollusca, Nemertea, Platyhelminthes, and Porifera. Only alignments covering 99% of the subject or query sequence and above the chosen identity threshold for each marker (18S: 99%; COI: 92%; 16S: 97%; see [chapter II.1](#)) were considered. If one ASV matched with several references, it was assigned to the one with the highest identity percentage. If two alignments with different references had the same identity percentage, the ASV was assigned to the lowest common taxonomic rank. All assignments at a rank higher than the family were classified as “unassigned”. Each OTU was given the assignment of its most abundant ASV.

### c. Alpha and beta-diversity analyses

The quadrat dataset was composed of colonial taxa for which semi-abundance data (rank from 0 to 3) were obtained and solitary ones for which individual counts were obtained. In order to analyse together the whole dataset, individual counts were fourth-root transformed prior to analyses. For the eDNA datasets, after read processing and filtering, the number of reads was significantly higher in samples from spring 2018 than in those from fall 2017 for all markers (paired Student t test, 18S:  $t_{28}=3.37$ ,  $P=0.001$ ; COI:  $t_{41}=3.15$ ,  $P=0.002$ ; 16S:  $t_{27}=6.28$ ,  $P<0.001$ ). To avoid introducing biases when comparing the two seasons, the three datasets were thus rarefied using the `rrarefy` function in the VEGAN-2.5.2 R package (Oksanen, et al. 2018). The total number of reads for each sample was set to match the one from the sample with the lowest number of reads (18S: 97782; COI: 98201; 16S: 32334). These cut-off values were located after reaching the plateau of the rarefaction curve for most samples (Figs. S1-S3).

Diversity analyses were performed on three to four subsets for each DNA marker (i.e. 11 in total). The first subset is composed of all OTUs retained after filtering, without taxonomic assignment. The three others considered distinct taxonomic or functional groups, with the second subset composed of OTUs assigned to the kingdom Chromista (hereafter named ‘Chromista’), the third subset composed

of OTUs assigned to metazoan taxa that are holopelagic or highly mobile (hereafter named 'mobile'), and the fourth subset composed of OTUs assigned to metazoan taxa comprising a benthic stage fixed to hard substrates or poorly mobile (hereafter named 'benthic'). Table S2 gives the list of taxa selected in each subset. The 'Chromista' subset was absent from the 16S dataset because this marker targets exclusively metazoans. Finally, analyses were also performed on a presence/absence dataset composed of BLAST® assignment results for all three markers combined and targeting the 'benthic' subset. To avoid redundancy, assignments to the family were discarded if any species or genus was already present for this family, and assignments to the genus were discarded if several species of the same genus were already present. If only one species of this genus was already identified, both assignments were pooled to the genus level.

Alpha-diversity was estimated based on the OTU/taxon richness and the Shannon index. They were calculated for each dataset and subset, using the diversity function of the VEGAN-2.5.2 R package. For the presence/absence taxa dataset (all markers combined), only richness was computed. A Wilcoxon-Mann-Whitney test was used to test for equality in alpha diversity between pontoons, seasons (paired test), or region pairs.

Community compositions were compared using Principal Component Analyses (PCA) as suggested in Borcard, et al. (2018). For each dataset and subset the Hellinger transformation was applied prior to analyses (Legendre and Gallagher 2001). The effect of the season, marina and region, and their interactions, was evaluated using a permutational multivariate analysis of variance (PERMANOVA, 9999 permutations) with marina nested within region. All analyses were conducted using the VEGAN-2.5.2 R package. Further pairwise comparisons were performed for all three biogeographic regions using the pairwise.adonis2 function in R (Martinez Arbizu 2020).

In order to evaluate the correspondence between community dissimilarity and geographic distance, beta-diversity indices were calculated on the quadrat dataset (Bray-Curtis) and the metabarcoding all markers dataset (Jaccard) with the VEGAN R package. These calculations were performed for all comparisons between each pontoon for each season separately. These indices were compared to the geographic distance between marinas using a Mantel test (9999 permutations) with the ADE4 R package. The approximate distances between each marina were calculated on Google maps by drawing straight lines around the coast.

The local (for each pontoon) and species contributions to beta diversity (LCBD and SCBD) were calculated on the quadrat dataset and the eDNA dataset based on OTU assignments across all markers for each season separately using the *beta.div* function of the ADESPATIAL-0.3.8 R package (Dray, et al. 2020).

#### d. Non-indigenous species diversity and contribution

In order to evaluate the contribution of non-indigenous species (NIS) on community diversity and structure patterns, the introduction status was determined for all species identified in the quadrat dataset and the OTU-based presence/absence eDNA benthic subset (all markers combined). Species were classified as either NIS, native or cryptogenic according to various databases (Table S3). For one species, the status could not be determined with certainty and it was classified as “undetermined”. For 16 other NIS and cryptogenic species, the taxonomic assignment was questionable mainly because of a lack of reference sequences for a closely related native species. These 17 species (out of 198) were not included in the analyses.

Taxon richness was calculated for the two species-based datasets for NIS/cryptogenic, and native species separately. Principal component analyses as well as PERMANOVA were performed as above for NIS/cryptogenic and native species separately.

## Results

#### a. OTUs distribution and taxonomic assignments

In total, the MiSeq runs yielded 15,754,233 reads, 13,108,322 reads, and 13,396,162 reads, for 18S, COI, and 16S, respectively, resulting in a total of 8708, 12738, and 896 unique ASVs. They were further clustered into 5859, 8811, and 480 OTUs, respectively. The index-jump and replicate filtering steps led to 3101, 3186, and 240 OTUs for 18S, COI, and 16S, respectively (Table 1).

The clustering of ASVs in OTUs resulted in a decrease in diversity of 32.7%, 30.8%, and 46.4% for 18S, COI, and 16S, respectively. Most resulting OTUs grouped



ASVs with similar taxonomic assignments (18S: 98.6%, COI: 99.0%, and 16S: 85.8%), suggesting that clustering is effective in converging towards an OTU/species ratio of 1. However one and eight OTUs grouped ASVs assigned to taxa from different families for 18S and 16S, respectively (Table S7). In addition, 37, 44, and 49 OTUs clustered together ASVs with different taxonomic assignment within the same family or genus for 18S, COI and 16S respectively. Most of these differences, however, were due to one of the ASV or the OTU being only assigned to the genus or family-level (18S: 25 out of 37, COI: 40 out of 44, and 16S: 39 out of 49). No posterior modification was done on OTU clustering based on taxonomic assignment.

**Table 1** Number of OTUs found in each of the eleven eDNA subsets for all three markers used in this study as well as the number of taxa assigned combining assignment across all markers.

	18S	COI	16S	All markers
All OTUs	3101	3186	240	
Chromista	1655	753	-	
Mobile	81	90	34	
Benthic	219	334	92	251

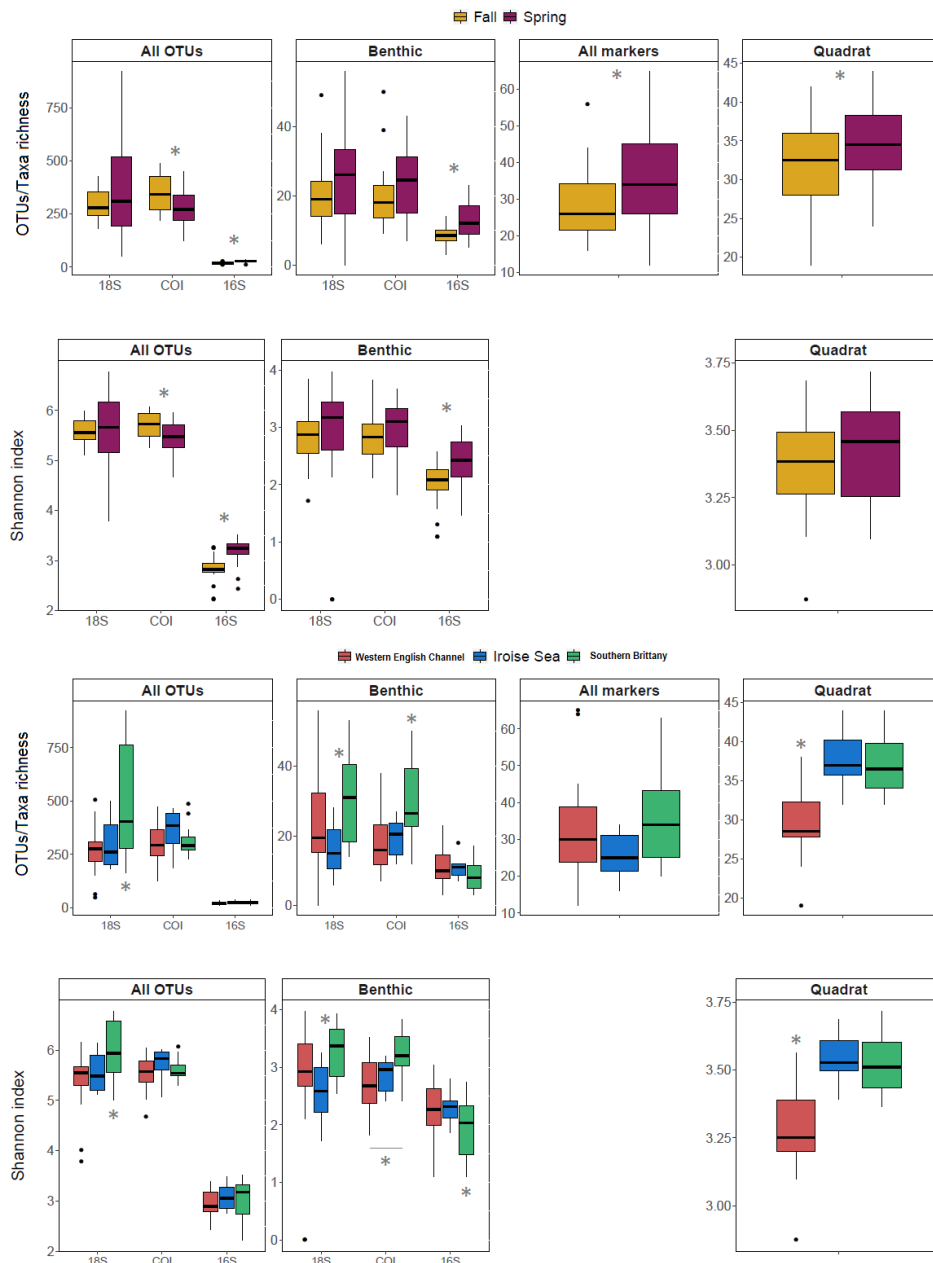
When aligning OTUs against our restricted database, 9.5%, 18.1%, and 43.3% of them were assigned to a species, genus or family of the targeted metazoan phyla for 18S, COI, and 16S, respectively. The number of taxa identified, however, was lower for 16S (93) than for 18S (267) and COI (297). Out of the 198 species identified with metabarcoding, across all markers, in the benthic fraction, 21 were NIS, 12 were cryptogenic and 148 were native (Table S8).

## b. Alpha diversity patterns

Over all datasets and subsets, the OTU/taxa richness was higher in spring (Fig. 2), although the difference was not significant for 18S and the benthic subset for COI. The same pattern was observed for the Shannon index, albeit not significant for the quadrat dataset. For the "All OTUs" subset with COI, the opposite scenario was observed, with a richness and Shannon diversity significantly higher in fall ( $V=177$ ,  $P=0.008$ ;  $V=191$ ,  $P<0.001$ , respectively).

When focusing on differences among regions, the Southern Brittany and the Iroise Sea displayed higher diversity than the Western English Channel, with both indices for the quadrat dataset. With eDNA-metabarcoding, only the Southern

Brittany exhibited significantly higher diversity than the two other regions, with 18S for the all OTUs subset, and COI for the benthic subset. In addition, this region was also significantly more diverse than the Iroise Sea with 18S for the benthic subset and had a higher Shannon index than the Western English Channel with 16S for the same subset.

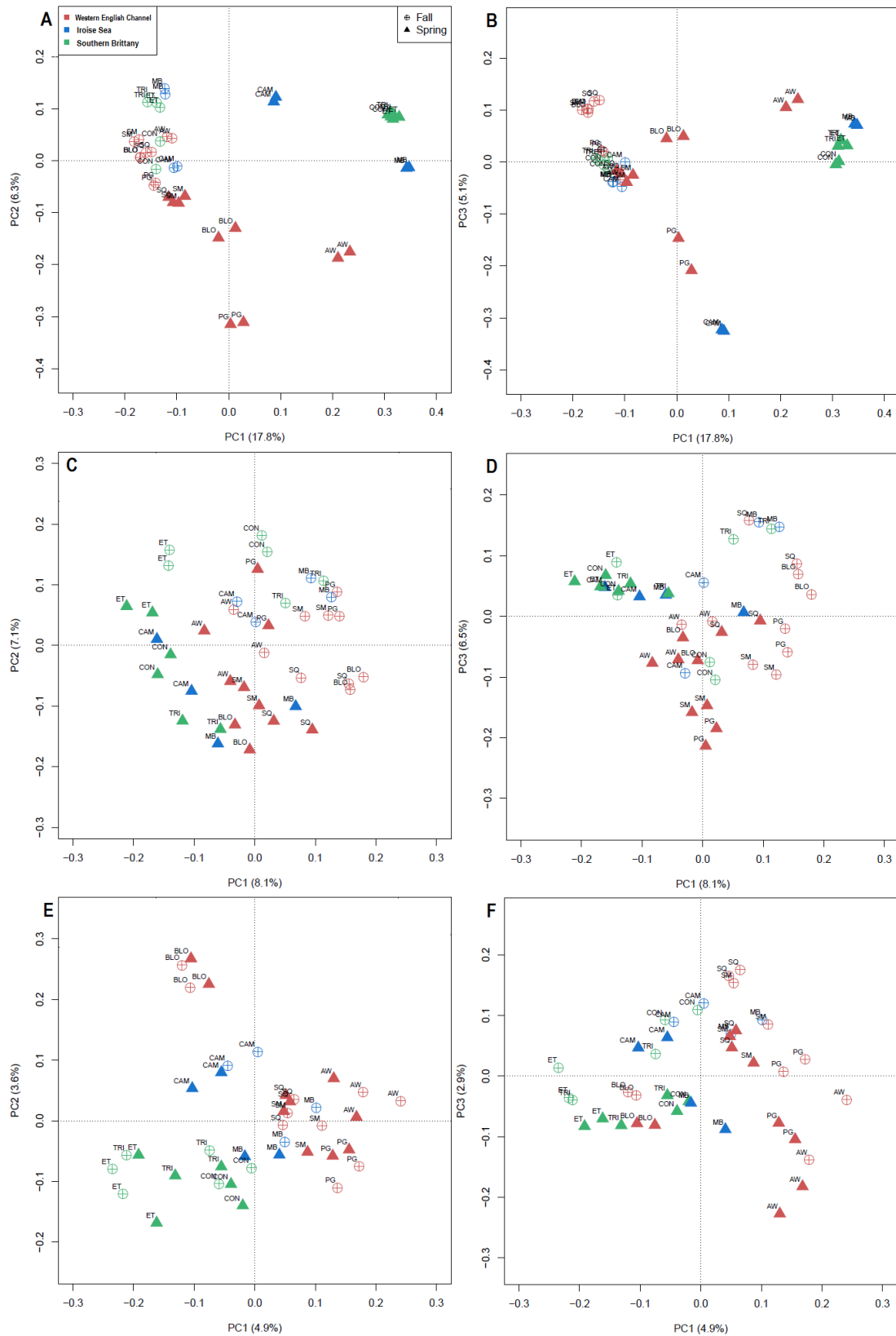


**Figure 2** Alpha-diversity between seasons (top panels) and regions (bottom panels), for 1) all OTUs, 2) OTUs of the 'benthic' subset, for the three markers, 3) All markers, benthic taxa assigned across the three markers combined (richness only), and 4) benthic taxa identified from quadrats. Significant differences (Wilcoxon test,  $P < 0.05$ ) are indicated with grey stars. A star between two bars indicates a significant difference between these two bars. A star over a bar indicates a significant difference between this bar and the two others.

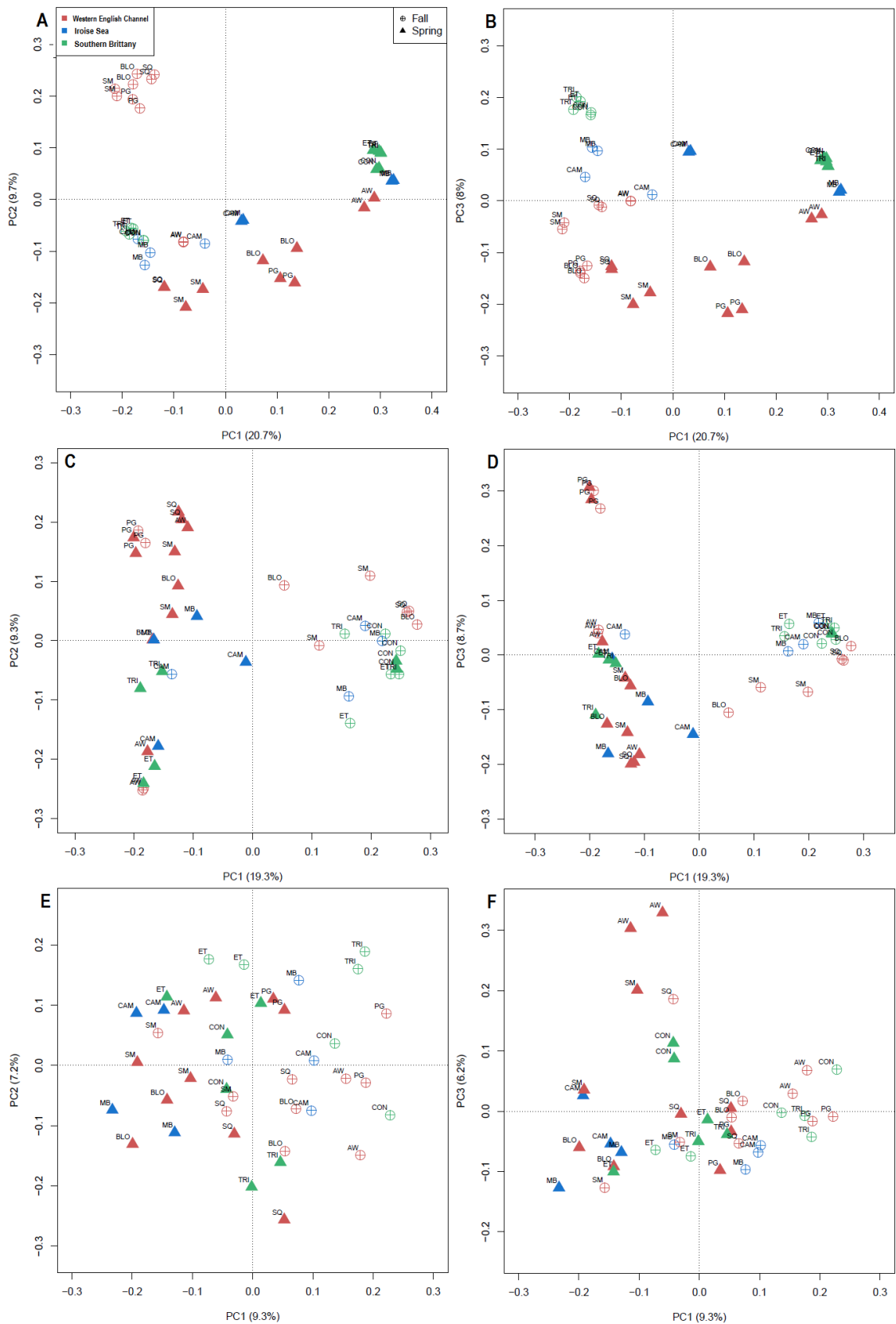
### c. Community dissimilarities

For each marker, considering all OTUs, the community structure was significantly influenced by time (season) and space, and their interactions, with a stronger spatial effect (PERMANOVA,  $P < 0.001$ ). Marinas were responsible for the main effect (18S:  $R^2 = 0.294$ ; COI:  $R^2 = 0.248$ ; 16S:  $R^2 = 0.305$ ; Table S7). The seasonal effect was stronger than the regional effect for 18S and COI. The 16S dataset exhibited lower yet still significant seasonal and regional effects. Looking at sample ordinations, for each season, the two pontoons of a given marina were clustered together, and the marinas were globally distributed according to their region (Figs. 3A-B, S4A-B, and S5A-B). When testing dissimilarities between pairs of regions, all three pairs were significantly different with all three markers and the Southern Brittany region was consistently more different from the two other regions (Table S7).

When examining separately the three subsets (Chromista, Mobile, and Benthic) for each of the three markers, the spatial and temporal effects were still significant (Table S8), with an important effect of marinas ( $R^2$  ranging from 0.237 to 0.420). For Chromista, season (18S:  $R^2 = 0.160$ ; COI:  $R^2 = 0.216$ ) and region (18S:  $R^2 = 0.127$ ; COI:  $R^2 = 0.145$ ) also displayed a significant effect (Figs. 4A-B and S4C-D). The two metazoan subsets, however, exhibited lower effects of season and region ( $R^2$  ranging from 0.025 to 0.122; Table S10; Figs. 4C-F, S4E-H, and S5C-F). Pairwise tests between regions were significant for all region pairs, with all subsets for COI and 18S. For 16S, however, not all pairwise differences were significant and depend on the subset (Table S11). When analysing the 'All markers' dataset, composed of taxa identified across the three markers (Table 2; Fig 3C-D), marinas again explained most of the structure ( $R^2 = 0.292$ ,  $P < 0.001$ ) but, contrary to what was found with the other eDNA datasets, the regional effect was stronger ( $R^2 = 0.101$ ) than the seasonal effect ( $R^2 = 0.062$ ). Pairwise tests showed a significant difference between all regions with a greater difference between Southern Brittany than the two other regions (Table 3).



**Figure 3** Ordination plots of principal component analysis results from Hellinger-transformed OTUs/taxa abundances (or occurrences) for pontoons from each locality according to their region (color) and the season of sampling (shape) on three datasets: All OTUs obtained with COI (A and B), benthic taxa assigned across all markers (C and D), and taxa from quadrat sampling (E and F). Sample scores are displayed in scaling 1. See figure 1 for location codes. Left panel shows ordination with axes PCA1 and PCA2, right panel shows ordination with axes PCA1 and PCA3.



**Figure 4** Ordination plot of principal component analysis results from Hellinger-transformed OTU abundances for pontoons from each locality according to their region (colors) and the season of sampling (shape) on the three COI subsets: Chromista (A and B), Mobile (C and D), and Benthic (E and F). Sample scores are displayed in scaling 1. See figure 1 for location codes. Left panel shows ordination with axes PCA1 and PCA2, right panel shows ordination with axes PCA1 and PCA3.

**Table 2** PERMANOVA results comparing community composition between biogeographic regions, between marinas within regions, and between seasons, for the quadrat dataset and eDNA dataset based on taxonomic assignments across all three markers (9999 permutations). Non-significant values are in bold.

Quadrat					
	df	Sum of squares	R <sup>2</sup>	F	P
<b>Region</b>	2	2.049	0.190	8.820	<0.001
<b>Season</b>	1	0.609	0.056	5.241	<0.001
<b>Marina (Region)</b>	7	4.456	0.413	5.481	<0.001
<b>Region x Season</b>	2	0.369	0.034	1.587	0.027
<b>Marina (Region) x Season</b>	<b>7</b>	<b>0.975</b>	<b>0.090</b>	<b>1.199</b>	<b>0.101</b>
<b>Residuals</b>	20	2.323	0.215		
All markers (benthic taxa assigned from eDNA dataset across all markers)					
	df	Sum of squares	R <sup>2</sup>	F	P
<b>Region</b>	2	2.885	0.101	3.437	<0.001
<b>Season</b>	1	1.752	0.062	4.174	<0.001
<b>Marina (Region)</b>	7	8.321	0.292	2.832	<0.001
<b>Region x Season</b>	2	1.651	0.058	1.967	<0.001
<b>Marina (Region) x Season</b>	7	5.478	0.192	1.864	<0.001
<b>Residuals</b>	20	8.394	0.295		

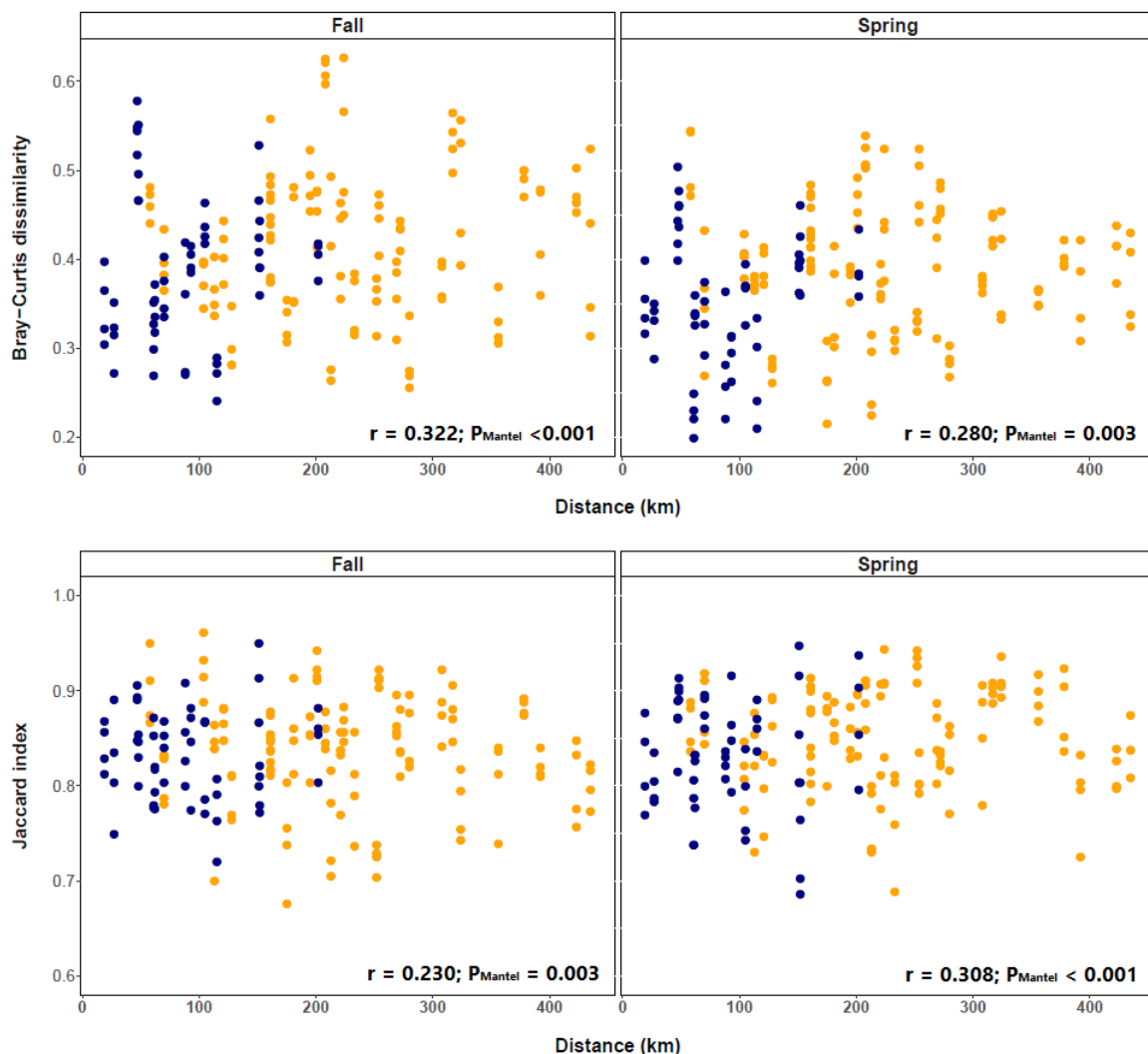
**Table 3** PERMANOVA pairwise comparisons for the three biogeographic regions. The test was computed on the quadrat dataset and eDNA dataset based on taxonomic assignments across all three markers. Due to the positive interaction between seasons and regions (see Table 2), permutations were constrained within seasons for each region (9999 permutations). WEC = Western English Channel, SB = Southern Brittany, and IS = Iroise Sea.

Quadrat					
	df	Sum of squares	R <sup>2</sup>	F	P
<b>WEC vs. SB</b>	1	1.445	0.165	5.933	<0.001
<b>WEC vs. IS</b>	1	0.710	0.100	2.880	0.001
<b>SB vs. IS</b>	1	0.791	0.174	3.803	<0.001
eDNA dataset from taxonomical assignments across all markers					
	df	Sum of squares	R <sup>2</sup>	F	P
<b>WEC vs. SB</b>	1	1.633	0.072	2.340	<0.001
<b>WEC vs. IS</b>	1	1.331	0.068	1.904	<0.001
<b>SB vs. IS</b>	1	1.295	0.097	1.929	<0.001

Similarly to what was obtained with the eDNA datasets and subsets, PERMANOVA on the quadrat dataset revealed significant spatial and temporal effects on the community structure (Table 2), with marinas responsible for the major part of the structure ( $R^2=0.413$ ). The effect of the region was stronger ( $R^2=0.190$ ) than the

effect of the season ( $R^2=0.056$ ; Table 2; Fig. 3E-F). Pairwise tests revealed a significant difference between all regions with a greater difference between Southern Brittany than the two other regions (Table 3).

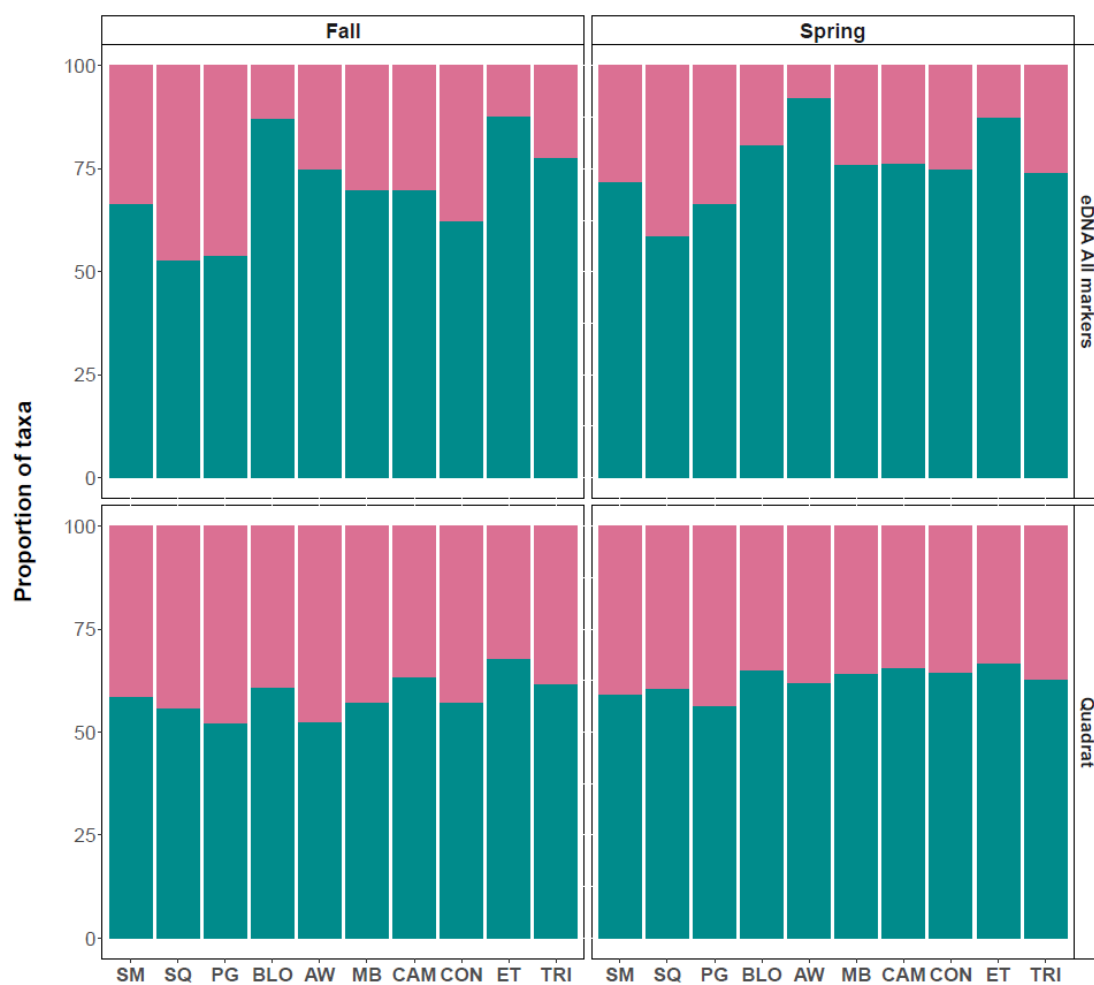
Mantel tests showed that beta-diversity matrices and the geographic distance matrix were not independent (Fig. 5) for all seasons. The corresponding correlation coefficients were, however, low, ranging from 0.230 to 0.322. Based on LCBD calculations on the same two datasets, for each season separately, no locality contributed more significantly than the others to the total beta diversity (Table S12).



**Figure 5** Correlation between the Bray-Curtis dissimilarity index for the quadrat dataset (top) or Jaccard index for the eDNA dataset ('All markers' subset) (bottom) and the "sailing" distance between marinas. Indices were calculated between pontoons for each season separately. Comparisons of pontoons from marinas from the same biogeographical region are coloured in dark blue whereas comparisons of pontoons from marinas from different biogeographical regions are coloured in orange.

### d. Non-indigenous species contribution to community patterns

The proportion of identified non-indigenous and cryptogenic species was higher in the quadrat dataset than in the eDNA benthic all markers subset (Fig. 6). In the quadrat dataset, all marinas had proportions of NIS similar in both seasons, with values ranging from 32% for ET in fall to 50% for PG in fall. The variations were more important for the eDNA dataset but still similar in both seasons with values ranging from 8% for AW in spring to 47% for SQ in fall.



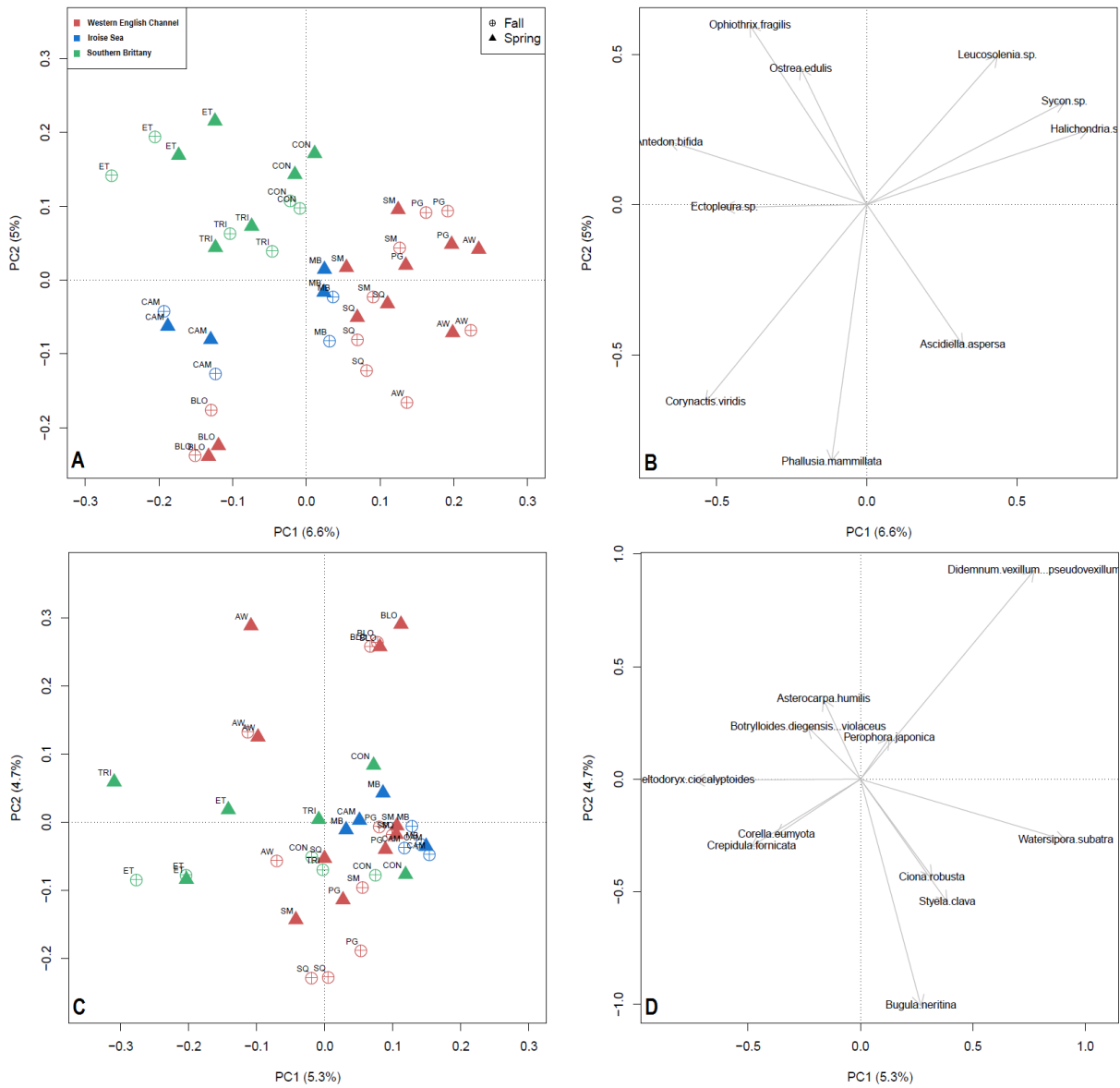
**Figure 6** Proportion of non-indigenous and cryptogenic species (pink) and native species (seagreen) in every marina, in fall (left panels) and spring (right panels). Only data of the benthic taxa identified from assignment across all markers (upper panels) and data of the quadrat dataset (lower panels) are displayed. See figure 1 for location codes.



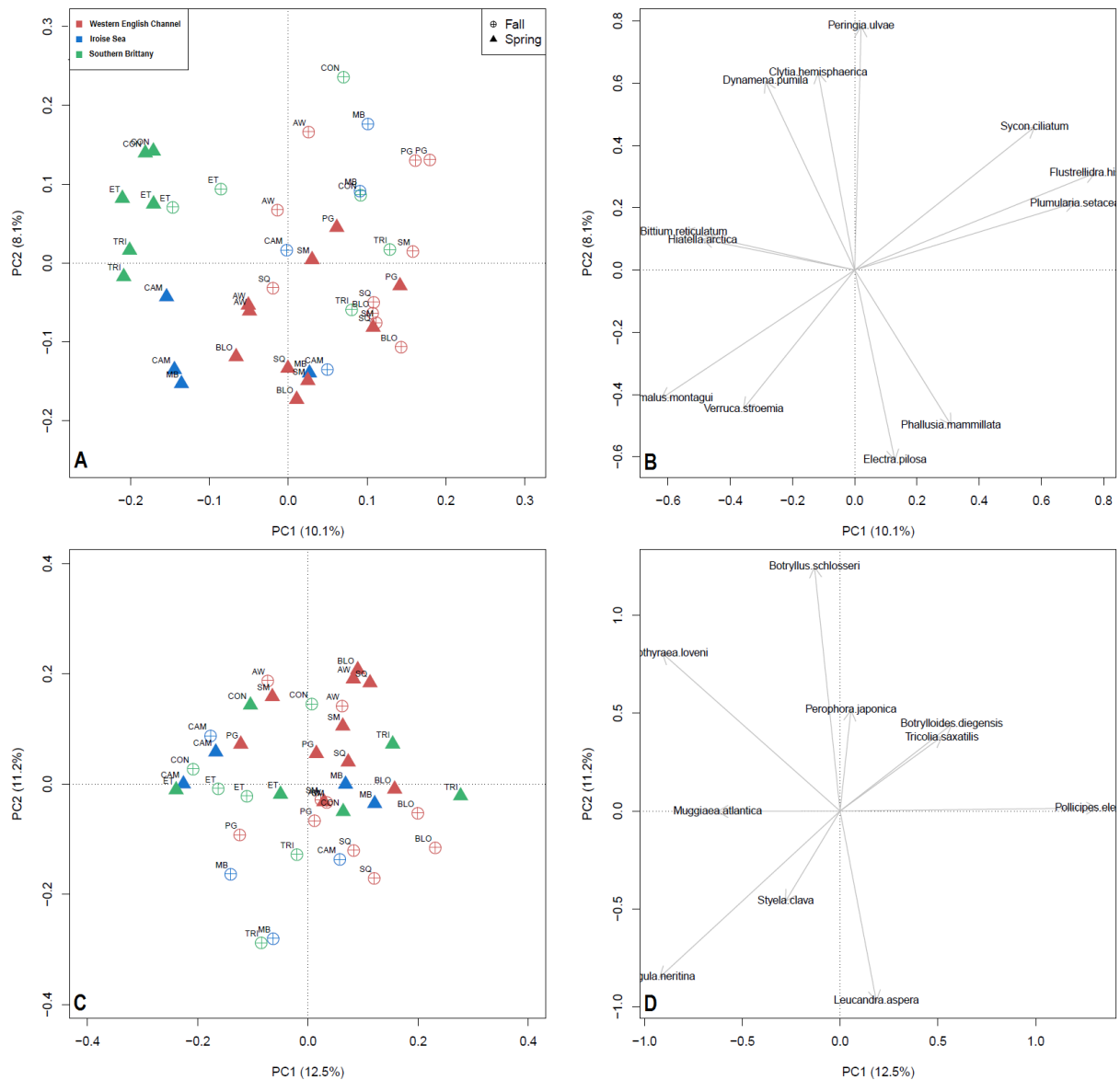
Re-analyzing the community structure from these two datasets considering the native species and the non-indigenous and cryptogenic species separately, significant spatial and temporal effects were observed (Table 4). There was no significant interaction between seasons and spatial factors for the NIS and cryptogenic species, with the quadrat dataset. Marinas were still the factor explaining the highest proportion of the total variance ( $R^2$  ranging from 0.288 to 0.414), with a stronger effect in the quadrat dataset. In the two datasets, the spatial and temporal factors explained a lower proportion of NIS community structure, as compared to native species. This pattern is well-illustrated by the PCA ordination plots (Figs. 7 and 8), in which the first two axes did not efficiently discriminate regions for NIS and cryptogenic species as compared to native species.

**Table 4** PERMANOVA results comparing community composition between biogeographic regions, between marinas within regions, and between seasons, for the quadrat dataset and eDNA dataset based on taxonomic assignments across all three markers (9999 permutations). Calculations were performed considering either native species or non-indigenous and cryptogenic species. Non-significant values are in bold.

Quadrat									
	Native species					NIS and cryptogenic species			
	df	Sum of squares	R <sup>2</sup>	F	P	Sum of squares	R <sup>2</sup>	F	P
<b>Region</b>	2	2.049	0.190	9.124	<0.001	1.216	0.156	5.921	<0.001
<b>Season</b>	1	0.609	0.056	3.761	<0.001	0.271	0.035	2.639	0.015
<b>Marina (Region)</b>	7	4.456	0.413	5.552	<0.001	3.239	0.414	4.506	<0.001
<b>Region x Season</b>	<b>2</b>	<b>0.369</b>	<b>0.034</b>	<b>1.364</b>	<b>0.116</b>	<b>0.253</b>	<b>0.032</b>	<b>1.232</b>	<b>0.243</b>
<b>Marina (Region) x Season</b>	<b>7</b>	<b>0.975</b>	<b>0.090</b>	<b>1.148</b>	<b>0.197</b>	<b>0.786</b>	<b>0.101</b>	<b>1.094</b>	<b>0.327</b>
<b>Residuals</b>	20	2.323	0.215			2.054	0.263		
eDNA dataset from taxonomical assignments across all markers									
	Native species					NIS and cryptogenic species			
	df	Sum of squares	R <sup>2</sup>	F	P	Sum of squares	R <sup>2</sup>	F	P
<b>Region</b>	2	3.203	0.107	3.778	<0.001	2.526	0.095	3.053	<0.001
<b>Season</b>	1	2.000	0.067	4.718	<0.001	1.462	0.055	3.535	<0.001
<b>Marina (Region)</b>	7	8.789	0.292	2.961	<0.001	7.670	0.288	2.649	<0.001
<b>Region x Season</b>	2	1.756	0.058	2.071	<0.001	1.270	0.048	1.535	0.021
<b>Marina (Region) x Season</b>	7	5.824	0.194	1.962	<0.001	5.416	0.203	1.871	<0.001
<b>Residuals</b>	20	8.480	0.282			8.273	0.311		



**Figure 7** Ordination plots of principal component analysis results from Hellinger-transformed taxa abundances between localities (left) and the ten species contributing the most to each axis (right) according to their region (colors) and the season of sampling (shape) on the quadrat dataset. Analyses were performed on either native species (A-B) or non-indigenous and cryptogenic species (C-D). Both sample scores and species scores are displayed in scaling 1. See figure 1 for location codes.



**Figure 8** Ordination plots of principal component analysis results from Hellinger-transformed taxa occurrences between localities (left) and the ten species contributing the most to each axis (right) according to their region (colors) and the season of sampling (shape) on the eDNA dataset based on taxonomic assignments across all markers. Analyses were performed on either native species (A-B) or non-indigenous and cryptogenic species (C-D). Both sample scores and species scores are displayed in scaling 1. See figure 1 for location codes.

The ten taxa contributing the most to the spatial structure of the communities at each season accounted for approximately 30% and 15% of the total beta diversity for the quadrat and the eDNA benthic subset across all markers, respectively, as shown by the SCBD analysis (Table 5). Out of the 14 and 32 NIS or cryptogenic taxa identified within the quadrat dataset and the eDNA dataset,

respectively, four and three were present in the ten highest contributors for each dataset, and none in common. NIS and cryptogenic taxa accounted for 23% and 13% of the total taxa from the quadrat dataset and the eDNA dataset respectively. These same taxa contributed for 23% and 21% to the total beta diversity in the fall and spring respectively for the quadrat dataset. For the eDNA dataset, they contributed for 19% and 16% to the total beta diversity in the fall and spring respectively.

**Table 5** List of the ten taxa showing the highest contribution to beta diversity for each season within either the quadrat dataset or the eDNA dataset based on OTU assignments across all markers. Species identified as non-indigenous or cryptogenic are in bold.

Quadrats			
Fall		Spring	
Taxon	SCBD	Taxon	SCBD
<i>Haliclona</i> sp.	0.034	<i>Haliclona</i> sp.	0.038
<i>Halichondria</i> sp.	0.033	<i>Oscarella</i> sp.	0.035
<b><i>Didemnum vexillum/pseudovexillum</i></b>	0.031	<i>Leucosolenia</i> sp.	0.032
<b><i>Watersipora subatra</i></b>	0.031	<i>Sycon</i> sp.	0.031
<i>Antedon bifida</i>	0.030	<i>Antedon bifida</i>	0.031
Other Serpulidae	0.028	<i>Corynactis viridis</i>	0.030
<i>Aplidium glabrum</i>	0.028	<i>Perforatus perforatus</i>	0.029
<b><i>Ciona robusta</i></b>	0.027	Metridioidea	0.029
<i>Oscarella</i> sp.	0.026	<b><i>Bugula neritina</i></b>	0.028
<i>Phallusia mammillata</i>	0.025	<b><i>Watersipora subatra</i></b>	0.028
eDNA dataset from assignment across all markers			
Fall		Spring	
Taxon	SCBD	Taxon	SCBD
<i>Clytia hemisphaerica</i>	0.017	<i>Malacoceros fuliginosus</i>	0.015
<i>Phallusia mammillata</i>	0.017	<i>Electra pilosa</i>	0.013
<i>Lagis koreni</i>	0.016	<i>Flustrellidra hispida</i>	0.013
<b><i>Austrominius modestus</i></b>	0.016	<b><i>Bugulina fulva</i></b>	0.013
<i>Laomedea flexuosa</i>	0.015	<i>Spio</i> sp.	0.012
<i>Halichondria</i> sp.	0.015	<b><i>Gonothyraea loveni</i></b>	0.012
<i>Asciella</i> sp.	0.015	<i>Hymeniacidon</i> sp.	0.012
<i>Ciona intestinalis</i>	0.015	<i>Plumularia</i> sp.	0.012
<i>Peringia ulvae</i>	0.015	<i>Plumularia setacea</i>	0.012
<i>Plumularia setacea</i>	0.015	<i>Peringia ulvae</i>	0.011

## Discussion

Marinas are singular habitats displaying particular species assemblages. They are hotspots of biological introductions, especially for fouling organisms because of both the dispersal vectors (boat hulls) and the artificial hard substrates they offer to sessile benthic species. Their biodiversity needs to be further monitored, to understand how they function and evaluate the impact of non-indigenous species (NIS) on local communities. We used a combination of a morphology-based assessment by quadrat scraping under floating pontoons, and a molecular-based approach through eDNA metabarcoding to describe the diversity within and among ten French marinas spread over two biogeographic regions and a transition zone. All datasets showed a significant temporal (fall *vs.* spring) and spatial differentiation of their community compositions. This result was observed not only for benthic sessile species assemblages but also for mobile ones. Further, non-indigenous and native species showed similar diversity patterns, although differences among biogeographic regions were lower for NIS.

### **a. Spatial differences suggest that natural processes are not fully overcome by human-mediated processes**

With the increase of species transfers at a global scale (Seebens, et al. 2017), the ecosystems around the world tend to homogenize (Smart, et al. 2006; Magurran, et al. 2015). Furthermore, species range distributions are expected to be less limited by their dispersal abilities and natural barriers (overcome by human-mediated transport) than by the environmental conditions and species interactions that would control their establishment (Capinha, et al. 2015). In this context, marinas might be particularly affected because of the high proportion of NIS they harbour and the high propagule and colonization pressure they sustain through shipping (Clarke Murray, et al. 2011). Their particular environmental conditions might favour the settlement of a few highly tolerant and invasive species to the detriment of less resilient native ones (Piola and Johnston 2009; Rivero, et al. 2013; Johnston, et al. 2017), or the establishment of species already selected for being efficient foulers during their transport (Briski, et al. 2018). This environmental filtering might not only select NIS, but also particular native species, such as those limited by light, that would benefit from the shade of pontoons.

Consequently, we expected to find evidence of biotic homogenization across the study marinas, particularly for sessile species established on hard substrates. However, at the scale of our study, we did not find any evidence supporting this hypothesis. Moreover all “functional” groups considered (i.e. Chromista, mobile and benthic) exhibited a significant spatial structure, with differences between biogeographic regions and between marinas within regions, a result observed whatever the approach used (either traditional or eDNA-based; Tables 2, S7 and S8, Figs. 3 and 4). Finding regional structure for Chromista and metazoan pelagic groups could have been expected because they do not rely only on their transport via recreational boating, which is mainly done through hull fouling (Clarke Murray, et al. 2011). Nevertheless, the benthic subsets, comprising species living fixed on hard substrates, exhibited a significant spatial effect, although weaker than the other fractions. Marina (nested in biogeographic region) was the factor that explained the highest proportion of the total variance for every subset, whatever the “functional” group considered. This suggested that human-mediated dispersal is not strong enough to homogenize the species assemblages at a regional scale, not only between regions but also within regions. This could be explained either by a human-mediated dispersal weaker than expected, or by the existence of characteristics specific to each marina that have a strong impact on the communities they harbour. In Brittany, sailing habits appear to mostly consist in short-term trips, typically one day, longer cruises being rarer, so that sailing distances might be limited to intra-regional short distances in most cases (Sonnich 2008; Baudelle, et al. 2011).

The dissimilarity between benthic communities fixed on hard substrates from all marinas, was correlated with the geographic distance (Fig. 5). This result contrasts with population genetics studies of both native (e.g. the ascidian *Ciona intestinalis*; Hudson, et al. 2016), and non-native species (e.g. the seaweed *Undaria pinnatifida*; Guzinski, et al. 2018), carried out in the same region, which showed a chaotic genetic structure, best explained by unpredictable shipping routes. This result is, however, congruent with the study by López-Legentil, et al. (2015) who showed a significant correlation between non-indigenous ascidian diversity and geographic distance between marinas in Spain. These authors suggested that boats in this area might only travel at short range, thus limiting the spread of these ascidian species. In our case, the correlation between dissimilarity and distance is likely explained by the regional effect (the two most distinct regions are those that are the most distant). As a matter of fact, marinas were distributed across distinct regions, known to be naturally separated by strong currents and to display contrasted environments (Gallon, et al. 2014), which might contribute to the observed community structuring. Environmental

factors might, thus, play an important role in the dissimilarity between assemblages across regions. Marinas located in Southern Brittany exhibited a higher diversity than those in the Western English Channel for several subsets and both traditional and eDNA-metabarcoding approaches, whereas the diversity of the Iroise Sea was closer to one or the other regions depending on the approach and the subset considered (Fig. 2). These results are congruent with the very different environmental conditions of the two regions (e.g. lower temperatures in the Western English Channel) and the intermediate situation of the Iroise Sea. Altogether, in our study region, human-mediated dispersal and specificities from highly anthropogenic habitats seem not strong enough to fully overcome natural processes. We lack, however, anterior descriptions of these same assemblages for comparison and we are thus not in a position for detecting a possible shift that might have occurred over time due to anthropogenic pressures.

### **b. Marinas are highly disturbed habitats that may lead to fast community turn-over**

Besides spatial structuring, marina communities were also strongly variable over time, with an overall higher diversity in spring, especially for the metazoan subsets (Fig. 2). The temporal effect (here spring *vs.* fall sampling) was even stronger than the regional effect for the 'all OTUs' and 'Chromista' subsets (Tables 2, S7 and S8, Figs. 3 and 4). This seasonal pattern might be explained by the high proportion of short-lived organisms in all the functional groups considered. The only subset displaying a weak seasonal pattern is the 16S mobile subset. It was, however, composed of ca. 25% of fish taxa because of the high affinity of the primers used for vertebrates, species that are long-lived organisms with a distribution less dependent on seasonality. Moreover, marinas are highly disturbed environments with frequent cleaning and pollution events which might increase species turnover. Species established under pontoons can rapidly collapse with changes in salinity due to rainfall and river input, as documented for the native and non-indigenous *Ciona* species in Brittany (Bouchemousse, et al. 2017), which would further contribute to seasonal patterns. Our results, especially those from the 'Chromista' and 'Mobile' subsets, are congruent with other studies that have documented, with metabarcoding, rapid changes in the zooplankton community structure in ports (Chain et al. 2016). They also support results obtained with traditional methods, that showed variation in species abundance during *in situ* monitoring (Albano and Obenat

2019), or when examining the colonization dynamics of novel substrates, such as experimental settlement plates (e.g. Leclerc, et al. 2020).

### c. The impact of non-indigenous species in marina community structure

Marinas are known to harbour a high proportion of NIS, especially in fouling communities fixed on artificial shallow substrates such as pontoons (Dafforn, et al. 2009b; Rivero, et al. 2013; López-Legentil, et al. 2015). For instance, Leclerc, et al. (2020) showed that NIS contribution to the community structure established on floating habitats was 10 times higher than on pilings. In our study, almost half of the species richness of fouling communities under pontoons in each marina was composed of non-indigenous or cryptogenic species (Fig. 6). The proportion was lower and more variable across marinas for the 'Benthic' subset from eDNA-metabarcoding with values ranging from 8% to 47%, which could be explained by the undetection of some NIS in some marinas with eDNA metabarcoding (see [Chapter II.1](#)), and the higher number of taxa identified in eDNA. These proportions were much lower than those reported in other studies where NIS represented the majority of species in their samples (Dafforn, et al. 2009a; López-Legentil, et al. 2015). López-Legentil, et al. (2015) however only considered ascidians which is well known for having a high invasibility potential, with many species globally distributed and invasive around the world (Zhan, et al. 2015). In our study, many taxa could not be identified to the species level (26% in the quadrat dataset and 33% in the eDNA dataset) and some taxa identified at the genus or family-level could include NIS. Moreover, several of the native species in our area are highly invasive species in other parts of the world, such as the ascidians *Ciona intestinalis* (Linnaeus, 1767) or *Ascidella aspersa* (Müller, 1776), both recognized as invasive species along the Atlantic coast of Canada (e.g. McDonald 2004; Ma, et al. 2019).

With both traditional and eDNA-metabarcoding approaches, the benthic metazoans fixed on hard substrates exhibited a strong spatial structuring, with both marina and region factors explaining 60% of the community structure for the quadrat dataset (Table 2). The regional effect was, however, lower for the benthic subsets with 18S and COI than for the other subsets (Tables S7-8, Figs. 4 and S4). Under the assumption that homogenization is driven by the spread of NIS, their low proportion in our datasets could explain the lack of evidence for community homogenization.



When looking at NIS and native species separately, however, a significant spatial effect was detected in the two subsets ('quadrat' and 'all markers'), although slightly weaker when considering only NIS (Table 5, Figs. 7-8). This suggested that NIS, at least some of them, play a role in the spatial structuring of the studied marinas. One to three NIS were found among the ten species contributing most to beta diversity, for each season and dataset (Table 6). It thus appears that, despite being potentially transported between all marinas of our study region and despite their potential invasiveness, NIS settlement and survival seem to be impacted by local biotic and abiotic conditions. For example, *Ciona robusta* Hoshino & Tokioka, 1967, a NIS present in many of our studied marinas, was never observed in l'Aber Wrac'h (AW) despite being observed in the neighbouring marina Moulin Blanc (MB, Brest) for many years (Bouchemousse, et al. 2016). This would indicate either that trips between these two marinas are not frequent enough, or that, even if individuals are transported to this marina, the local conditions do not allow this species to settle there. Some of the NIS however might be highly tolerant to various factors and are found in all the studied marinas. This is the case of the ascidian *Perophora japonica* Oka, 1927 which was observed in all ten marinas at both seasons, and thus is a poor contributor to the observed spatial patterns (All markers: 0.011 and 0.008 for Fall and Spring respectively; Quadrats: 0.012 and 0.005 for Fall and Spring respectively).

#### **d. eDNA metabarcoding as a monitoring tool in marinas**

The use of eDNA metabarcoding to describe biodiversity patterns in the marine environment is more and more widespread and has produced promising results (e.g. Lacoursière-Roussel, et al. 2018; Bakker, et al. 2019; Rey, et al. 2020). Aside from its practical advantages for sampling and processing the data, this approach offers the possibility to assess a broader taxonomic range than traditional methods, with a high spatial and temporal resolution. Some issues, however, such as reference database completion or primer bias could not be resolved at the moment and should be taken into account when interpreting the results (Zaiko, et al. 2018; Duarte, et al. 2020).

The conclusions drawn from results obtained by both traditional methods and eDNA metabarcoding were equivalent regarding spatial and temporal beta diversity patterns. Differences were, however, more important regarding alpha diversity estimates with metabarcoding. These differences could be the result of the limitations

of the quadrat sampling which obviously offers a more restricted view of the total community. They could also be linked to an erroneous increase of diversity with metabarcoding because of the persistence of sequencing or PCR errors or by the stochastic amplification of rare species that might bias the data obtained for each location independently. One further problem in metabarcoding is the primer amplification bias, which might also bias alpha diversity estimates. One way to circumvent this issue is to target several markers with several primers. The eDNA dataset created from taxonomic assignments across all markers allowed us to capitalize on their complementarity but resulted in the loss of the abundances information.

The only major difference that could be noted between the morphology-based and the molecular approaches is the relative importance of the seasonal factor in community structures. The three eDNA benthic subsets (i.e. based on OTUs for each marker) exhibited a stronger seasonal than regional effect, a pattern that was reversed in both the quadrat and the 'All markers' (i.e. after assignment of OTUs across the three markers) datasets. Moreover, this higher seasonal effect was not associated with a significant increase in richness, for either season, for the OTU-based benthic subsets. One explanation could be that the relative abundances for each OTU are contingent on the abundances of the others. The highest proportion of reads associated to Chromista in spring thus might have systematically resulted in the decrease of the number of reads associated to benthic taxa as compared to fall. On the contrary, spring being the reproductive period for many taxa, their read abundances could have been significantly higher in spring because of the sampling of a large amount of gametes released in the water. The relation between individual counts and reads abundances in metabarcoding have been tested by several studies with various outcomes (Lamb, et al. 2019) and quantitative information should be used with caution, especially for eDNA.

## Conclusion

Marinas are very disturbed marine environments with specific physical and chemical characteristics and the communities they shelter are not comparable to the ones found in surrounding natural habitats (Bulleri and Chapman 2010; Rivero, et al. 2013). They contain a diversity of micro-habitats that should all be monitored together in order to get an accurate representation of the global diversity and to be

able to detect shifts in communities. In total, the eDNA metabarcoding approach allowed us to retrieve information on three totally different functional groups with only one sample and a limited time and cost. Significant seasonal and regional effects were shown in all groups, whatever the method used. Our results suggest that, despite a human-mediated transport of fouling organisms between marinas, the biotic and abiotic conditions specific to each location play a major role in species assemblages. In this study we focused mainly on fouling communities under floating pontoons because they are supposed to harbour a higher number of NIS than other groups. In this context, water sampling was performed close to the pontoons, at 1 m under the surface to increase the chances of retrieving DNA from sessile organisms fixed to these substrates. This sampling methodology was thus poorly adapted for monitoring other communities such as benthic organisms dwelling in the sand. In order to get a complete image of marinas diversity, eDNA should be sampled in various locations and depth within each marina. Nevertheless, this approach would still be easier to perform than traditional sampling and could be done routinely for biomonitoring. Our results provide a baseline for further biodiversity assessments in the study area, in order to detect shifts in communities and the efforts should continue to gather information on a larger time scale.

## Acknowledgments

We are thankful to the Diving and Marine core service from the Roscoff Biological Station, and particularly Yann Fontana, Wilfried Thomas and Mathieu Camusat for the quadrat sampling, as well as Jérôme Coudret and Stéphane Loisel for their help for the field work. We thank Gwenn Tanguy from the Biogenouest Genomer Platform for advice and access to the sequencing facilities, and the Biogenouest ABIMS Platform for access to the calculation resources. This project was supported by the TOTAL Foundation (project AquaNIS2.0). MC acknowledges a PhD grant by Région Bretagne (ENIGME ARED project) and Sorbonne Université (ED 227 "Sciences de la Nature et de l'Homme").

## Authors' contributions

MC, LL, TC, and FV designed the study and/or the analyses; TC, LL, and FV supervised field work; MC, and CDT collected the molecular data; FV and LL supervised morphological identifications in the field and LL carry out further analyses in the lab; MC analysed the data; MC, TC, and FV led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

## References

- Airoldi L, Beck M. 2007. Loss, Status and Trends for Coastal Marine Habitats of Europe. *Oceanography and Marine Biology: An Annual Review* 45:345-405.
- Albano MJ, Obenat SM. 2019. Fouling assemblages of native, non-indigenous and cryptogenic species on artificial structures, depths and temporal variation. *Journal of Sea Research* 144:1-15.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403-410.
- Bakker J, Wangensteen OS, Baillie C, Buddo D, Chapman DD, Gallagher AJ, Guttridge TL, Hertler H, Mariani S. 2019. Biodiversity assessment of tropical shelf eukaryotic communities via pelagic eDNA metabarcoding. *Ecology and Evolution* 9:14341-14355.
- Baudelle G, Sonnic E, Alkan D, Quantin P-Y, Duhayon J-J. 2011. L'accueil des navires de plaisance dans la perspective d'une gestion intégrée des zones côtières.
- Bishop JDD, Wood C, Yunnice A, Griffiths C. 2015a. Unheralded arrivals: non-native sessile invertebrates in marinas on the English coast. *Aquatic Invasions* 10:249-264.
- Bishop JDD, Wood CA, Lévêque L, Yunnice ALE, Viard F. 2015b. Repeated rapid assessment surveys reveal contrasting trends in occupancy of marinas by non-indigenous species on opposite sides of the western English Channel. *Marine Pollution Bulletin* 95:699-706.
- Borcard D, Gillet F, Legendre P. 2018. *Numerical ecology with R*: Springer.
- Borrell YJ, Miralles L, Do Huu H, Mohammed-Geba K, Garcia-Vazquez E. 2017. DNA in a bottle—Rapid metabarcoding survey for early alerts of invasive species in ports. *PLOS ONE* 12:e0183347.
- Borrell YJ, Miralles L, Martínez-Marqués A, Semeraro A, Arias A, Carleos CE, García-Vázquez E. 2018. Metabarcoding and post-sampling strategies to discover non-indigenous species: A case study in the estuaries of the central south Bay of Biscay. *Journal for Nature Conservation* 42:67-74.
- Bouchemousse S, Lévêque L, Dubois G, Viard F. (Bouchemousse2016 co-authors). 2016. Co-occurrence and reproductive synchrony do not ensure hybridization between an alien tunicate and its interfertile native congener. *Evolutionary Ecology* 30:69-87.
- Bouchemousse S, Lévêque L, Viard F. 2017. Do settlement dynamics influence competitive interactions between an alien tunicate and its native congener? *Ecology and Evolution* 7:200-213.
- 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* 16:176-182.
- Brandt MI, Trouche B, Quintric L, Wincker P, Poulain J, Arnaud-Haond S. 2020. A flexible pipeline combining clustering and correction tools for prokaryotic and eukaryotic metabarcoding. *Peer Community in Ecology*:717355.
- Briski E, Chan FT, Darling JA, Lauringson V, MacIsaac HJ, Zhan A, Bailey SA. 2018. Beyond propagule pressure: importance of selection during the transport stage of biological invasions. *Frontiers in Ecology and the Environment* 16:345-353.
- Briski E, Ghabooli S, Bailey SA, MacIsaac HJ. 2012. Invasion risk posed by macroinvertebrates transported in ships' ballast tanks. *Biological Invasions* 14:1843-1850.
- Brown EA, Chain FJJ, Zhan A, MacIsaac HJ, Cristescu ME. 2016. Early detection of aquatic invaders using metabarcoding reveals a high number of non-indigenous species in Canadian ports. *Diversity and Distributions* 22:1045-1059.
- Bulleri F, Chapman MG. 2004. Intertidal assemblages on artificial and natural habitats in marinas on the north-west coast of Italy. *Marine Biology* 145:381-391.
- Bulleri F, Chapman MG. 2010. The introduction of coastal infrastructure as a driver of change in marine environments. *Journal of Applied Ecology* 47:26-35.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581.
- Capinha C, Essl F, Seebens H, Moser D, Pereira HM. 2015. The dispersal of alien species redefines biogeography in the Anthropocene. *Science* 348:1248-1251.

- Clarke Murray C, Pakhomov EA, Therriault TW. 2011. Recreational boating: a large unregulated vector transporting marine invasive species. *Diversity and Distributions* 17:1161-1172.
- Couton M, Comtet T, Le Cam S, Corre E, Viard F. 2019. Metabarcoding on planktonic larval stages: an efficient approach for detecting and investigating life cycle dynamics of benthic aliens. *Management of Biological Invasions* 10:657-689.
- Cristescu ME. 2014. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution* 29:566-571.
- Dafforn KA, Glasby TM, Johnston EL. 2009a. Links between estuarine condition and spatial distributions of marine invaders. *Diversity and Distributions* 15:807-821.
- Dafforn KA, Johnston EL, Glasby TM. 2009b. Shallow moving structures promote marine invader dominance. *Biofouling* 25:277-287.
- Dray S, Bauman D, Blanchet G, Borcard D, Clappe S, Guenard G, Jombart T, Larocque G, Legendre P, Madi N, et al. 2020. *adespatial: Multivariate Multiscale Spatial Analysis*. Version R package version 0.3-8.
- Duarte CM, Pitt KA, Lucas CH, Purcell JE, Uye S-i, Robinson K, Brotz L, Decker MB, Sutherland KR, Malej A, et al. 2013. Is global ocean sprawl a cause of jellyfish blooms? *Frontiers in Ecology and the Environment* 11:91-97.
- Duarte S, Vieira PE, Lavrador AS, Costa FO. 2020. Status and prospects of marine NIS detection and monitoring through (e)DNA metabarcoding. *bioRxiv:2020.2005.2025.114280*.
- Fletcher LM, Zaiko A, Atalah J, Richter I, Dufour CM, Pochon X, Wood SA, Hopkins GA. (Fletcher2017 co-authors). 2017. Bilge water as a vector for the spread of marine pests: a morphological, metabarcoding and experimental assessment. *Biological Invasions* 19:2851-2867.
- Fonseca V, Carvalho G, Sung W, Johnson HF, Power D, Neill S, Packer M, Blaxter ML, Lamshead PJ, Thomas W, et al. 2010. Second-generation environmental sequencing unmask marine metazoan biodiversity. *Nature Communications* 1:98.
- Gallon RK, Robuchon M, Leroy B, Le Gall L, Valero M, Feunteun E. 2014. Twenty years of observed and predicted changes in subtidal red seaweed assemblages along a biogeographical transition zone: inferring potential causes from environmental data. *Journal of Biogeography* 41:2293-2306.
- Glasby T. 1999. Interactive effects of shading and proximity to the seafloor on the development of subtidal epibiotic assemblages. *Marine Ecology-progress Series - MAR ECOL-PROGR SER* 190:113-124.
- Glasby TM, Connell SD, Holloway MG, Hewitt CL. (Glasby2007 co-authors). 2007. Nonindigenous biota on artificial structures: could habitat creation facilitate biological invasions? *Marine Biology* 151:887-895.
- Guzinski J, Ballenghien M, Daguin-Thiébaud C, Lévêque L, Viard F. 2018. Population genomics of the introduced and cultivated Pacific kelp *Undaria pinnatifida*: Marinas—not farms—drive regional connectivity and establishment in natural rocky reefs. *Evolutionary Applications* 11:1582-1597.
- Hudson J, Viard F, Roby C, Rius M. 2016. Anthropogenic transport of species across native ranges: unpredictable genetic and evolutionary consequences. *Biology Letters* 12:20160620.
- Hufbauer RA, Facon B, Ravigné V, Turgeon J, Foucaud J, Lee CE, Rey O, Estoup A. 2012. Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. *Evolutionary Applications* 5:89-101.
- Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, Nakamura A, Kitching R, Dolman PM, Woodcock P, Edwards FA, et al. 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters* 16:1245-1257.
- Johnston E, Dafforn K, Clark G, Rius M, Floerl O. 2017. How Anthropogenic Activities Affect the Establishment and Spread of Non-Indigenous Species Post-Arrival: An Annual Review. In. p. 389-419.
- Kelly RP, O'Donnell JL, Lowell NC, Shelton AO, Samhuri JF, Hennessey SM, Feist BE, Williams GD. 2016. Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ* 4:e2444.
- Lacoursière-Roussel A, Bock DG, Cristescu ME, Guichard F, Girard P, Legendre P, Mckindsey CW. 2012. Disentangling invasion processes in a dynamic shipping–boating network. *Molecular Ecology* 21:4227-4241.

- Lacoursière-Roussel A, Howland K, Normandeau E, Grey EK, Archambault P, Deiner K, Lodge DM, Hernandez C, Leduc N, Bernatchez L. 2018. eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. *Ecology and Evolution* 8:7763-7777.
- Lamb PD, Hunter E, Pinnegar JK, Creer S, Davies RG, Taylor MI. 2019. How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology* 28:420-430.
- Leclerc J-C, Viard F, González Sepúlveda E, Díaz C, Neira Hinojosa J, Pérez Araneda K, Silva F, Brante A. 2020. Habitat type drives the distribution of non-indigenous species in fouling communities regardless of associated maritime traffic. *Diversity and Distributions* 26:62-75.
- Leclerc J-C, Viard F, Sepúlveda E, Díaz C, Hinojosa J, Araneda K, Silva F, Brante A. 2018. Non-indigenous species contribute equally to biofouling communities in international vs local ports in the Biobío region, Chile. *Biofouling* 34:1-16.
- Legendre P, Gallagher ED. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129:271-280.
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ. (Leray2013 co-authors). 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology* 10:34.
- López-Legentil S, Legentil ML, Erwin PM, Turon X. (López-Legentil2015 co-authors). 2015. Harbor networks as introduction gateways: contrasting distribution patterns of native and introduced ascidians. *Biological Invasions* 17:1623-1638.
- Ma KC, Hawk HL, Goodwin C, Simard N. 2019. Morphological identification of two invading ascidians: new records of *Asciidiella aspersa* (Müller, 1776) from Nova Scotia and *Diplosoma listerianum* (Milne-Edwards, 1841) from New Brunswick and Quebec. *BioInvasions Record* 8.
- Magurran AE, Dornelas M, Moyes F, Gotelli NJ, McGill B. 2015. Rapid biotic homogenization of marine fish assemblages. *Nature Communications* 6:8405.
- Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* 3:e1420.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*; Vol 17, No 1: Next Generation Sequencing Data Analysis DO - 10.14806/ej.17.1.200.
- Martinez Arbizu P. 2020. pairwiseAdonis: Pairwise multilevel comparison using adonis. Version R package version 0.4.
- McDonald J. 2004. The invasive pest species *Ciona intestinalis* (Linnaeus, 1767) reported in a harbour in southern Western Australia. *Marine Pollution Bulletin* 49:868-870.
- Mineur F, Cook EJ, Minchin D, Bohn K, MacLeod A, Maggs CA. 2012. Changing coasts: marine aliens and artificial structures. In: Gibson RN, Atkinson RJA, Gordon JDM, Hughes RN, editors. *Oceanography and Marine Biology: an Annual Review, Volume 50*: CRC Press. p. 189-234.
- Molnar JL, Gamboa RL, Revenga C, Spalding MD. 2008. Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment* 6:485-492.
- Núñez AL, Katsanevakis S, Zenetos A, Cardoso AC. 2014. Gateways to alien invasions in the European seas. *Aquatic Invasions* 9:133-144.
- O'Shaughnessy KA, Hawkins SJ, Yunnice ALE, Hanley ME, Lunt P, Thompson RC, Firth LB. 2020. Occurrence and assemblage composition of intertidal non-native species may be influenced by shipping patterns and artificial structures. *Marine Pollution Bulletin* 154:111082.
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, et al. 2018. vegan: community ecology package. Version R package version 2.5-2.
- Olden JD. 2006. Biotic homogenization: a new research agenda for conservation biogeography. *Journal of Biogeography* 33:2027-2039.
- Piola RF, Johnston EL. 2009. Comparing differential tolerance of native and non-indigenous marine species to metal pollution using novel assay techniques. *Environmental Pollution* 157:2853-2864.
- Rey A, Basurko OC, Rodriguez-Ezpeleta N. 2020. Considerations for metabarcoding-based port biological baseline surveys aimed at marine nonindigenous species monitoring and risk assessments. *Ecology and Evolution* 10:2452-2465.

- Rivero NK, Dafforn KA, Coleman MA, Johnston EL. 2013. Environmental and ecological changes associated with a marina. *Biofouling* 29:803-815.
- Ruiz G, Freestone A, Fofonoff P, Simkanin C. 2009. Habitat Distribution and Heterogeneity in Marine Invasion Dynamics: the Importance of Hard Substrate and Artificial Structure. In. p. 321-332.
- Sardain A, Sardain E, Leung B. 2019. Global forecasts of shipping traffic and biological invasions to 2050. *Nature Sustainability* 2:274-282.
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, Winter M, Arianoutsou M, et al. 2017. No saturation in the accumulation of alien species worldwide. *Nature Communications* 8:14435.
- Shanks AL. 2009. Pelagic larval duration and dispersal distance revisited. *The Biological Bulletin* 216:373-385.
- Shanks AL, Grantham BA, Carr MH. 2003. Propagule dispersal distance and the size and spacing of marine reserves. *Ecological Applications* 13:159-169.
- Smart SM, Thompson K, Marrs RH, Duc MGL, Maskell LC, Firbank LG. 2006. Biotic homogenization and changes in species diversity across human-modified ecosystems. *Proc Biol Sci* 273:2659-2665.
- Soares MdO, Campos CC, Santos NMO, Barroso HdS, Mota EMT, Menezes MOBd, Rossi S, Garcia TM. 2018. Marine bioinvasions: Differences in tropical copepod communities between inside and outside a port. *Journal of Sea Research* 134:42-48.
- Sonnich E. 2008. La navigation de plaisance : une activité touristique "amphibie" entre espaces de pratiques et territoires de gestion.
- Spalding MD, Fox HE, Allen GR, Davidson N, Ferdaña ZA, Finlayson M, Halpern BS, Jorge MA, Lombana A, Lourie SA, et al. 2007. Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. *Bioscience* 57:573-583.
- Sylvester F, Kalaci O, Leung B, Lacoursière-Roussel A, Murray CC, Choi FM, Bravo MA, Therriault TW, MacIsaac HJ. 2011. Hull fouling as an invasion vector: can simple models explain a complex problem? *Journal of Applied Ecology* 48:415-423.
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* 21:2045-2050.
- Turon X, Casso M, Pascual M, Viard F. 2020. Looks can be deceiving: *Didemnum pseudovexillum* sp. nov. (Ascidacea) in European harbours. *Marine Biodiversity* 50:48.
- Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, et al. 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology* 25:929-942.
- von Ammon U, Wood SA, Laroche O, Zaiko A, Tait L, Lavery S, Inglis GJ, Pochon X. 2018. Combining morpho-taxonomy and metabarcoding enhances the detection of non-indigenous marine pests in biofouling communities. *Sci Rep* 8:16290.
- Wasson K, Zabin CJ, Bedinger L, Cristina Diaz M, Pearse JS. 2001. Biological invasions of estuaries without international shipping: the importance of intraregional transport. *Biological Conservation* 102:143-153.
- Zaiko A, Pochon X, Garcia-Vazquez E, Olenin S, Wood SA. 2018. Advantages and limitations of environmental DNA/RNA tools for marine biosecurity: management and surveillance of non-indigenous species. *Frontiers in Marine Science* 5.
- Zaiko A, Schimanski K, Pochon X, Hopkins GA, Goldstien S, Floerl O, Wood SA. 2016. Metabarcoding improves detection of eukaryotes from early biofouling communities: implications for pest monitoring and pathway management. *Biofouling* 32:671-684.
- Zhan A, Briski E, Bock DG, Ghabooli S, MacIsaac HJ. (Zhan2015 co-authors). 2015. Ascidiaceans as models for studying invasion success. *Marine Biology* 162:2449-2470.

# CHAPTER III:

Can metabarcoding detect hidden non-indigenous species in the wild?







## Preamble

In [Chapters I and II](#), my work focused on NIS communities in marinas. They are considered as introduction hotspots, and should be prioritized in surveys for detecting new arrivals (either new introduced species or spread of already reported ones). Marinas are, however, also known as bridge-heads for the spread of these introduced species into the wild. In addition, numerous studies have shown that, if NIS abundances are substantial in marinas (as shown in [Chapter II.1](#), and [II.2](#)), their numbers in natural habitats are relatively low. Whereas dispersal limitations, competitive interactions with resident species or/and habitat selection may explain these low numbers, our ability to detect NIS in the wild by traditional methods (e.g. diving surveys, scrapping etc.) may also be questioned.

In this last chapter, I thus used metabarcoding to look for NIS in natural communities using two approaches. First I studied zooplankton samples in one location over a time series of 22 months, in order to assess the potential for larval dispersal of NIS (both long and short dispersers). Then I examined specimens that have settled on experimental panels deployed in 6 natural subtidal locations at two seasons, and compared the fouling communities with those found in the neighboring marina. As in the preceding chapter, I had the opportunity to compare the outcomes of these two studies with results gathered by traditional methods (based on species identification using morphological or DNA barcoding).



Pictures of plankton sampling (left), settlement structures *in situ* (top right), and organisms fixed on one settlement plate after 10 weeks of immersion (bottom right).

## Are fouling communities in marinas, and particularly the non-indigenous fraction, singularly different from communities on natural rocky habitats?

An experimental study, at a bay scale, joining metabarcoding and morphological-based analyses of settlement panels.

Marjorie Couton<sup>1</sup>, Laurent Lévêque<sup>2</sup>, Claire Daguin-Thiébaud<sup>1</sup>, Thierry Comtet<sup>1</sup>, Frédérique Viard<sup>1\*</sup>

<sup>1</sup> Sorbonne université, CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

<sup>2</sup> Sorbonne université, CNRS, FR 2424, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

\* Correspondence author: viard@sb-roscoff.fr; +33 2 98 29 23 12

**Preliminary remark:** Following the structure of the other chapters, this part is built as a scientific paper. However, the results are based on preliminary analyses only, and will require further work. The study design was made by Thierry Comtet, Laurent Lévêque and Frédérique Viard. The field work was implemented by the Diving and Marine facilities at the Station Biologique of Roscoff. Morphological-based taxonomic identification was conducted by Laurent Lévêque. I produced environmental DNA and bulk DNA, as well as sequencing data, using the local facilities (GENOMER Platform) with advice and help from Claire Daguin-Thiébaud. I made the statistical analyses with the help of my PhD supervisors, and led the writing of this chapter.

## Abstract

Marinas provide suitable habitats for many hard-substrate sessile species, in particular non-indigenous species (NIS), which may then spread to nearby natural rocky habitats. Nevertheless, a low number and proportion of NIS have often been reported in natural environments as compared to marinas. This pattern might be due to particular features of marinas, which are selecting for singular species assemblages, and/or by high levels of competition with resident species in natural habitats. Our ability to detect non-indigenous species in the wild might also be questioned. To determine to which extent benthic sessile species assemblages, at early stages, differ between the two types of habitats, in particular regarding the NIS fraction, settlement panels were deployed in six localities (one marina, five natural rocky habitats) within a same bay, over two 10-week periods. The settled organisms were identified by both morphological-based and metabarcoding approaches. Metabarcoding on environmental DNA samples from the same localities was also carried out to obtain general biodiversity assessments for each locality. While very few taxa were found solely in the marina, it showed a very distinct species assemblage, a result expected from the literature. However, conversely to expectations, the proportion of NIS in the marina was similar or only slightly higher than in natural sites, even in distant sites less influenced by human-activities. This study shows that comparison of marinas and adjacent natural habitats benefit from metabarcoding, particularly when coupled with traditional surveys, as the species identified by the two methods only partially overlapped. This work also provides supports for further studies to confirm the observed pattern at larger scales and investigate the processes governing these patterns.

## Introduction

Marine coastal areas are severely impacted by human activities. The constant human population growth, the industrial increase, and the growing number of coastal cities (multiplied by 4.5 since the mid-20<sup>th</sup> century; Barragán and de Andrés 2015) have intensified coastal degradation, especially at the ecosystem level (Airoldi and Beck 2007; Firth, et al. 2016). The proliferation of artificial structures, such as seawalls, artificial reefs, aquaculture facilities, jetties or floating pontoons in marinas (known as ocean sprawl) has led to numerous impacts on natural ecosystem structure and functioning (e.g. Firth, et al. 2016; Bishop, et al. 2017). In particular, comparisons

between urban and non-urban coastal areas have shown negative impacts on natural populations, such as decreased population density (e.g. *Mytilus galloprovincialis*; Veiga, et al. 2020), changes in ecological connectivity (for a review, see Bishop et al. 2017), and establishment of non-indigenous species (NIS) (Dafforn 2017).

Shipping being the major vector for marine biological introductions, ports and marinas are considered as invasion hubs, sustaining a high colonization and propagule pressure (Molnar, et al. 2008; Nuñez, et al. 2014). Hull fouling is responsible for the majority of NIS transport in marinas through recreational boating which is not as regulated as international shipping (Clarke Murray, et al. 2011; Ulman, et al. 2019). Selection during transport might also increase the likelihood of NIS establishment on artificial hard substrates found in ports and marinas (Briski, et al. 2018) as well as commonly shared abiotic factors by these anthropogenic environments, such as high turbidity, shading, pollution, and reduced water flow (Glasby 1999; Rivero, et al. 2013). Finally, the availability of a wide variety of artificial substrates in marinas offers settling possibilities to newly arriving species where they can subsequently become established and reproduce. For these reason, marinas can be considered as “stepping stones” for introduced species (Bishop, et al. 2017).

Artificial habitats can both enhance NIS spread at a regional scale, combined to leisure boating, but also constitute bridge-heads for dispersal into surrounding natural habitats (Glasby, et al. 2007), which can lead to a number of ecological and economic consequences (Seebens, et al. 2017). NIS can alter ecosystem services such as food provisioning (fisheries, aquaculture), can be major drivers of ecological changes, through predation or competition with native species (Katsanevakis, et al. 2014; Ojaveer, et al. 2015), and of evolutionary change, through hybridization with native species (Ayres, et al. 1999; Viard, et al. 2020). It is therefore of the utmost importance to evaluate the potential spread of NIS from marinas to the surrounding coastal areas and to examine to which extent they can successfully establish outside of their entry point (e.g. Simkanin, et al. 2012).

Communities found within ports and marinas are not similar to those observed in surrounding natural habitats (Connell 2001; Soares, et al. 2018). This is particularly true for benthic sessile assemblages fixed on hard substrates, which usually harbor a higher proportion of NIS (Bulleri and Chapman 2004; Wasson, et al. 2005; Ruiz, et al. 2009; Airoldi, et al. 2015), or higher abundances of given NIS (Simkanin, et al. 2012), than natural rocky reefs. Several hypotheses could explain this phenomenon, such as the different quality and complexity of artificial substrates or the particular abiotic

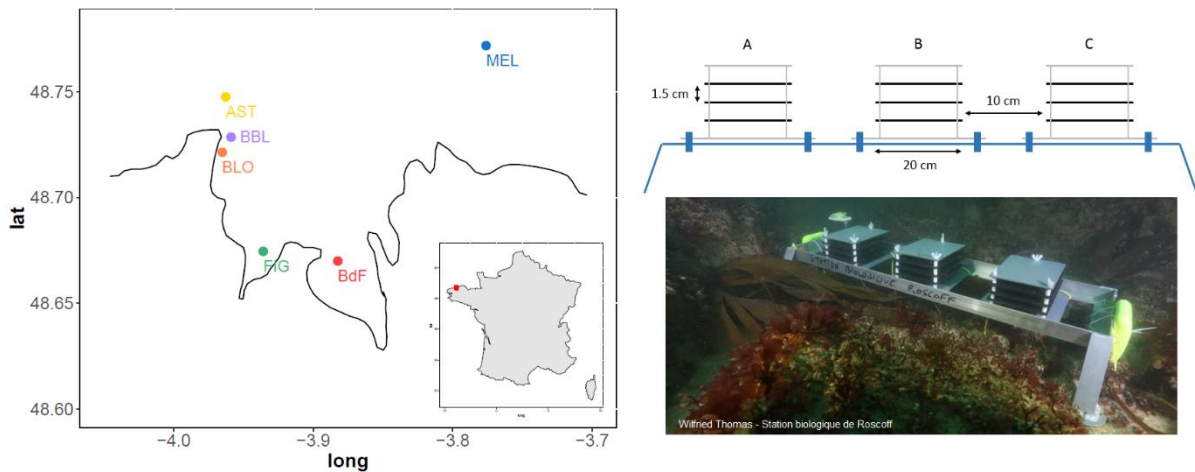
and biotic conditions of marinas (see [Chapter II.2](#)). However, this difference might also be due, at least partly, to limitations of the methods that are usually implemented to survey subtidal rocky reefs. They are, indeed, complex habitats made of numerous microhabitats, such as cracks and crevices, in which many species can hide, and are thus difficult to sample or observe, especially when using *in situ* observation and time-limited diving censuses. For example, O'Shaughnessy, et al. (2020) showed that the number of NIS identified in both natural and artificial habitats may differ depending on the observation technique used (quadrat sampling vs. rapid assessment survey). Other methods such as the deployment of settlement panels in natural reefs, with morphological identification of settlers, can also be hampered by the inherent difficulty to identify NIS, particularly when they are morphologically similar to native species (i.e. (pseudo) cryptic species), and at young stages (e.g. early settlers). NIS might thus actually be more abundant in natural settings than previously reported with these approaches.

Our objectives, here, were first to compare the benthic sessile communities in a marina and in neighboring natural habitats. Then, focusing on NIS, we evaluated their dispersal and settlement potential in natural environments. Based on current knowledge and monitoring made in the study bay and region, we expected to report a reduced number of NIS in the wild as compared to the marina, presumably those with highest dispersal abilities. High-throughput sequencing techniques have been increasingly used to detect and study introduced species and have been proposed as a valuable addition to traditional methods (Comtet, et al. 2015; Darling, et al. 2017; Zaiko, et al. 2018). To achieve our objectives, we used a combined approach of morphology-based assessment and metabarcoding to compare early establishment of benthic sessile communities present on settlement plates deployed in natural rocky reefs and in one marina from the same bay. In addition, environmental DNA (eDNA) was obtained from seawater samples collected near the settlement panels, at the same localities, for a more global assessment of the species found in these diverse habitats. We hypothesized that more NIS would be detected by metabarcoding approaches than by the traditional morphology-based method, in particular when examining community at early stages of recruitment (ca. 2-3 months).

## Materials and methods

### a. Survey design

Sampling was performed in six locations in the bay of Morlaix (Brittany, France; Fig. 1). The sites included one marina (BLO) and a locality close by (BBL), two points in the inner part of the bay (FIG and BdF) at each river mouth, and two points in the outer part of the bay (AST and MEL). These locations were chosen because of the presence of natural rocky reefs suited for the recruitment of benthic hard bottom organisms. The recruitment was monitored using standardized structures inspired by the Autonomous Reef Monitoring Structures (ARMS) described in Plaisance, et al. (2011) with modifications. Each structure grouped three piles composed of five 15x17 cm plates (Fig. 1), fixed to an aluminium structure. Only the three middle plates, made in Correx® plastic, shown to be effective in similar experiments (e.g. Bouchemousse 2015; Leclerc and Viard 2018), were examined. The settlement structures were immersed for approximately 10 weeks during two periods in 2018, from April to July and from August to November (Table 1). All structures were placed at the bottom, in a natural rocky reef habitat, except the one inside the marina which was suspended below a floating pontoon.



**Figure 1** Location of each sampling point (left), and schematic and picture (from BBL site) of the experimental settlement structure (right). Photo credit: Wilfried Thomas – Station biologique de Roscoff. Location codes are as follows: MEL: Méloine, AST: Astan, BBL: Basse Blocon, BLO: Blocon (Roscoff’s marina), FIG: Figuier, BdF: Barre des Flots.

Following the retrieval of the structure by divers, each pile was placed in separate plastic containers filled with marine water collected on site to be transported back to the lab. Upon arrival each pile was placed inside an aquarium with a constant flow of filtered water (5  $\mu\text{m}$ ) at 15°C to keep the organisms alive until further processing. The pile from the middle of each structure (label 'B'; Fig. 1) was devoted to morphological identification of organisms whereas piles A and C were scraped for DNA extraction and metabarcoding. All aquaria, plastic containers and materials used were immersed in 12.5 % commercial bleach (0.65 % hypochlorite) for at least 30 min and rinsed before field work.

**Table 1** Details of the six sampling sites within and outside the Bay of Morlaix. The first deployment period will further be referred to as "Summer" whereas the second period will be referred to as "Fall".

Site	Longitude	Latitude	Depth	Summer		Fall	
				Deployment	Retrieval	Deployment	Retrieval
BLO	-3.966	48.718	< 1 m	2018/04/25	2018/07/12	2018/08/28	2018/11/08
BBL	-3.959	48.729	13.3 m	2018/04/25	2018/07/11	2018/08/30	2018/11/06
AST	-3.963	48.747	21.5 m	2018/04/24	2018/07/06	2018/08/22	2018/11/14
MEL	-3.777	48.775	22.1 m	2018/04/23	2018/07/04	2018/08/29	2018/11/15
FIG	-3.936	48.674	3.8 m	2018/04/24	2018/07/02	2018/08/29	2018/11/12
BdF	-3.882	48.670	13.1 m	2018/04/26	2018/07/09	2018/08/27	2018/11/05

In addition to settlement plates, seawater was collected by divers at each site to extract environmental DNA. Three 3-L plastic containers were filled with water from the exact location and depth nearby the structures, both during deployment and retrieval, leading to a total of 72 samples. The water was stored at 4°C until filtration, which was performed a couple of hours later, on a Millipore® Sterivex™ filter unit (0.22  $\mu\text{m}$ ) using a Masterflex® L/S® economy drive peristaltic pump. To standardize volume across sites, 2 L were processed for each container. After filtration, 2 mL of lysis buffer (sucrose 0.75 M, Tris 0.05 M pH 8, EDTA 0.04 M) were added before storing the Sterivex unit at -20°C until DNA extraction.

## b. Processing of settlement plates

Organisms attached to the three Correx® plates from pile B (Fig. 1) were identified under a dissecting microscope to the lowest taxonomic level possible. Only



the bottom facing part of a plate was observed. A grid of 11 x 11 squares was placed on top of the plate and the occurrence of each taxon in each square was recorded. Squares from the outer part of the grid were not included in the occurrence records leaving a total of 81 squares (9x9; surface of 144 cm<sup>2</sup>) for the observations. Only sessile organisms were considered. The abundance for every taxon identified thus corresponded to the number of squares in which this taxon was observed. An additional abundance of 0.5 was attributed to taxa which were only observed in squares from the outer part of the plate.

The bottom facing part of the three Correx® plates from piles A and C were scraped to collect all organisms in a 2-L glass beaker. They were subsequently mixed with an immersion blender (ErgoMixx MSM66020, BOSCH®) until homogenization, and filtered with a 48-µm mesh size nylon filter. The solid fraction (> 48 µm) was stored in a 50 mL tube filled with a DMSO solution (DMSO 3.2M, EDTA 0.25M pH 8, NaCl saturated) at -20°C until DNA extraction. Before the field work, all consumables not sold as DNA-free (including the blender axis) were immersed in 12.5 % commercial bleach (0.65 % hypochlorite) for at least 30 min, rinsed with ultrapure water and placed under UV light for at least 15 min.

### c. DNA extraction, amplification and sequencing

At every step, precautions were taken to avoid contamination of DNA extracts with external DNA, as detailed above for consumables not sold as DNA-free. All equipment and bench surfaces were also DNA decontaminated before use.

DNA extraction from the solid fraction (settlement plates) was performed using the NucleoSpin® Soil kit (Macherey-Nagel). Prior to lysis, the samples were centrifuged and the excess DMSO was removed from the tube. The bead beating step from the manufacturer's protocol was replaced by a two-step lysis procedure. First, approximately 300 mg of wet material were placed in a tube with 700 µL of SL2 lysis buffer and an additional 50 µL of proteinase K (20 mg.mL<sup>-1</sup>). The tubes were placed in an Eppendorf ThermoMixer® at 56°C for three hours. At the end of the first lysis process, the samples were centrifuged, the supernatant was placed in a new tube with 150 µL of SL3 buffer and stored at 4°C overnight. The remaining solid fraction was resuspended with 700 µL of SL2 lysis buffer and 50 µL of proteinase K (20 mg.mL<sup>-1</sup>), and was placed in an Eppendorf ThermoMixer® at 56°C overnight. The following day, the next steps of the extraction protocol were performed following the

manufacturer's protocol with the two lysis products from the same sample processed separately. Elution was performed in 40  $\mu$ L of buffer pre-heated at 70 °C, placed twice on the column for a better yield. DNA solutions extracted from the same sample were then pooled resulting in a total volume after extraction of 80  $\mu$ L per sample. DNA was quantified by absorbance in a Spark TECAN reader using a NanoQuant Plate™, and samples were stored at -20°C.

DNA from the water filtered on Sterivex® units was extracted using a custom protocol based on the NucleoSnap® Finisher Midi extraction kit (Macherey-Nagel) which is detailed in [chapter II.1](#).

All library preparation steps were performed in the respect of strict rules for preventing contamination of samples, such as UV irradiation of all tips, plates and tubes and use of filter tips. Library preparation was performed using a dual-barcoded, dual-indexed two-step PCR procedure detailed in [chapter II.1](#) with five tagged PCR replicates for each sample. DNA was amplified with three primer pairs, each targeting a different marker: i) a 365-bp COI portion amplified using primers designed by Leray, et al. (2013), ii) a 389-bp to 489bp portion of the 18S rRNA V1-V2 region amplified using the forward primer designed by Fonseca, et al. (2010) and the reverse modified primer designed by Sinniger, et al. (2016), and iii) a 159-bp portion of the 16S rRNA amplified using primers designed by Kelly, et al. (2016). These three markers were chosen because they were designed to target metazoans and complement each other in terms of taxonomic resolution and amplification biases. Negative controls were included at each step (2 for the scraping, 6 for the DNA extraction and 6 for the PCR). Sequencing was performed in six different runs on an Illumina® MiSeq sequencer. The COI and 18S markers were sequenced using a 600 cycle v3 protocol with two index reads and the 16S marker was sequenced using a 300 cycle v2 protocol with two index reads.

#### d. Denoising and taxonomic assignment

Sequencing data were produced for each of the three markers, and for both types of samples (scrapped material and water). DNA extracted from water samples will further be referred to as "Water eDNA" or "W" and DNA extracted from scraped material will be referred to as "Plate bulkDNA" or "P". These datasets were processed similarly, as detailed in [Chapter II.1](#), using CUTADAPT v-2.8 (Martin 2011) to remove primers and tags, and DADA2 v-1.13.1 (Callahan, et al. 2016) to produce a set of

amplicon sequence variants (ASVs). The ASV table was then filtered for index-jump correction and only ASVs found in at least two out of five PCR replicates per sample were retained (Table S1).

A first taxonomic assignment was performed in order to evaluate the proportion of reads assigned to the different kingdoms and metazoan phyla for each marker. Each ASV was aligned against references retrieved from the GenBank nt database using the `ecotag` command from the `OBITOOLS v-1.2.11` package (Boyer, et al. 2016) with no minimum identity threshold. Then an assignment at a lower taxonomic level was performed by aligning all ASVs against a restricted reference database ([Chapter II.1](#)) using the `BLAST®` command line tool (Altschul, et al. 1990). This reference database focused on ten metazoan phyla, most of them commonly observed in the scraped samples, and/or found on or near floating pontoons: Annelida, Arthropoda (only crustaceans), Bryozoa, Chordata (only fish and tunicates), Cnidaria, Echinodermata, Mollusca, Nemertea, Platyhelminthes, and Porifera. Only alignments covering 99% of the subject or query sequence and above the chosen identity threshold for each marker (18S: 99%; COI: 92%; 16S: 97%; see [chapter II.1](#)) were considered. If one ASV matched with several references, it was assigned to the one with the highest identity percentage. If two alignments with different references had the same identity percentage, the ASV was assigned to the lowest common taxonomic rank. All assignments at a rank higher than the family were classified as “unassigned”.

### e. Diversity analyses

Three datasets were analysed. The first one, labelled “Morphology”, is composed of abundances of taxa collected from settlement plates. Then, for each type of DNA (water eDNA and plate bulkDNA), a presence-absence dataset from all three markers combined was produced with assigned taxa corresponding to benthic sessile organisms fixed on hard substrates (see Table S5 from [chapter II.2](#) for a detailed list of included taxa). To avoid redundancy, assignments to the family were discarded if any species or genus was already present for this family, and assignments to the genus were discarded if several species of the same genus were already present. If only one species of this genus was already identified, both assignments were pooled to the genus level.

For taxa identified at the species level with metabarcoding, that were already reported in marinas studied in [Chapter II](#) (marinas at the regional scale), their “native” *versus* “NIS or cryptogenic species” status was assigned using the list previously described<sup>1</sup>. The status was also assigned for all species found in the morphology-based dataset, in all sites.

Alpha-diversity was estimated based on taxon richness, as only presence-absence data were used for the metabarcoding datasets. Community compositions were compared using a Principal Component Analyses (PCA) as suggested in Borcard, et al. (2018). For each dataset the Hellinger transformation was applied prior to analyses. All analyses were conducted using the VEGAN-2.5.2 R package.

## Results

### a. Identification of taxa across methods

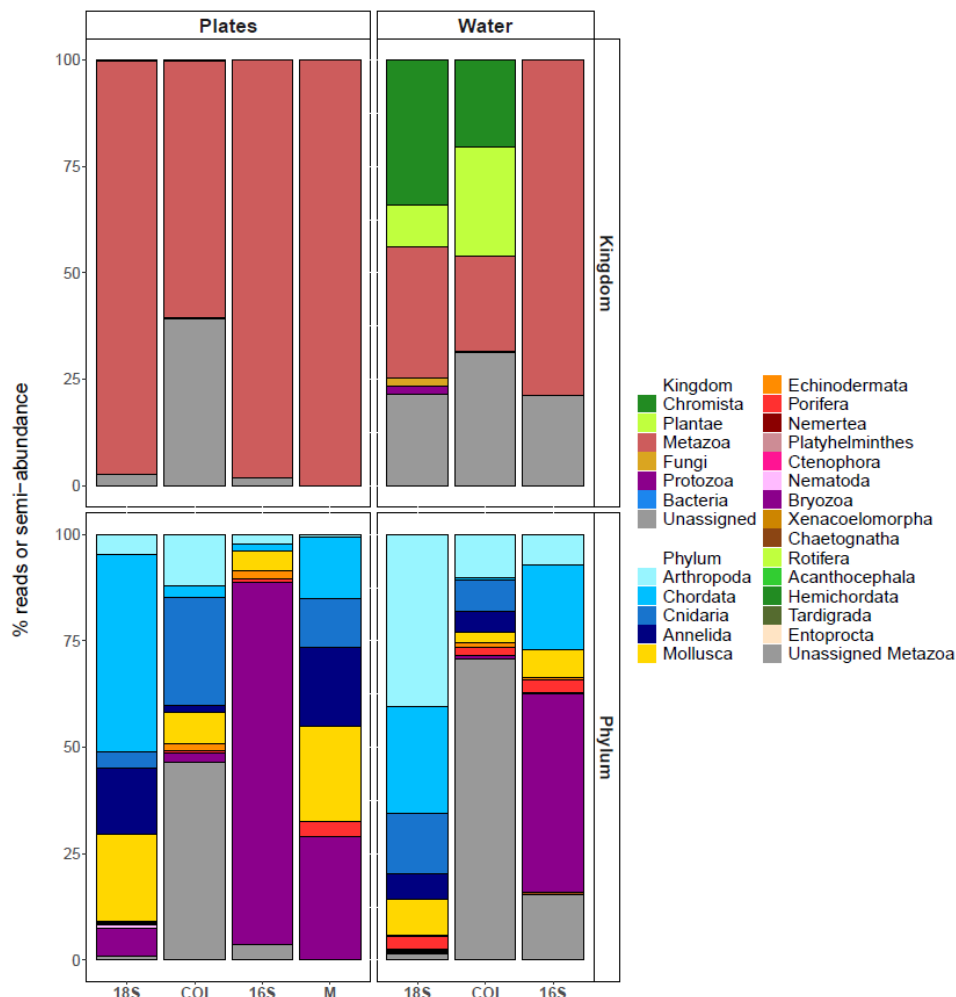
Laboratory observations of settlement plates allowed the identification of 70 taxa, based on morphological criteria, across all study sites (Table S2). More than half (n=36) were bryozoans. Taxonomic identifications were mainly done at the species level (n=42), whereas 15 were assigned to the genus level, and 7 to the family level. Remaining assignments were done at higher taxonomic levels (1 order, 4 classes, and 1 phylum). Among the 42 species identified, 8 were NIS and 7 were cryptogenic species.

High-throughput sequencing of water eDNA yielded 16,723,338 reads, 20,252,982 reads, and 18,878,245 reads for the 18S, COI, and 16S markers, respectively. After processing, this resulted in 3524, 4029, and 503 unique ASVs. The sequencing of plate bulkDNA yielded values in the same range for reads (17,816,314, 16,284,040, and 21,973,771 reads for 18S, COI, and 16S, respectively), but resulted in

---

<sup>1</sup> As acknowledged in the “Preliminary remarks” section of this chapter, some of the analyses done here are preliminary. In particular, the analyses made to examine the effect of the species status (native *vs.* NIS) were done on a subset of species, corresponding to those found in marinas studied in [Chapter II](#). These analyses will be expanded later by determining the status of the 90 remaining species not found previously in marinas (27 and 63 species found in Plates bulkDNA and water eDNA datasets, respectively; see Results section). Note however that our primary aim was to check the presence of species reported in the BLO marina in adjacent natural habitats: the native/NIS status was determined for all of them based on results of the previous chapter.

a lower number of unique ASVs (i.e. 1560, 3001, and 441 unique ASVs). After the filtering steps, none of the 14 negative controls of scraping (2), extraction (6) and PCR (6) contained any reads for 18S and COI. For 16S, one PCR negative control from the water eDNA dataset contained 43,140 reads assigned to a human reference and another 1951 reads unassigned (blast best hit) to an insect of the family Chironomidae (96.5% identity, 100% query cover). One extraction control and one scraping control from the plate bulkDNA dataset also had 24,731 and 28,823 reads, respectively, assigned to a human reference. Finally, the second scraping control, from Fall, still exhibited 2599 reads assigned to *Bugulina fulva* (Ryland, 1960), the results concerning the presence of this particular species thus need to be interpreted with caution.



**Figure 2** Percentage of reads assigned to a kingdom (top) or a metazoan phylum (bottom), over all samples, from the plate bulkDNA or the water eDNA datasets, for each of the three markers used in this study. An extra column in the “plates” panels displays the abundance percentage of taxa assigned to each of the kingdoms and metazoan phyla listed for the morphology dataset (M).

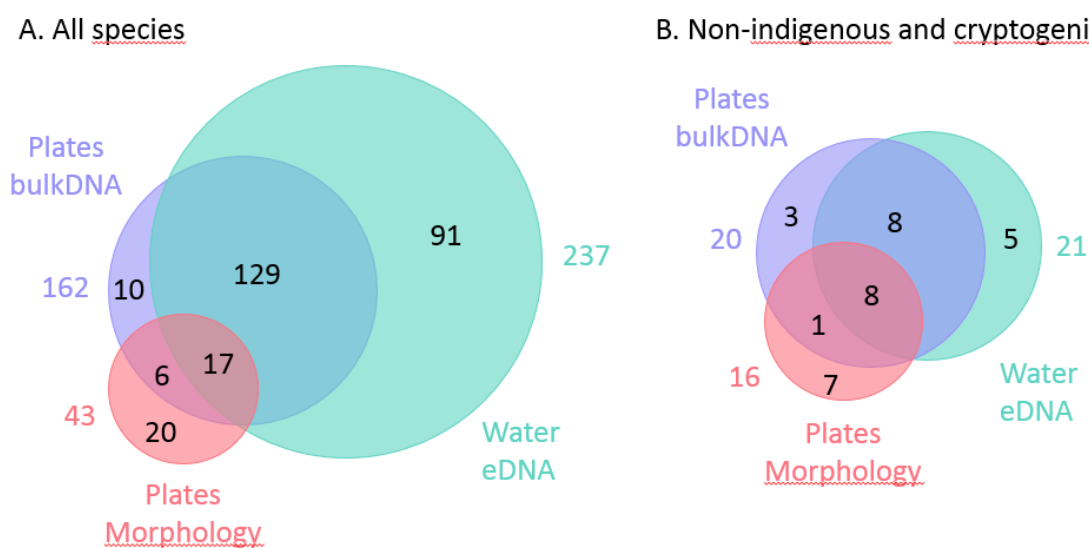
When aligned against the GenBank nt database using the ecotag tool, only a small proportion of reads were associated with metazoan taxa for 18S (31%) and COI (22%) for water eDNA, contrasting with 16S (79%) (Fig. 2). For the Plate bulkDNA datasets, however, the majority of reads were attributed to metazoan taxa for all markers (18S: 97%; COI: 60%; 16S: 98%). With COI, a high proportion of reads were unassigned at the kingdom level (W: 31%; P: 39%) or at the phylum level within metazoans (W: 71%; P: 46%). The most abundant metazoan phyla identified with the three markers for the plate bulkDNA datasets were different, thus showing their complementarity. 18S amplified preferentially Chordata (46%), Mollusca (20%) and Annelida (16%) whereas COI amplified mostly Cnidaria (25%) and 16S Bryozoa (85%). For the water eDNA datasets, both Arthropoda and Cnidaria were preferentially amplified by 18S (40% and 14%) and COI (10% and 7%). Chordata were abundant in both 18S (25%) and 16S (20%) but the former amplified mainly tunicates whereas the latter selected mainly fish. Finally, Bryozoa were again the most abundant phyla with 16S (47%).

When compared against the database restricted to benthic sessile taxa from 10 metazoan phyla, 8% (5%), 9% (3%), and 22% (57%) of ASVs (reads) were assigned to a species, genus, or family for the water eDNA datasets with 18S, COI, and 16S, respectively (Table 2). For the plate bulkDNA datasets, the percentages of assigned ASVs were higher for 18S and COI (14% (57%), 14% (19%) of ASVs (reads), respectively) but similar for 16S (19% (23%) of ASVs (reads)). The number of identified taxa was consistently higher within the water eDNA datasets than the plates bulkDNA datasets, with 16S identifying less taxa than the two other markers.

**Table 2.** Number of hard-bottom sessile taxa identified among ten metazoan phyla whatever the taxonomic level considered. The number of species, genera, and families assigned are also detailed. The number and proportion (in parentheses) of ASVs and reads assigned to one of the reference sequence, for each marker and each dataset, are indicated.

	Water eDNA			Plates bulkDNA		
	18S	COI	16S	18S	COI	16S
<b>Taxa</b>	195	130	43	112	105	38
<b>Species</b>	144	104	30	84	88	27
<b>Genus</b>	150	101	38	98	86	35
<b>Family</b>	92	73	32	63	55	28
<b>ASVs</b>	267 (8%)	348 (9%)	110 (22%)	213 (14%)	417 (14%)	85 (19%)
<b>reads</b>	337,034 (5%)	273,910 (3%)	5,149,589 (57%)	3,055,993 (57%)	883,117 (19%)	2,099,161 (23 %)

After combining datasets from all three markers, a total of 261 and 174 taxa were identified for water eDNA and plate bulkDNA respectively. Among these, 214 and 143 were species, 43 and 28 were genera, and 4 and 3 were families for water eDNA and plate bulkDNA, respectively. Among the species identified, the introduction status could be assigned for 151 and 116 of them for each dataset, respectively, based on the work presented in [Chapter II](#). Twenty-one and 16 were NIS, 7 and 7 were cryptogenic species, and 76 and 62 were native species for water eDNA and plates bulkDNA, respectively.



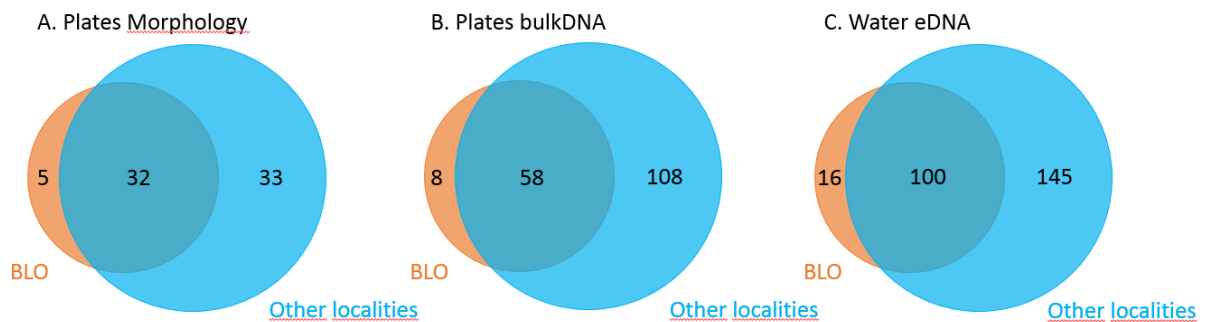
**Figure 3** Number of species (A) and non-indigenous or cryptogenic species (B) recovered by the three methods used in this study. Note that “native” vs. “NIS and cryptogenic species” was established only for the species previously observed in marinas from Brittany ([Chapter II.1](#), see Materials & Methods).

When comparing species detected with all three methods, 90% of species identified within the plates bulkDNA dataset were also recovered with water eDNA, whereas 38% of species detected with water eDNA were exclusively found in this dataset (Fig. 3A). Importantly, almost half (47%) of the species identified within the plate morphological identification dataset were not recovered within the two metabarcoding datasets, whereas 40% were recovered by the two metabarcoding methods. No species was shared only by “plate morphology” and “water eDNA”. Thirty-two NIS and cryptogenic species were identified across methods, with more than half (N=17; 53%) recovered by at least two methods, but only 8 (25%) recovered by all three methods (Fig. 3B). Similar to the “all species” analysis, a large proportion

(44%) of the non-indigenous and cryptogenic species identified in the “plate morphology” dataset were not observed within the metabarcoding datasets. They represent 22% of all NIS and cryptogenic species identified across all methods.

## b. Taxa found only in the marina

Many of the taxa identified (including those identified at species-, genus- or family-level) were observed both within and outside the marina for all three methods, although with differences among them (M: 46%, P: 57%, and W: 60%; Fig. 4).



**Figure 4** Number of taxa observed in the marina (BLO) or in other parts of the bay (Other localities) for each of the three methods used in this study.

Considering each method separately, 25 taxa were found only in the marina with at least one approach (Table 3). But almost half of them (10) were found in other localities with a different method. For instance, the ascidian *Clavelina lepadiformis* was only found in BLO with Plate bulkDNA but observed also in FIG with the Morphology dataset. Two NIS, the bryozoans *Bugulina fulva* and *Watersipora subatra*, were found in other localities, when combining results across methods. Thus, taking into account the different methods together, only 15 taxa were found exclusively in the marina. The most represented phyla in these species were Arthropoda (6) and Annelida (6). Among them, seven were classified as NIS (some of them being putative novel NIS for the Brittany region or false positives, see footnote in Table 3) and one as cryptogenic species.



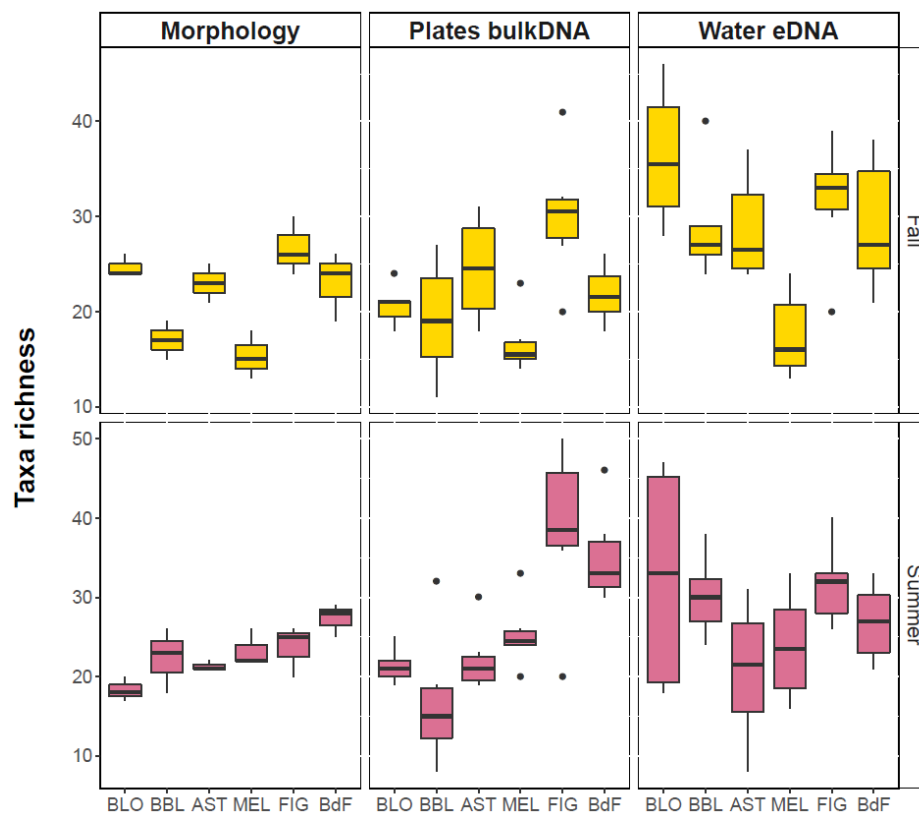
**Table 3** List of taxa observed solely within the marina with at least one of the three methods. Species in bold are found only in the marina even when combining all three methods.

Phylum	Family	Taxon	Status <sup>(1)</sup>	Morphology	Plates bulkDNA	Water eDNA
Annelida	Sabellidae	<b><i>Pseudopotamilla sp.</i></b>	undetermined	no	no	BLO
Annelida	Serpulidae	<b><i>Neodexiospira alveolata</i></b>	Novel NIS?	no	no	BLO
Annelida	Serpulidae	<b><i>Vermiliopsis striaticeps</i></b>	Novel NIS?	no	BLO	BLO
Annelida	Spionidae	<b><i>Boccardia proboscidea</i></b>	NIS	no	no	BLO
Annelida	Spionidae	<b><i>Dipolydora bidentata</i></b>	Novel NIS?	no	no	BLO
Annelida	Spionidae	<i>Malacoceros fuliginosus</i>	Native	no	BLO	BLO, BBL, FIG, BdF
Arthropoda	Chthamalidae	<i>Chthamalus montagui</i>	Native	no	BLO	BLO, BBL, AST, MEL, FIG
Arthropoda	Caprellidae	<b><i>Caprella acanthifera</i></b>	Native	no	no	BLO
Arthropoda	Gammaridae	<b><i>Gammarus locusta</i></b>	Native	no	BLO	no
Arthropoda	Idoteidae	<b><i>Idotea balthica</i></b>	Native	no	no	BLO
Arthropoda	Ischyroceridae	<b><i>Jassa slatteryi</i></b>	Cryptogenic	no	BLO	no
Arthropoda	Nuuanuidae	<b><i>Gammarella fucicola</i></b>	Native	no	no	BLO
Bryozoa	Bugulidae	<b><i>Bugula neritina</i></b>	NIS	BLO	BLO	no
Bryozoa	Bugulidae	<i>Bugulina fulva</i>	Cryptogenic	BLO, FIG	BLO	no
Bryozoa	Bugulidae	<b><i>Bugulina stolonifera</i></b>	NIS	BLO	no	no
Bryozoa	Calloporidae	<i>Callopora dumerilii</i>	Native	no	all localities	BLO
Bryozoa	Cryptosulidae	<i>Cryptosula pallasiana</i>	Native	BLO	BLO, BBL, MEL	BLO, BBL, MEL, FIG
Bryozoa	Watersiporidae	<i>Watersipora subatra</i>	NIS	BLO	BLO, BBL	BLO, MEL
Chordata	Clavelinidae	<i>Clavelina lepadiformis</i>	Native	BLO, FIG	BLO	no
Chordata	Molgulidae	<i>Molgula sp.</i>	undetermined	BLO	BLO, AST, MEL, FIG, BdF	all localities
Cnidaria	Campanularidae	<i>Obelia dichotoma</i>	Native	no	BLO, BBL, BdF	BLO
Cnidaria	Pandeidae	<i>Amphinema dinema</i>	Native	no	BLO and BdF	BLO
Porifera	Darwinellidae	<b><i>Aplysilla rubra</i></b>	Native	no	no	BLO
Porifera	Hymedesmiidae	<b><i>Phorbas fictitius</i></b>	Native	no	no	BLO
Porifera	Oscarellidae	<b><i>Oscarella microlobata</i></b>	Novel NIS?	no	no	BLO

(1) regarding the introduction status, “novel NIS?” refers to species that may be novel for the study region (Western English Channel) or false positives, whereas “NIS” refers to species for which literature showed its introduction in our region. Briefly: Native to North Japan Sea, *N. alveolata* has previously been reported in European seas (on marine litter along the Cantabrian coast; Miralles, et al. 2018). Similarly, *V. striaticeps* has been reported as a first record for the Netherlands in 2011 (Ligthart, et al. 2011). These two species might thus be true novel species for the study region. There are no previous records in cold-temperate waters for *Oscarella microlobata*, which has been described, and reported only in the Mediterranean Sea, making unlikely this introduction. Similarly, *Dipolydora bidentata* has only been reported in the N.E. Pacific, and not as an introduced species (Abe, et al. 2019). These two species are thus likely false positives. The other NIS in this table have all been previously reported in Brittany (e.g. *B. proboscidea* native to the north Pacific is introduced in many places around the world including in Brittany; Radashevsky, et al. 2019, as well as the two Bugulidae species; Ryland, et al. 2011).

### c. Alpha- and beta-diversity at the bay scale

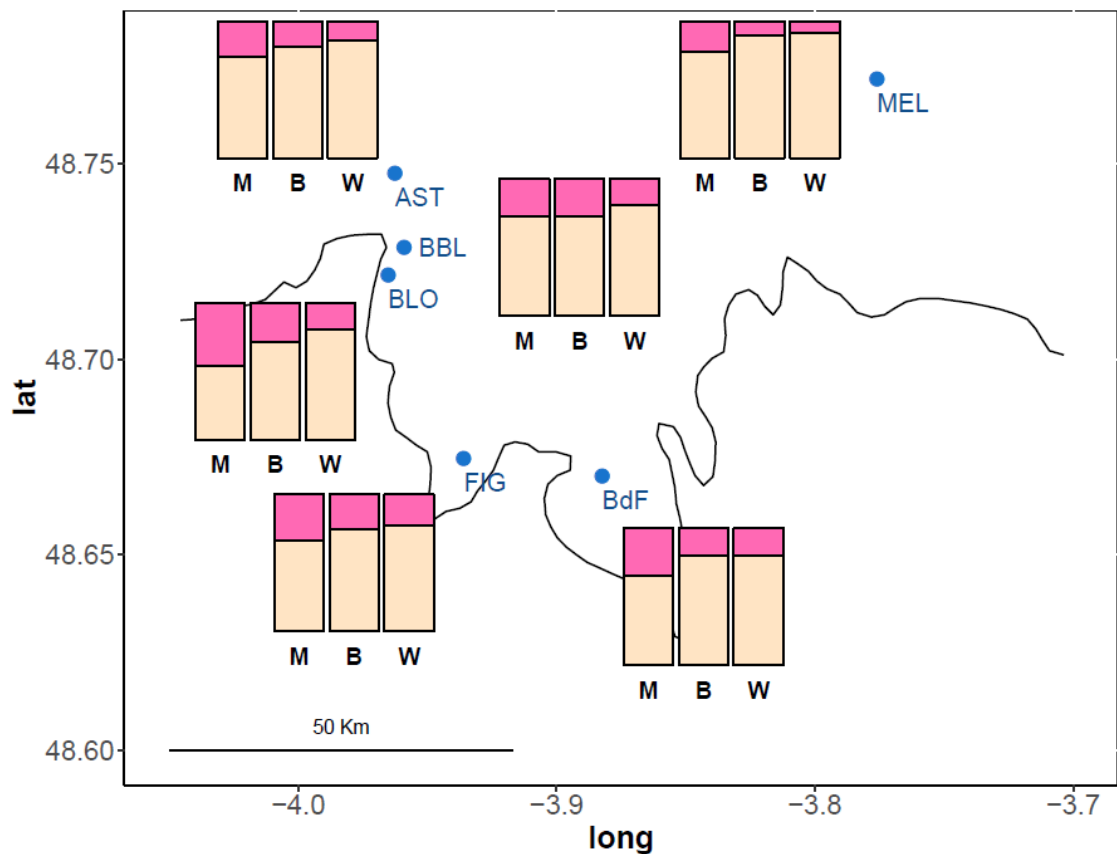
No clear pattern in richness was observed in the fall, whatever the method, except the consistent lower diversity in MEL (Fig. 5). In summer, the two localities in the most inner part of the bay, nearby river mouths, (FIG and BdF) tended to have a higher richness than the two localities situated in the outer part of the bay (AST and MEL) with all three methods, and particularly with plate bulkDNA (Fig. 5). The marina, BLO, showed high richness both in fall and summer especially with eDNA from water samples.



**Figure 5** Benthic taxa richness distribution for each locality at the two seasons of retrieval of the settlement structures.

The proportion of NIS and cryptogenic species, for both deployment seasons combined, varied across locations (Fig. 6). The lowest proportion was observed in MEL, the outermost site, whatever the method (23, 10, 8 %, for M, P and W, respectively; Table 4). The highest proportion for both approaches applied on

settlement plates was observed in the marina BLO, reaching 46% with morphology. For Water eDNA, however, BLO exhibited only 19% of NIS, whereas 23% were recorded in FIG where the highest value was observed). Nevertheless, none of the Chi-squared pairwise comparisons of all location for each method displayed a significant difference, including pairwise comparisons between MEL and BLO (Table S3-S5).

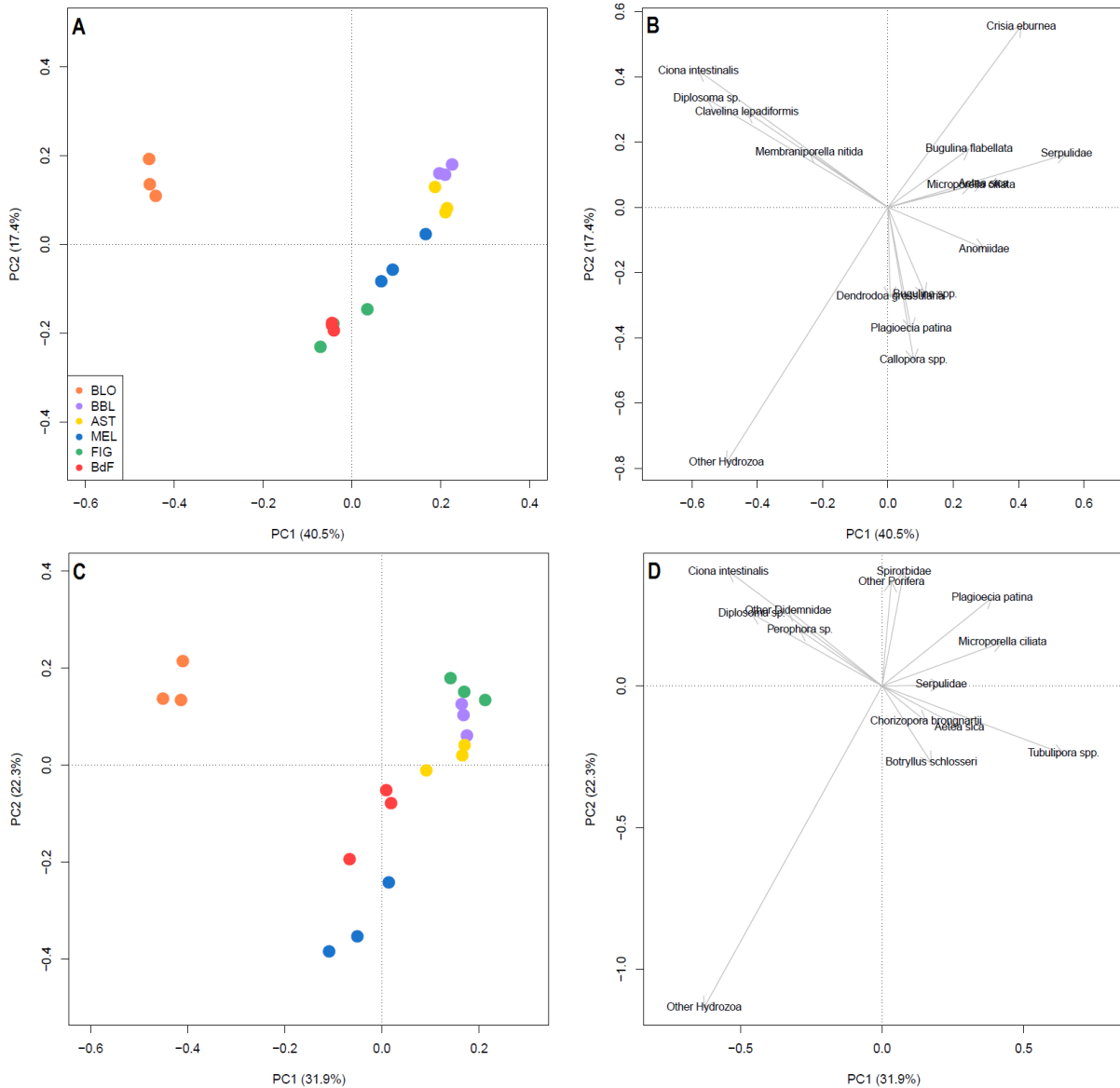


**Figure 6** Proportion of non-indigenous and cryptogenic species (pink) and native species (beige) as observed by the three methods used, Morphology (M), plates bulkDNA (B), and water eDNA (W). NIS that might be false positives (see Chapter II and Table 3) were excluded.

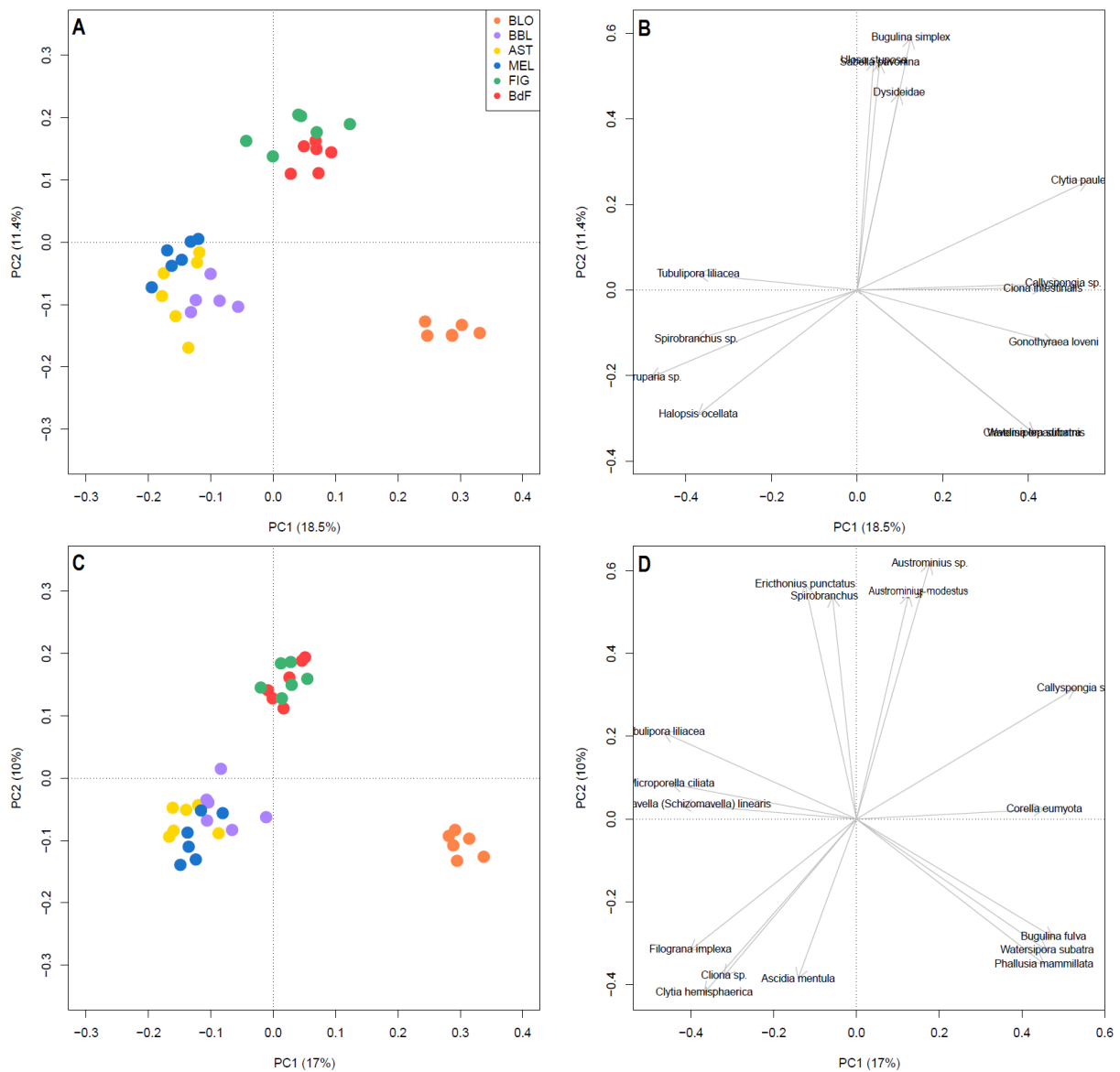
**Table 4** Summary statistics for the percentage of NIS and cryptogenic species recorded per method.

	Plate Morphology	Plate BulkDNA	Water eDNA
Range	23-46	10-29	8-23
Site with the lowest value	MEL	MEL	MEL
Site with the highest value	BLO	BLO	FIG (BLO: 19)
Mean	32	22	17
SD	8	7	5

Spatial diversity patterns were similar for both the morphology and the plate bulkDNA datasets. The marina (BLO) community was consistently very different from all other localities (Fig. 7A-8A, 7C-8C), whatever the season. The species explaining these differences were mainly ascidians (Fig. 7B-8B, 7D) and bryozoans (Fig. 8D). The other localities were clustered according to their position in the bay, either by grouping the outer sites and the inner sites (Fig. 7A; 8A-8C) or along an east-west axis (Fig. 7C). Interestingly, the same patterns were observed with water eDNA collected at the time of deployment of the settlement structures (Fig. S1) but not with those collected upon retrieval of the structures (Fig. 9). In this case, there was again a strong difference between inner and outer sites, but the marina was not distinct from its neighbouring localities (BBL and FIG) at the 2 seasons. The site BdF, however, appeared to be strongly separated from the other inner localities.



**Figure 7** Ordination plots of principal component analysis results from Hellinger-transformed tax abundances collected from morphological identification on settlement plates. The two seasons of sampling have been treated separately (Summer: A-B; Fall C-D). Sample scores are displayed in scaling 1. Colours indicate the different sampled localities.



**Figure 8** Ordination plots of principal component analysis results from Hellinger-transformed taxa occurrences collected from metabarcoding of bulkDNA from settlement plates. The two seasons of sampling have been treated separately (Summer: A-B; Fall C-D). Sample scores are displayed in scaling 1. Colours indicate the different sampled localities.



## Discussion

As points of entry for many non-indigenous species (NIS), marinas are invasion hubs, especially regarding benthic sessile organisms which can establish on artificial hard substrates and subsequently disperse in surrounding natural rocky habitats. However, a low number of NIS have often been reported in natural environments (Bulleri and Chapman 2004; Wasson, et al. 2005; Ruiz, et al. 2009), raising important ecological and methodological questions. The low abundances of NIS in natural environments could be related to NIS potential for dispersal and establishment outside of artificial habitats or to our ability to detect them in the wild. In both cases, there is a need for determining to which extent artificial substrates in marinas display unique communities. In this context, we used morphology- and molecular-based identifications on settlement plates and water samples to assess community dissimilarities between a French marina and five localities distributed in the adjacent bay. As expected, our results revealed a very distinct assemblage in the marina, but, conversely to our expectations, this might not be related to a higher proportion of NIS. Moreover, regardless of their native *vs.* non-indigenous status, many species observed within the marina were also found in other locations inside the bay suggesting substantial ecological connectivity between artificial and natural habitats.

### **a. Marina assemblages are not equivalent to communities observed in natural environments**

Communities observed on the settlement plates after ten weeks of immersion (i.e. at an early developmental stage) were strongly different between the marina and the other localities (Figs. 7-8). Both methods used to identify taxa, either based on morphology or on metabarcoding, revealed the same dissimilarities, despite detecting different taxa (Fig. 3). This result is concordant with numerous studies that reported differences in community composition for benthic sessile epifauna between artificial structures and natural rocky reefs (e.g. Bulleri and Chapman 2004; Perkol-Finkel and Benayahu 2007; Bulleri and Chapman 2010; Airoidi, et al. 2015).

Marinas being mostly composed of artificial hard structures (pilings, seawalls, floating pontoons...) the assemblages they harbor would likely differ from those of natural environments, offering solely rocky reefs as suitable habitats. In order to evaluate the potential for species found in the marina to occur outside of the marina,



and to eliminate biases due to the influence of different substrates on settlement, we used standardized settlement panels as observation unit. Such biases were reported for instance for ascidian larvae (Chase, et al. 2016), which were among the targeted taxonomic groups in this study, being major colonizers in marinas. Such experimental collectors are not only commonly used to survey marine fouling communities (Marraffini, et al. 2017) but are also routinely deployed for biodiversity monitoring coupled with (meta)barcoding (Ransome, et al. 2017). By doing this, any difference observed in our case would thus not be imputable to the artificial substrate, and would more likely be explained by a different species pool available at the time of recruitment (i.e. some species might only be present within the marina), and/or by different biotic and abiotic conditions that might favor or hinder the settlement of a given taxon in the two types of habitat. The settlement structures were, however, suspended under floating pontoons in marinas and placed on the bottom at the other localities. Dafforn, et al. (2009) showed that shallow moving surfaces did not exhibit the same communities than deep fixed surfaces, so our experimental procedure might also explain part of the differences observed. This design was chosen to reflect the reality of the conditions encountered at settlement by species in every locality and the difference that might be created by this distinction can be considered as inherent to these environmental conditions.

Interestingly, the singularity of the marina was less pronounced when focusing on data produced by metabarcoding of water eDNA (Figs. 9 and S1), as compared to Plate bulkDNA. Water eDNA recovered almost all benthic sessile taxa observed on settlement plates (with metabarcoding) but also a high number of additional taxa (Fig. 3). These might comprise species too rare to settle on our structures, species which settled preferentially on close natural rocky reefs (i.e. substrate selection), species which did not reproduce during the period of deployment of our settlement structures, or species with direct development, that did not reach the settlement plates. Even within marinas, where most of the available hard substrate is artificial, a higher number of taxa was recovered with water eDNA (Fig. 5). The settlement plates we used are good surrogates for fouling communities under floating pontoons (Marraffini, et al. 2017) but some species preferentially settle on other structures such as pilings or concrete walls (e.g. Dafforn, et al. 2012). Our results suggest that the difference between benthic sessile communities, including all hard substrates available, is lower than when comparing exclusively the communities established on a particular type of substrate. This is concordant with previous studies that examined the effects of microhabitats on biofouling communities (e.g. Leclerc, et al. 2020). Moreover, the marina did not particularly stand out from the other localities

in both July and November water eDNA datasets probably because of the greater difference between BdF and the other localities (Fig. 9). In these cases, a hydrozoan species from the genus *Nemertesia* was found exclusively in BdF samples. This taxon was not observed on the settlement plates and might thus preferentially settle on natural substrates. Two bivalves were also found almost exclusively in this locality, the queen scallop *Aequipecten opercularis* (Linnaeus, 1758) in July and an oyster from the genus *Crassostrea* (most likely the Pacific oyster *Crassostrea gigas* (Thunberg, 1793), conspicuous in the study bay because of the presence of many oyster farms) in July, August, and November. These identifications correspond to the reproductive period for both species (Román, et al. 1996; Enríquez-Díaz, et al. 2008) and the collection of released gametes within our water samples might be partly responsible for the distinction of BdF for these datasets. Moreover, although the taxa included in our datasets can be found on hard bottoms, some may also live on soft bottoms, such as the above example of *Aequipecten opercularis*, that lives attached by a byssus in its early life and lives as an active swimmer when adult (Tebble 1976).

Over all methods, the ordination of localities showed two types of spatial structure, either contrasting the inner and outer localities, or a west-east gradient. These patterns, and especially the inner-outer gradient, were consistent across all datasets. This was, for example, characterized by a lower richness in the outer localities (especially MEL) and a higher richness in the inner sites (Fig. 5). Several environmental conditions differ and may explain the gradients in community composition, for example FIG and BdF are more sheltered and located in two estuaries whereas AST, MEL and BBL are more exposed to strong currents. AST and MEL are also less influenced by river inputs and anthropogenic disturbances (including nutrient release from land-farming). MEL is an extreme case, this island being considered locally as a rather 'pristine site'. Finally, the discrimination between western and eastern of localities could be due to the influence of the two rivers, with the flow connecting localities situated at the exit of each estuary.

## **b. Are non-indigenous species observed outside the marina?**

The recurring arrival of NIS *via* hull fouling in marinas is expected to produce a high colonization and propagule pressure in these specific environments. Consequently, proportions of NIS are thought to be higher in marinas and harbours than in neighbouring natural habitats, especially for biofouling communities (e.g. Rivero, et al. 2013; Airoidi, et al. 2015). In this study, proportions of NIS and

cryptogenic species within the marina were similar or only slightly higher than in other sampling points in the bay (pairwise differences were not significant). Moreover, no gradient could be observed moving away from the marina, neither in terms of richness nor diversity. BBL, which is located just outside the marina, was not closer to the marina but to AST and MEL, the two outer localities. These results suggested that, either most NIS are already established in various parts of the bay (which has not been reported yet) and can colonize new substrates when available, or NIS are effectively exported outside of the marina but are only able to settle when a suitable substrate (here artificial) is offered. Such exportation could be achieved through natural larval dispersal (see examples discussed below), or through human-mediated transport. In fact recreational boating is the main dispersal vector for short-lived benthic sessile epifauna (Clarke Murray, et al. 2011) and the frequent travels of boats from the marina into the bay might be responsible for the major part of species dispersal from the marina. Aquaculture facilities, however, could also be another source for NIS as they offer a wide surface of artificial substrates. Several oyster farms are present within the estuaries of the bay of Morlaix and their proximity to FIG and BdF might have allowed NIS settlement in these localities. Interestingly, MEL exhibited always the lowest proportion of NIS whatever the method, in particular when compared to BLO (e.g. 10% vs. 30% for the Plate BulkDNA dataset; Table 4), although the difference was not significant. This suggests that NIS did not reach this point, which could be explained by the lower number of boats cruising in this location or by the restricted potential for natural dispersal of the targeted species. Another explanation could also be related to the more exposed environmental conditions which might not be suitable for most NIS to settle there.

The proportion of species found exclusively within the marina was very low, between 5% and 7% according to the method used. When combining results across all methods, only 15 species were observed solely in the marina. Some of these species might have been present in other localities but not found because of their low abundance. This was the case for *Bugulina fulva*, for example, which was only detected in BLO in the plate bulkDNA dataset but was also observed in FIG in the morphology dataset. Other species might be found solely on settlement plates in the marina because they settle preferentially on rocky substrates in natural environments, and/or in intertidal habitats. This is likely the case of the intertidal barnacle *Chthamalus montagui* Southward, 1976, which was found only in BLO in the plate bulkDNA dataset but was identified in all localities but one in the water eDNA dataset, reflecting its common occurrence in the bay. Seven species, observed within the marina only, were classified as NIS or cryptogenic. Some of them might be new

introductions or assignment errors due to a lack of reference sequence for a closely related native species or to a lack of resolution for the marker used. Either way, the ASV in question was still detected solely in the marina which might argue for the possibility of a novel introduction. Other species, however, are known in the region for a long time and their absence outside the marina suggested they might not be able to disperse and establish outside of the marina. One example is *Watersipora subatra* (Ortmann, 1890) which was found only in BLO in the morphology dataset. Both metabarcoding datasets identified this species outside the marina, in BBL and MEL, but with a very small amount of reads as compared to the read abundance in BLO. This species is known to have a short pelagic larval duration (typically 3-8 hours; (Marshall and Keough 2003; Sams, et al. 2015), that would fit a transport from the marina to BBL (where it was observed in plates bulkDNA) or even AST (not observed) over a single ebb tide (Cabioch and Douvillé 1979). Similar dispersal distances (a few km) were already observed for this species by Page et al. (2019). On the other hand, such a pelagic duration is not compatible with dispersal from the marina to the most distant site MEL (where it was observed in water eDNA) that would take longer. On some occasions, larval period might last longer (up to 24 h; Ng and Keough 2003), but this lengthened period could be at the expense of growth and survival of the colonies (Sams et al. 2015). The presence of *W. subatra* in our water eDNA samples might thus be due to either colonies in low abundance or simply free DNA dispersing from the marina. Further, Cacabelos, et al. (2020) have suggested that biofilms from natural habitats might inhibit the settlement of *W. subatra*, while this species seems highly tolerant to pollution (Ng and Keough 2003; Piola and Johnston 2009) and might be particularly suited to colonize artificial substrates in highly disturbed and polluted environments. This could explain why this species is able to thrive in marinas but not in natural habitats, despite being able to disperse to other localities. Another interesting example is the bryozoan *Bugula neritina* (Linnaeus, 1758). It was not observed in any other locality than the marina with none of the three methods used and could be an example of species that is not able to disperse outside of the marina. Its natural dispersal is very limited since its larvae are very short-lived, less than 2 hours if a suitable substrate is available (although in the laboratory they can swim for up to two days without a suitable substrate) (Keough 1989). This species might still be able to spread *via* hull fouling and more sampling would be necessary to confirm its absence in the bay.

As a conclusion, our results revealed a strong difference between early-stage assemblages collected on settlement plates within and outside the marina. This difference was, however, lower when considering all benthic sessile epifaunal taxa recovered by eDNA. Most species identified within the marina were also observed in other localities in the bay with either method used, suggesting that marinas can be both a source and a recipient of species established in the wild. This is concordant with the phenomenon of spillover (i.e. colonization of natural habitats from marinas) and spillback (i.e. colonization of artificial structures from natural habitats) shown for instance with the introduced alga *Undaria pinnatifida* (Epstein and Smale 2018; Salamon, et al. 2020). Some of the species found exclusively in the marina were NIS and might be unable to colonize other habitats in the bay. Our study was, however, not able to discriminate if their absence, or low abundance was due to a poor dispersal ability or a low settlement capacity due to predation or competition with native species. To answer this question, it would be interesting to assess the presence of these particular species in the plankton, and to carry out manipulative experiments in natural habitats. The next part of this [chapter \(III.2\)](#) partly answers this question, showing the presence of dispersal stages of several NIS in the bay of Morlaix, even for short dispersers like *W. subatra* or *B. neritina*. Plankton samples were also collected every two weeks from April to November at the same six sampling points where the settlement structures were immersed. The samples had not been processed yet but the information they might give us would be a valuable addition for understanding the dispersal ability of NIS outside of marinas.

## References

- Abe H, Takeuchi T, Taru M, Sato-Okoshi W, Okoshi K. 2019. Habitat availability determines distribution patterns of spionid polychaetes (Annelida: Spionidae) around Tokyo Bay. *Marine Biodiversity Records* 12:7.
- Airoidi L, Beck M. 2007. Loss, Status and Trends for Coastal Marine Habitats of Europe. *Oceanography and Marine Biology: An Annual Review* 45:345-405.
- Airoidi L, Turon X, Perkol-Finkel S, Rius M. 2015. Corridors for aliens but not for natives: effects of marine urban sprawl at a regional scale. *Diversity and Distributions* 21:755-768.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403-410.
- Ayres DR, Garcia-Rossi D, Davis HG, Strong DR. 1999. Extent and degree of hybridization between exotic (*Spartina alterniflora*) and native (*S. foliosa*) cordgrass (Poaceae) in California, USA determined by random amplified polymorphic DNA (RAPDs). *Molecular Ecology* 8:1179-1186.
- Barragán JM, de Andrés M. 2015. Analysis and trends of the world's coastal cities and agglomerations. *Ocean & Coastal Management* 114:11-20.

- Bishop MJ, Mayer-Pinto M, Airoidi L, Firth LB, Morris RL, Loke LHL, Hawkins SJ, Naylor LA, Coleman RA, Chee SY, et al. 2017. Effects of ocean sprawl on ecological connectivity: impacts and solutions. *Journal of Experimental Marine Biology and Ecology* 492:7-30.
- Borcard D, Gillet F, Legendre P. 2018. *Numerical ecology with R*: Springer.
- Bouchemousse S. 2015. Dynamique éco-évolutive de deux ascidies congénériques et interfertiles, l'une indigène et l'autre introduite, dans leur zone de sympatrie. [Université Pierre et Marie Curie-Paris VI.
- 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* 16:176-182.
- Briski E, Chan FT, Darling JA, Luringson V, MacIsaac HJ, Zhan A, Bailey SA. 2018. Beyond propagule pressure: importance of selection during the transport stage of biological invasions. *Frontiers in Ecology and the Environment* 16:345-353.
- Bulleri F, Chapman MG. 2004. Intertidal assemblages on artificial and natural habitats in marinas on the north-west coast of Italy. *Marine Biology* 145:381-391.
- Bulleri F, Chapman MG. 2010. The introduction of coastal infrastructure as a driver of change in marine environments. *Journal of Applied Ecology* 47:26-35.
- Cabioch L, Douvillé J-L. 1979. La circulation des eaux dans la baie de morlaix et ses abords: premières données obtenues par suivis de flotteurs dérivants. *Travaux de la Station Biologique de Roscoff* 26:11-20.
- Cacabelos E, Ramalhosa P, Canning-Clode J, Troncoso JS, Olabarria C, Delgado C, Dobretsov S, Gestoso I. 2020. The Role of Biofilms Developed under Different Anthropogenic Pressure on Recruitment of Macro-Invertebrates. *Int J Mol Sci* 21.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581.
- Chase AL, Dijkstra JA, Harris LG. 2016. The influence of substrate material on ascidian larval settlement. *Marine Pollution Bulletin* 106:35-42.
- Clarke Murray C, Pakhomov EA, Therriault TW. 2011. Recreational boating: a large unregulated vector transporting marine invasive species. *Diversity and Distributions* 17:1161-1172.
- Comtet T, Sandionigi A, Viard F, Casiraghi M. (Comtet2015 co-authors). 2015. DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions* 17:905-922.
- Connell SD. 2001. Urban structures as marine habitats: an experimental comparison of the composition and abundance of subtidal epibiota among pilings, pontoons and rocky reefs. *Marine Environmental Research* 52:115-125.
- Dafforn K. 2017. Eco-engineering and management strategies for marine infrastructure to reduce establishment and dispersal of non-indigenous species. *Management of Biological Invasions* 8:153-161.
- Dafforn KA, Glasby TM, Johnston EL. 2012. Comparing the Invasibility of Experimental "Reefs" with Field Observations of Natural Reefs and Artificial Structures. *PLOS ONE* 7:e38124.
- Dafforn KA, Johnston EL, Glasby TM. 2009. Shallow moving structures promote marine invader dominance. *Biofouling* 25:277-287.
- Darling JA, Galil BS, Carvalho GR, Rius M, Viard F, Piraino S. 2017. Recommendations for developing and applying genetic tools to assess and manage biological invasions in marine ecosystems. *Marine Policy* 85:54-64.
- Enríquez-Díaz M, Pouvreau S, Chávez-Villalba J, Le Pennec M. (Enríquez-Díaz2008 co-authors). 2008. Gametogenesis, reproductive investment, and spawning behavior of the Pacific giant oyster *Crassostrea gigas*: evidence of an environment-dependent strategy. *Aquaculture International* 17:491.
- Epstein G, Smale DA. 2018. Environmental and ecological factors influencing the spillover of the non-native kelp, *Undaria pinnatifida*, from marinas into natural rocky reef communities. *Biological Invasions* 20:1049-1072.
- Firth LB, Knights AM, Bridger D, Evans AJ, Mieszkowska N, Moore PJ, O'Connor NE, Sheehan EV, Thompson RC, Hawkins SJ. 2016. Ocean sprawl: challenges and opportunities for biodiversity management in a changing world. *Oceanography and Marine Biology: An Annual Review* 54:193-269.

- Fonseca V, Carvalho G, Sung W, Johnson HF, Power D, Neill S, Packer M, Blaxter ML, Lamshead PJ, Thomas W, et al. 2010. Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications* 1:98.
- Glasby T. 1999. Interactive effects of shading and proximity to the seafloor on the development of subtidal epibiotic assemblages. *Marine Ecology-progress Series - MAR ECOL-PROGR SER* 190:113-124.
- Glasby TM, Connell SD, Holloway MG, Hewitt CL. (Glasby2007 co-authors). 2007. Nonindigenous biota on artificial structures: could habitat creation facilitate biological invasions? *Marine Biology* 151:887-895.
- Katsanevakis S, Wallentinus I, Zenetos A, Leppäkoski E, Çinar ME, Oztürk B, Grabowski M, Golani D, Cardoso AC. 2014. Impacts of invasive alien marine species on ecosystem services and biodiversity: a pan-European review. *Aquatic Invasions* 9:391-423.
- Kelly RP, O'Donnell JL, Lowell NC, Shelton AO, Samhuri JF, Hennessey SM, Feist BE, Williams GD. 2016. Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ* 4:e2444.
- Keough MJ. 1989. Dispersal of the bryozoan *Bugula neritina* and effects of adults on newly metamorphosed juveniles. *Marine Ecology Progress Series* 57:163-171.
- Leclerc J-C, Viard F, González Sepúlveda E, Díaz C, Neira Hinojosa J, Pérez Araneda K, Silva F, Brante A. 2020. Habitat type drives the distribution of non-indigenous species in fouling communities regardless of associated maritime traffic. *Diversity and Distributions* 26:62-75.
- Leclerc JC, Viard F. 2018. Habitat formation prevails over predation in influencing fouling communities. *Ecology and Evolution* 8:477-492.
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ. (Leray2013 co-authors). 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology* 10:34.
- Ligthart M, ten Hove H, Faasse M. 2011. De kalkkokerwormen *Apomatus cf. similis* Marion & Bobretzky, 1875 en *Vermiliopsis striaticeps* (Grube, 1862) autochtoon aangetroffen in Nederland (Annelida: Polychaeta: Serpulidae). *Het Zeepaard* 71:88 - 95.
- Marraffini M, Ashton G, Brown C, Chang A, Ruiz G. 2017. Settlement plates as monitoring devices for non-indigenous species in marine fouling communities. *Management of Biological Invasions* 8:559-566.
- Marshall DJ, Keough MJ. 2003. Variation in the dispersal potential of non-feeding invertebrate larvae: the desperate larva hypothesis and larval size. *Marine Ecology Progress Series* 255:145-153.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*; Vol 17, No 1: Next Generation Sequencing Data Analysis DO - 10.14806/ej.17.1.200.
- Miralles L, Gomez-Agenjo M, Rayon-Viña F, Gyraitė G, Garcia-Vazquez E. 2018. Alert calling in port areas: Marine litter as possible secondary dispersal vector for hitchhiking invasive species. *Journal for Nature Conservation* 42:12-18.
- Molnar JL, Gamboa RL, Revenga C, Spalding MD. 2008. Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment* 6:485-492.
- Ng TYT, Keough MJ. 2003. Delayed effects of larval exposure to Cu in the bryozoan *Watersipora subtorquata*. *Marine Ecology Progress Series* 257:77-85.
- Núñez AL, Katsanevakis S, Zenetos A, Cardoso AC. 2014. Gateways to alien invasions in the European seas. *Aquatic Invasions* 9:133-144.
- O'Shaughnessy KA, Hawkins SJ, Yunn ALE, Hanley ME, Lunt P, Thompson RC, Firth LB. 2020. Occurrence and assemblage composition of intertidal non-native species may be influenced by shipping patterns and artificial structures. *Marine Pollution Bulletin* 154:111082.
- Ojaveer H, Galil BS, Campbell ML, Carlton JT, Canning-Clode J, Cook EJ, Davidson AD, Hewitt CL, Jelmert A, Marchini A, et al. 2015. Classification of non-indigenous species based on their impacts: considerations for application in marine management. *PLOS Biology* 13:e1002130.
- Perkol-Finkel S, Benayahu Y. 2007. Differential recruitment of benthic communities on neighboring artificial and natural reefs. *Journal of Experimental Marine Biology and Ecology* 340:25-39.

- Piola RF, Johnston EL. 2009. Comparing differential tolerance of native and non-indigenous marine species to metal pollution using novel assay techniques. *Environmental Pollution* 157:2853-2864.
- Plaisance L, Brainard R, Caley MJ, Knowlton N. 2011. Using DNA Barcoding and Standardized Sampling to Compare Geographic and Habitat Differentiation of Crustaceans: A Hawaiian Islands Example. *Diversity* 3:581-591.
- Radashevsky VI, Pankova VV, Malyar VV, Neretina TV, Wilson RS, Worsfold TM, Diez ME, Harris LH, Hourdez S, Labrune C, et al. 2019. Molecular analysis and new records of the invasive polychaete *Boccardia proboscidea* (Annelida: Spionidae). 2019 20:16.
- Ransome E, Geller JB, Timmers M, Leray M, Mahardini A, Sembiring A, Collins AG, Meyer CP. 2017. The importance of standardization for biodiversity comparisons: A case study using autonomous reef monitoring structures (ARMS) and metabarcoding to measure cryptic diversity on Mo'orea coral reefs, French Polynesia. *PLOS ONE* 12:e0175066.
- Rivero NK, Dafforn KA, Coleman MA, Johnston EL. 2013. Environmental and ecological changes associated with a marina. *Biofouling* 29:803-815.
- Román G, Campos MJ, Acosta CP. 1996. Relationships among environment, spawning and settlement of Queen scallop in the Ría de Arosa (Galicia, NW Spain). *Aquaculture International* 4:225-236.
- Ruiz G, Freestone A, Fofonoff P, Simkanin C. 2009. Habitat Distribution and Heterogeneity in Marine Invasion Dynamics: the Importance of Hard Substrate and Artificial Structure. In. p. 321-332.
- Ryland J, Bishop J, De Blauwe H, Nagar A, Minchin D, Wood C, Yunnice A. 2011. Alien species of *Bugula* (Bryozoa) along the Atlantic coasts of Europe. *Aquatic Invasions* 6:17-31.
- Salamon M, Lévêque L, Ballenghien M, Viard F. 2020. Spill-back events followed by self-sustainment explain the fast colonization of a newly built marina by a notorious invasive seaweed. *Biological Invasions* 22:1411-1429.
- Sams MA, Warren-Myers F, Keough MJ. 2015. Increased larval planktonic duration and post-recruitment competition influence survival and growth of the bryozoan *Watersipora subtorquata*. *Marine Ecology Progress Series* 531:179-191.
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, Winter M, Arianoutsou M, et al. 2017. No saturation in the accumulation of alien species worldwide. *Nature Communications* 8:14435.
- Simkanin C, Davidson IC, Dower JF, Jamieson G, Therriault TW. 2012. Anthropogenic structures and the infiltration of natural benthos by invasive ascidians. *Marine Ecology* 33:499-511.
- Sinniger F, Pawlowski J, Harii S, Gooday AJ, Yamamoto H, Chevaldonné P, Cedhagen T, Carvalho G, Creer S. 2016. Worldwide analysis of sedimentary DNA reveals major gaps in taxonomic knowledge of deep-sea benthos. *Frontiers in Marine Science* 3.
- Soares MdO, Campos CC, Santos NMO, Barroso HdS, Mota EMT, Menezes MOBd, Rossi S, Garcia TM. 2018. Marine bioinvasions: Differences in tropical copepod communities between inside and outside a port. *Journal of Sea Research* 134:42-48.
- Tebble N. 1976. British bivalve seashells; a handbook for identification-2nd Edition. In. *Natural History*. London: Trustees of the British Museum. p. 212.
- Ulman A, Ferrario J, Forcada A, Seebens H, Arvanitidis C, Occhipinti-Ambrogi A, Marchini A. 2019. Alien species spreading via biofouling on recreational vessels in the Mediterranean Sea. *Journal of Applied Ecology* 56:2620-2629.
- Veiga P, Ramos-Oliveira C, Sampaio L, Rubal M. 2020. The role of urbanisation in affecting *Mytilus galloprovincialis*. *PLOS ONE* 15:e0232797.
- Viard F, Riginos C, Bierne N. 2020. Anthropogenic hybridization at sea: three evolutionary questions relevant to invasive species management. *Philos Trans R Soc Lond B Biol Sci* 375:20190547.
- Wasson K, Fenn K, Pearse JS. 2005. Habitat Differences in Marine Invasions of Central California. *Biological Invasions* 7:935-948.
- Zaiko A, Pochon X, Garcia-Vazquez E, Olenin S, Wood SA. 2018. Advantages and limitations of environmental DNA/RNA tools for marine biosecurity: management and surveillance of non-indigenous species. *Frontiers in Marine Science* 5.



## Metabarcoding on planktonic larval stages: an efficient approach for detecting and investigating life cycle dynamics of benthic aliens

**Running headline:** NIS detection and survey using metabarcoding of zooplankton samples

Marjorie Couton<sup>1</sup>, Thierry Comtet<sup>1</sup>, Sabrina Le Cam<sup>1</sup>, Erwan Corre<sup>2</sup>, Frédérique Viard<sup>1\*</sup>

<sup>1</sup> Sorbonne université, CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

<sup>2</sup> Sorbonne université, CNRS, FR 2424, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France

\* Correspondence author: viard@sb-roscoff.fr; +33 2 98 29 23 12

**Keywords:** zooplankton, time-series, non-indigenous species, estuary, high-throughput sequencing, surveillance, monitoring

**This part has been published in Management of Biological Invasions:**

Couton M, Comtet T, Le Cam S, Corre E, Viard F (2019) Metabarcoding on planktonic larval stages: an efficient approach for detecting and investigating life cycle dynamics of benthic aliens. *Management of Biological Invasions* **10**:657-689.

**DOI:** 10.3391/mbi.2019.10.4.06

## Abstract

High-throughput sequencing (HTS) technologies offer new promise to support surveillance programs targeting marine non-indigenous species (NIS). Metabarcoding might surpass traditional monitoring methods, for example through its ability to detect rare species, a key feature in early detection of NIS. Another interest of this approach is the identification of organisms difficult to identify based on morphology only (e.g. early developmental stages), making it relevant in the context of management programs. Because many marine benthic NIS have a bi-phasic benthic-pelagic life cycle, targeting their pelagic larval stages in zooplankton may allow early detection and assessment of their establishment and potential spread. We illustrate this approach with an analysis of bulk-DNA retrieved from a time-series of zooplankton samples collected over 22 months in one bay in Brittany (France). Using HTS of amplicons obtained with two markers (COI and 18S) and a metabarcoding approach, 12 NIS were identified and their temporal larval dynamics were monitored. Importantly, we chose to focus on a closed list of species, from four metazoan classes encompassing 52 NIS reported within the study area or nearby seas, with molecular references available or obtained locally for 42 of them. The use of a custom-designed database allowed the detection of three NIS otherwise not detected when using public databases. Interestingly, NIS known to have a short-lived larval stage were detected (e.g. the bryozoan *Bugula neritina* or the tunicate *Corella eumyota*). For two molluscs *Ruditapes philippinarum* and *Crepidula fornicata*, metabarcoding results were compared to those obtained using traditional methods (i.e. barcoding of individual larvae and morphology, respectively) to show the reliability of the approach in detecting and assessing the extent of their reproductive periods. Our results also revealed that the Pacific oyster *Crassostrea gigas*, a notorious invasive species, failed to reproduce in the study bay, showing that metabarcoding on larval stages also provides information regarding the establishment success (or failure) of NIS. While metabarcoding has its limitations and biases, this study demonstrates its effectiveness for surveillance of targeted NIS, notably to support management strategies like the European Marine Strategy Framework Directive (MSFD).

## Introduction

The number of marine non-indigenous species (NIS) has been increasing globally since the beginning of the 20<sup>th</sup> century. This trend is an outcome of increasing maritime traffic and trade, and is expected to last (Sardain et al. 2019; Seebens et al. 2017). NIS can cause a wide variety of ecological (e.g. biodiversity loss, changes in ecosystem dynamics) and economical (e.g. infrastructure maintenance, aquaculture losses) damages (Molnar et al. 2008), which entail a wide range of actions, from prevention to long-term management (Simberloff et al. 2013). Monitoring NIS is therefore crucial in order to set-up handling strategies adapted to the different phases of the invasion sequence (Blackburn et al. 2011). Early detection will promote action at the earliest stage, during which NIS control is likely to be the most efficient, particularly in the marine environment (Ojaveer et al. 2015). On the other hand, monitoring NIS establishment and spread will allow long-term management and evaluation (e.g. reinvasion after eradication; Simberloff et al. 2013).

In coastal areas, shipping (commercial trade and leisure boating) and aquaculture are the most important introduction pathways (Molnar et al. 2008; Nuñez et al. 2014). Consequently, ports and aquaculture facilities, with their numerous artificial substrates, are points-of-entry for NIS and promote the settlement of new species (Bishop et al. 2015b; Connell 2001; Glasby et al. 2007), especially encrusting and sessile fauna (Firth et al. 2016). These infrastructures and facilities act as bridgeheads for the escape of newly introduced species into nearby natural habitats (Airoldi et al. 2015; Firth et al. 2016), where most ecological damage is observed. The lag phase between the primary introduction (arrival) of new NIS into artificial habitats and their escape into the wild is variable across species, imposing the need for regular temporal surveys in nearby natural environments. Such surveys could be achieved by applying NIS detection tools to existing long-term monitoring programmes (Ojaveer et al. 2015).

One particular feature of many marine coastal invertebrate species is the existence of a biphasic, benthic-pelagic life cycle, during which the benthic adult stage alternates with a pelagic larval phase (Mileikovsky 1971), living in the plankton for hours, weeks, sometimes months (Shanks 2009). In marine benthic NIS such a pelagic larval stage plays a major role at all steps of the invasion process (i.e. introduction, establishment and spread, *sensu* Blackburn et al. 2011). Given their small size, larvae can be transported in ballast water (Carlton and Geller 1993) or they can be released from adults within hull fouling communities (e.g. species brooding

their embryos before releasing swimming larvae like some barnacle species) and thus can be major actors of primary introductions. Larvae may also facilitate the long-term establishment of introduced species by promoting the demographic reinforcement of their local populations through reproduction and recruitment (Dunstan and Bax 2007). Finally, they are the main vector for natural dispersal (Cowen et al. 2007), thus playing a major role in secondary spread and expansion of NIS in novel introduction areas. This is illustrated by notorious invasive species with long-lived larval stages, such as the green crab *Carcinus maenas* (Linnaeus, 1758) (Pringle et al. 2011; Tepolt et al. 2009).

Targeting larvae in monitoring programs may provide key information about NIS introduction status. When sampling in close vicinity of entry points such as harbours, the identification of larvae belonging to a formerly unreported species delivers early detection of a newly arrived NIS. Likewise, the presence of larvae assigned to an already reported NIS will prove its reproductive ability in the novel environment, its potential for spread, and will suggest that this species is now established. Moreover, monitoring the larvae of a target NIS over time may allow a better understanding of its reproductive biology in the introduced area. In particular, it may shed light on the period and environmental conditions that favour its reproduction, as well as its reproductive effort (abundance of larvae). The collected data could support predictive models, such as ecological niche models. Finally, because NIS larvae are also non-indigenous within the local plankton, observing and counting them may allow to better assess their potential impact within the pelagic community, a largely understudied topic.

Monitoring programs targeting larvae are quite rare and regular programs monitoring zooplankton usually neglect invertebrate larvae, or consider them as broad taxonomic groups, like “lamellibranch larvae” as a whole (e.g. Southward et al. 2005). Identifying species at the larval stage is indeed challenging, especially with traditional methods based on larval sorting and identification with morphological criteria, a time-consuming task which requires well-trained taxonomists. Difficulties may also arise from a lack of description for the larval stages of some taxa, but even when they exist, many species are indistinguishable from each other based on simple morphological criteria. This is especially true for groups like bivalves (Garland and Zimmer 2002), or in taxa comprising cryptic species which are numerous among marine invertebrates (Appeltans et al. 2012). To overcome these issues, several single-species DNA-based tools have been used for the identification of NIS larvae (e.g. Darling and Tepolt 2008; Harvey et al. 2009; Le Goff-Vitry et al. 2007; Sánchez et al.

2015). However, these approaches allow the identification (and sometimes quantification) of only one or a few target species. The DNA metabarcoding approach consists of the high-throughput sequencing (HTS) of selected barcodes obtained from environmental or bulk-DNA, and their taxonomic assignment based on a reference sequence database. It can identify simultaneously a large number of species from many specimens, thus being a promising method to study NIS larvae in the plankton (Comtet et al. 2015; Cristescu 2014; Elbrecht and Leese 2015; Valentini et al. 2016; Viard and Comtet 2015; Zaiko et al. 2015, 2018). Moreover, DNA metabarcoding is more sensitive than traditional approaches when detecting rare organisms, a key advantage when investigating newly arrived species at low environmental concentrations. When dealing with larvae, metabarcoding has proven efficient for detecting very few individuals, down to a single larva, in plankton and sediment samples containing a wide array of eukaryotes (Pochon et al. 2013; Sun et al. 2015; Zhan et al. 2013).

Several studies have shown the power of bulk-DNA metabarcoding from plankton samples to assess zooplankton biodiversity and describe the structure of pelagic communities (Abad et al. 2016, 2017; Bucklin et al. 2016; Chain et al. 2016; Deagle et al. 2017; Harvey et al. 2017; Lindeque et al. 2013; Lopez-Escardo et al. 2018). Some of them highlighted the complementarity between morphological identification and molecular HTS approaches at various taxonomic levels (e.g. Harvey et al. 2017; Lindeque et al. 2013), but most of them did not focus on larvae of benthic species. Nonetheless, they conceded that such methods are useful for the identification of larval stages from both pelagic and benthic species, when not possible with other approaches (e.g. Mohrbeck et al. 2015, and references above). To our knowledge, only one study clearly focused on bivalve larvae (Jung et al. 2018), and a few others further demonstrated the interest of plankton DNA metabarcoding for NIS identification at the larval stage, from both pelagic (Abad et al. 2016) and benthic species (Ardura et al. 2015; Brown et al. 2016; Zaiko et al. 2015). In these papers, the main objectives were the early detection of NIS in areas where they had not yet been reported, and the evaluation of metabarcoding as an efficient tool to detect NIS in areas where they had already been reported.

In this study, we focused entirely on the larval stages of benthic NIS (1) to evaluate the use of plankton DNA metabarcoding for the detection of new introductions and (2) to assess the establishment, expansion and reproductive features of already identified benthic NIS. For this purpose we used a metabarcoding

approach on a 22-month plankton time series, to identify larvae of potential and known NIS, targeting four taxonomic groups.

## Materials and methods

### a. Sampling and DNA extraction

The samples used in this study are part of a zooplankton time-series survey, which started in 2004, and is conducted in the bay of Morlaix (48°40'11.1"N; 3°53'9.7"W), Brittany, France. Zooplankton samples are collected bi-monthly around time of high tide ( $\pm$  30 min) during neap tides. Sampling is performed using a vertical haul from the bottom to the surface with a modified WP2 plankton net (UNESCO 1968) with a mesh size of 63  $\mu$ m and a mouth area of 0.25 m<sup>2</sup>. A flow meter (KC Denmark A/S, model 23091) is attached at the centre of the net aperture to determine the volume of water filtered. Samples are preserved in 96% ethanol and stored at room temperature. For this study, to get insights about seasonal patterns of larval presence, and taking into account the large diversity of marine invertebrates reproductive modes, we used samples collected every two weeks from March 2012 to September 2012, and once a month until December 2013, for a total of 29 sampling dates.

Total DNA was extracted from each plankton sample using the PowerWater® DNA Isolation Kit (MoBio). The manufacturer's protocol was slightly modified by adding a drying step after filtration of the samples in order to evaporate all residual ethanol, and all volumes of reagents used were doubled. All equipment was either autoclaved or placed under UV light for 15 minutes before use. One blank was produced following the exact same extraction protocol on ultrapure water (18M $\Omega$ , 0.22 $\mu$ m) to ensure the absence of cross-contamination when processing the samples. Extracted DNA was stored at  $-20$  °C.

### b. Molecular analyses

For each sample, and the extraction blank, amplicons were generated in triplicates to minimize PCR biases. We used the primers SSU\_FO4 and SSU\_R22 from

Fonseca et al. (2010) targeting a ca. 400-bp portion of the V1–V2 region of the small subunit rRNA (18S) coding gene, and the primers mlCOIintF (Leray et al. 2013) and jgHCO2198 (Geller et al. 2013), following Leray et al. (2013), amplifying a 313-bp portion of the Cytochrome Oxidase I (COI) coding gene (mitochondrial DNA). In addition to the 90 PCR replicates from extracted samples, three PCR blanks were performed to check for any contamination at this step. Every PCR replicate of every sample and every blank was individually tagged using eight-nucleotide sequences (tags) added at the 5'-end of both forward and reverse primers. Each replicate was thus identified by the unique combination of its forward and reverse tag. For 18S, each reaction volume (25  $\mu$ L) contained 0.5 U Phusion® High-Fidelity DNA polymerase (New England BioLabs), 1X reaction buffer, 200  $\mu$ M dNTPs, 1  $\mu$ M of each tagged primer and 2 ng DNA template. Amplification involved an initial denaturation step at 98 °C for 30 s, followed by 30 cycles at 98 °C for 10 s, 57 °C for 30 s and 72 °C for 30 s, and followed by a final extension step at 72 °C for 10 min. For COI, each reaction volume (25  $\mu$ L) contained 1X Multiplex PCR Master Mix (QIAGEN), 1  $\mu$ M of each tagged primer and 2 ng DNA template. Amplification involved an initial denaturation step at 95 °C for 5 min, followed by 35 cycles at 94 °C for 50 s, 57 °C for 90 s and 72 °C for 30 s, and followed by a final extension step at 68 °C for 10 min. PCR products were then purified using NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel) and their concentration was measured using fluorescence by picogreen™. All PCR products of a same marker were then pooled at equimolar concentrations. Paired-end sequencing of the amplicons was performed by the company FASTERIS (Switzerland) using MiSeq Illumina technology (2x300 bp).

### c. Data processing

After sequencing, reads were processed using the OBITools v1.2.11 pipeline (Boyer et al. 2016). Briefly, paired reads were assembled and then grouped by replicate sample (i.e. demultiplexed) according to their unique tag pair. The primers and tags were then removed and all sequences outside of a specified size range were discarded (350–450 bp for 18S and 300–320 bp for COI). Finally, singletons (sequences present only once and in only one of the PCR replicates) were discarded. PCR and sequencing errors were detected using *obiclean* with a value of 2 for the parameter -d (maximum number of differences allowed for two reads to be considered deriving from one another) and a value of 0.025 for the parameter -r (threshold ratio between counts of two reads under which the less abundant is

classified as an error deriving from the more abundant). These values were determined following *in silico* tests on an artificial dataset (see Supplementary material Figure S1 for details). The end point of these processing steps is a set of unique variants (i.e. sequences all different from one another and expected to be the “true” sequences that were present in the extracted samples).

#### d. Taxonomic assignment

Given the occurrence of numerous errors and missing data in public databases (Briski et al. 2016; Harris 2003), a custom-designed database was used for taxonomic assignment. This database was specific to the study area and curated with caution. A list of 670 species across the four target taxonomic groups (ascidians, bryozoans, bivalves and gastropods) was considered for inclusion in our custom-designed database. These species were native, non-indigenous or cryptogenic. The native species list came from “Fauna inventories of the Station Biologique de Roscoff” (available at <http://www.sb-roscoff.fr/fr/observation/biodiversite/especes/inventaires/inventaires-de-la-faune-et-de-la-flore-marines>) and was completed with new data from local surveys and inventories. The non-indigenous and cryptogenic species list was composed of species that had already been reported in the study bay or were known to be present in the surrounding areas, based on both reports made by French authorities for the European Marine Strategy Framework Directive reporting, and local surveys and inventories (e.g. Bishop et al. 2015b). All available 18S and COI reference sequences associated with these taxa were retrieved from public databases (GenBank: Benson et al. 2013; BOLD: Ratnasingham and Hebert 2007; SILVA: Quast et al. 2013; PR2: Guillou et al. 2013) and manually checked. Only good quality sequences, with no more than one ambiguous base, were included. In order to increase taxonomic coverage, incomplete sequences were also added, where a length of at least 75% of the complete sequence was present. Reference sequences produced locally were also included, in particular for taxa with missing data in the above cited public databases (see Data availability section). For a better discrimination of known or expected NIS, reference sequences for species belonging to the same genus were also included. Finally, all species names were checked with the World Register of Marine Species (WoRMS; WoRMS Editorial Board 2019) to ensure that all names used were valid. In the case of the Pacific oyster, both *Magallana gigas* and *Crassostrea gigas* are accepted names, so we chose to use *Crassostrea gigas* (Thunberg, 1793) as advised by Bayne et al. (2017). In the end, our



custom-designed reference database was composed of 408 and 3131 sequences covering 314 and 410 species, for 18S and COI, respectively (Table 1). Our dataset included 42 different NIS, 35 with reference sequences for both markers; the remaining seven lacked a reference for 18S (Table 1, Table S1).

Unique variants resulting from the OBITools pipeline were then compared to sequences from our custom-designed reference database using BLAST® (Altschul et al. 1990). Only unique variants found in at least two PCR replicates of the same sample were considered for taxonomic assignment. Only alignments covering at least 99% of the reference sequence were considered. Variants were then assigned to the species whose reference sequence had the highest identity percentage. If two or more alignments had the same identity percentage, the variant was assigned to the lowest common taxonomic level. Identity thresholds were defined for each marker in order to consider an assignment to the species level as valid. To select an appropriate threshold, we explored the barcoding gap of our target taxa for the two markers. To that end, each reference sequence was aligned with all the references in the database using the EMBOSS *needle* global alignment tool (version: 6.5.7.0; Rice et al. 2000). For each alignment, the identity percentage was calculated using R 3.4.4 (R Core Team 2018) and the values of intraspecific and interspecific identity were plotted as a density curve using the R package ggplot2 (Wickham 2016). For COI a clear gap between intra- and interspecific identity was observed at 92% (Figure S2B), a threshold thus selected for species assignment. For 18S, the distribution of the intraspecific identity is aggregated towards 100% with a clear gap visible at 98% (Figure S2A). However, to avoid false positive, we chose to apply a more conservative threshold of 99% for this marker. Note that the number of NIS detected using 18S was the same with both thresholds (not shown).

To assess the value of using a custom-designed database for NIS detection, we also performed a taxonomic assignment with two other methods frequently used in the literature and based on public databases, either with a BLAST® approach or by using the *ecotag* command implemented in the OBITools pipeline. For the first method, unique variants were aligned against the GenBank nucleotide database (nt accessed 11.01.19) using BLAST® (v.2.7.1+). Only alignments with at least 99% (18S) or 92% (COI) identity and 99% query cover were considered. Each unique variant was then assigned to the taxon name of the reference sequence with the highest identity percentage. If several reference sequences had identical identity percentages, the unique variant was assigned to the lowest common taxonomic level. For the second approach, we used the *ecotag* command of the OBITools package (Boyer et al. 2016)

with a reference database created using the *ecoPCR* tool available in the same package by retrieving all sequences from GenBank, BOLD, Silva or PR2, and by adding all reference sequences produced locally. In the *ecotag* approach, unique variants are aligned with references using a global alignment algorithm, thus requiring only full-length reference sequences to be included in the database. Then, unique variants are assigned to the taxonomic level corresponding to the lowest common ancestor between all reference sequences that are closer to each other than to the selected unique variant (Boyer et al. 2016). As no minimum identity is required, all unique variants are assigned to a taxon (although sometimes at a high taxonomic level, such as family or class). To summarize, three assignment approaches were used: (1) BLAST® against our custom-designed database, (2) BLAST® against GenBank nt, (3) *ecotag* against its specific database containing only full-length sequences available.

#### e. Tests for amplification failure

Following preliminary results and discrepancies observed between the two markers (see results), and to better interpret the results, we investigated potential taxonomic amplification failures. These tests were performed on the 24 target NIS for which we had access to tissue or DNA. PCR was carried out with the primers mlCOIintF and jgHCO2198 for the COI marker and the primers SSU\_FO4 and SSU\_R22 for the 18S marker, following the same protocols (see above). Amplicons were observed on a 1.5% agarose gel.

#### f. Cross-validation with traditional methods

In order to compare the results from the metabarcoding approach to more traditional ones, two additional datasets were used. They consisted in the parallel sampling, identification and quantification of larvae associated to the bivalve *Ruditapes philippinarum* (Adams and Reeve, 1850) and the gastropod *Crepidula fornicata* (Linnaeus, 1758) during the year 2012.

For *R. philippinarum* larvae, a traditional barcoding approach was used. Briefly, an additional monthly plankton sample was collected at the study site, from March to November 2012, using the same sampling protocol. In the laboratory, each

sample was adjusted to a volume of 150 mL with 96% ethanol. Bivalve larvae were then sorted in three 5 mL subsamples (10% of the sample) under a dissecting microscope aiming to randomly sort at least 100 bivalve larvae, when possible, depending on their overall abundance. Finally, 64–200 bivalve larvae were obtained for each sample to be identified through individual barcoding. The DNA of single larvae was extracted following the method of Lasota et al. (2013). A 550 bp portion of the 5'-end of the 18S coding gene was amplified by PCR using the primers Myt18S-F (Espíñeira et al. 2009) and 18S-571-R (5'-CACCAGACTTGCCCTCCA-3'; C. Roby and T. Comtet, *unpublished*). Each reaction volume (25  $\mu$ L) contained 1 U Thermoprime Plus DNA polymerase (Abgene), 1X reaction buffer, 200  $\mu$ M dNTPs, 3.5 mM MgCl<sub>2</sub>, 0.4  $\mu$ M of each primer, 0.01 mg mL<sup>-1</sup> bovine serum albumin, and 2  $\mu$ L DNA template. Amplification involved an initial denaturation step at 94 °C for 4 min, followed by a 6-cycle touch-down at 94 °C for 40 s, 62–57 °C for 40 s, 72 °C for 1 min, followed by 30 cycles at 94 °C for 40 s, 57 °C for 40 s, 72 °C for 1 min, and followed by a final extension step at 72 °C for 10 min. PCR products were then sequenced in both directions by Sanger sequencing. Overall we obtained an amplification-sequencing success of 64% of all sorted larvae. Obtained sequences were then compared using BLAST® (Altschul et al. 1990) to reference sequences from GenBank and sequences produced locally for local species. Only alignments with 100% cover were considered. A larval sequence was assigned to a species if it differed by less than 2 base pairs (bp) (99.8% identity) from reference sequence(s) of a single species. In all other cases (difference of 2 bp or more and/or several possible species), then the larval sequence was assigned to a family. This 2-bp threshold was based on Blaxter et al. (2005) and empirical observations. In this study, we only focused on larvae assigned to *R. philippinarum*, and results were expressed as their relative abundance compared to all bivalve larvae.

For *C. fornicata*, a traditional morphology-based approach was used. To estimate larvae abundances, additional mesozooplankton samples were collected using a WP2 plankton net with a 200- $\mu$ m mesh size (UNESCO 1968) towed vertically from the bottom to the surface at each sampling date in 2012. Immediately after collection, samples were preserved in 96% ethanol. In the laboratory, larvae of *C. fornicata* were identified and sorted using a dissecting microscope, based on morphology, following early descriptions by Werner (1955), Thiriot-Quévèreux and Scheltema (1982), and using laboratory-reared reference larvae obtained during previous works (e.g. Taris et al. 2010). Our ability to correctly identify *C. fornicata* larvae was validated in a previous study with specific microsatellite loci (Riquet et al. 2017). Since larvae abundances for this species are usually low in the study bay, we

were able to process the whole sample at each date. Pearson correlation coefficients between larvae concentrations and metabarcoding results for all 2012 samples were calculated using the *stats* package implemented in R 3.4.4.

## Results

### a. Overall taxonomic assignment

After sequencing, 6,686,364 and 4,865,347 pairs of raw reads were obtained for 18S and COI, respectively. A total of 5,608,992 and 3,571,274 successfully passed the pairing, demultiplexing and primer removal steps, for 18S and COI, respectively. Only 30,639 (18S) and 1,032 (COI) reads were removed because they did not satisfy the size requirements. Finally, 1,670,807 and 1,022,459 reads corresponding to singletons were discarded, for 18S and COI, respectively. The few reads ( $n = 20$ ) assigned to the 6 blank samples were discarded when checking for singletons. At the end of the processing steps 2,503,893 (37% of the initial number of reads) and 1,450,532 (30%) reads were retained for 18S and COI, respectively. They were composed of 48,037 (18S) and 10,662 (COI) unique variants that were used for taxonomic assignment.

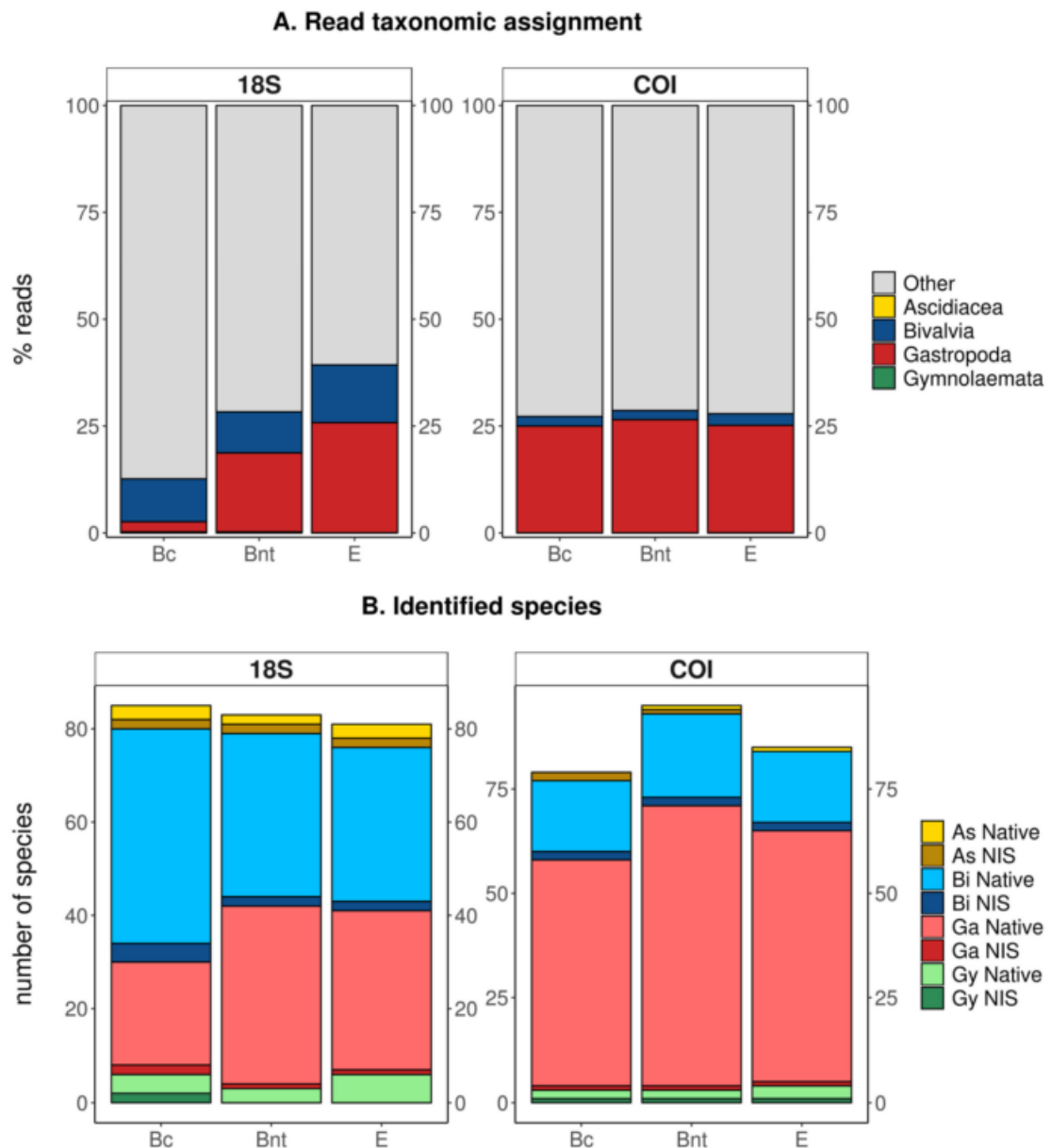
Out of these unique variants, 556 and 2,144 were assigned to an accepted species, genus or family within the four taxonomic groups of interest, when compared to the reference sequences in our custom-designed database, for 18S and COI, respectively (Table 1). These accounted for a total of 317,070 and 395,533 processed reads meaning that *ca.* 13% and 27% of the 18S and COI datasets, respectively, were assigned to one of the target taxa (Figure 1A). With our method, all COI unique variants were assigned to the species level whereas some 18S unique variants were only assigned to higher levels (genus, family or class; Table 1). Overall, within the four taxonomic groups of interest, 86 and 79 taxa were identified at the species level, with 18S and COI respectively (Table 1). Most identified species belonged to Bivalvia (18S) and Gastropoda (COI) (Figure 1A, B), which are the two classes with the highest number of species with a reference sequence in our database (Table 1).

Using alternative methods for taxonomic assignment (i.e. BLAST® against the GenBank nt database or assignment with the OBITools suite) did not change the

proportion of reads assigned to the four target classes for COI (*ca.* 25% of reads assigned) (Figure 1A). For 18S, alternative methods allowed an increase in the proportion of reads assigned to a taxon (up to 40% with *ecotag*) (Figure 1A). However, the number of species identified with 18S was similar across the methods, although the distribution of assigned species per class and/or status species varied (Figure 1B). With COI, the number of species identified was a bit higher due to the number of identified gastropod species, the other class and/or status being similar (Figure 1B).

**Table 1** Number of taxa (with the number of NIS among them indicated in bold and in parentheses), number of unique variants and number of reads identified at different taxonomic levels (species, genus, family or class) for each marker (18S and COI), within the four taxonomic groups of interest (Gymnolaemata, Gastropoda, Bivalvia, Ascidiacea) and using a BLAST® approach against a custom-designed database. The number of taxa (at the species, genus and family levels) with at least one reference sequence in the custom-designed database is given in the Ref column, with the number of reference for NIS in bold and in parentheses.

		18S				COI			
		Taxa with available references	Identified taxa	Unique variants	Reads	Taxa with available references	Identified taxa	Unique variants	Reads
<b>Gymnolaemata</b>				14	6,522			7	645
	Family	20	5	14	6,522	17	3	7	645
	Genus	23	6	14	6,522	24	3	7	645
	Species	34 ( <b>7</b> )	6 ( <b>2</b> )	14	6,522	46 ( <b>10</b> )	3 ( <b>1</b> )	7	645
<b>Gastropoda</b>				147	59,063			2,055	361,835
	Family	40	22	146	59,055	59	33	2,055	361,835
	Genus	52	25	134	40,532	94	43	2,055	361,835
	Species	84 ( <b>5</b> )	24 ( <b>2</b> )	128	38,785	199 ( <b>11</b> )	55 ( <b>1</b> )	2,055	361,835
<b>Bivalvia</b>				385	251,197			80	33,035
	Family	41	24	384	251,195	34	13	80	33,035
	Genus	88	45	317	151,143	67	18	80	33,035
	Species	140 ( <b>13</b> )	50 ( <b>4</b> )	281	140,340	118 ( <b>13</b> )	19 ( <b>2</b> )	80	33,035
<b>Ascidiacea</b>				10	288			2	18
	Family	9	4	10	288	10	1	2	18
	Genus	18	6	10	288	22	1	2	18
	Species	56 ( <b>10</b> )	5 ( <b>2</b> )	5	81	47 ( <b>8</b> )	2 ( <b>2</b> )	2	18
<b>Total</b>				556	317,070			2,144	395,533
	Family	110	55	554	317,060	120	50	2,144	395,533
	Genus	181	82	475	198,485	207	65	2,144	395,533
	Species	314 ( <b>35</b> )	86 ( <b>10</b> )	428	185,728	410 ( <b>42</b> )	79 ( <b>6</b> )	2,144	395,533



**Figure 1 A.** Proportion of reads assigned to each of the target classes with the three tested assignment methods, namely BLAST<sup>®</sup> against a custom-designed database (Bc), BLAST<sup>®</sup> against the GenBank nt database (Bnt), and the *ecotag* tool from the OBITools suite (E). The number of reads assigned to Ascidiacea and Gymnolaemata are too low to be noticeable. **B.** Number of species identified with each method for the four target classes (Ascidiacea: As, Bivalvia: Bi, Gastropoda: Ga, Gymnolaemata: Gy). The proportion of NIS (dark colour) and native species (light colour) is indicated.

## b. Identification of non-indigenous species

Of the 42 NIS of interest for which reference sequences were available, 12 were detected with at least one marker using our custom-designed database (Table 2). Most of them belonged to ascidians (4) and bivalves (4), whereas two were gastropods and two bryozoans. Most species exhibit a pelagic phase with either long-lived (5) or short-lived (6) larvae. The only identified species with no pelagic larval stage is *Crepipatella dilatata* (Lamarck, 1822). Among these 12 NIS, nine were already reported in the study area and two had no previous record. In addition, *Ruditapes philippinarum* was not present in the Roscoff fauna inventories but has been observed recently. The 12 NIS were identified by at least one unique variant with 100% identity (99% cover) to a reference sequence, except three with 99% identity with an 18S reference sequence, namely *Crassostrea gigas*, *Mercenaria mercenaria* (Linnaeus, 1758) and *Crepipatella dilatata*.

Ten out of the 12 NIS were identified based on 18S. Among these ten species, six were detected by 18S only. Three of them (the ascidians *Asterocarpa humilis* (Heller, 1878) and *Corella eumyota* Traustedt, 1882, and the gastropod *Crepipatella dilatata*) were unable to be amplified by the COI primer pair (Table 2, Figure S3). *Botrylloides violaceus* Oka, 1927 and *B. diegensis* Ritter and Forsyth, 1917 were both detected by COI but could not be identified to the species level with 18S. However, three reads corresponding to three unique variants were assigned to the genus *Botrylloides*. These *Botrylloides* sequences were observed in only one replicate of each of two sampling dates (Feb-13 and Nov-13). Those of February 2013 were congruent with the observation of COI reads assigned to *B. diegensis*, whereas those of November 2013 were not. Conversely, no 18S reads assigned to *Botrylloides* were identified in March and May 2012, when COI reads were assigned to *B. violaceus*. Six species were identified with COI, four of which also identified with 18S.

**Table 2** Non-indigenous species (NIS) in the four target classes detected by 18S, COI or both in at least one plankton sample collected in the bay of Morlaix. For each marker, the number of unique variants is given, with the number of reads in parentheses. For each NIS, the type of dispersal mode (Dispersal) is indicated with short and long disperser describing species with a life cycle including a pelagic larval stage lasting less or more than 2 days, respectively. 'Reported' indicates if the species has previously been reported in the study bay. For each marker, the total number (in bold) of reference sequences, retrieved from public databases or produced locally (in parenthesis), available in the custom-designed database is specified ( $N_{ref}$ ). Individual DNA was tested for amplification failure with the COI primers (see Figure S3 for amplification results). In case of COI amplification failure, N/A was added to the COI detection column.

Class	Species	18S detection	COI detection	Dispersal	Reported	$N_{ref}$ 18S	$N_{ref}$ COI
Ascidiacea	<i>Asterocarpa humilis</i>	1(54)	N/A	short disperser	yes	<b>1</b> (1)	<b>4</b> (2)
	<i>Botrylloides diegensis</i>	0 <sup>a</sup>	1(4)	short disperser	yes	<b>1</b> (1)	<b>2</b> (2)
	<i>Botrylloides violaceus</i>	0 <sup>a</sup>	1(14)	short disperser	yes	<b>1</b> (1)	<b>5</b> (1)
	<i>Corella eumyota</i>	1(6)	N/A	short disperser	yes	<b>1</b> (1)	<b>11</b> (11)
Bivalvia	<i>Crassostrea gigas</i>	1(3)	0 <sup>a</sup>	long disperser	yes	<b>4</b> (1)	<b>70</b> (1)
	<i>Mercenaria mercenaria</i>	1(319)	0	long disperser	no	<b>5</b> (-)	<b>8</b> (-)
	<i>Mya arenaria</i>	18(5,839)	8(711)	long disperser	yes	<b>2</b> (1)	<b>12</b> (-)
	<i>Ruditapes philippinarum</i>	14(32,677)	8(20,936)	long disperser	No <sup>b</sup>	<b>2</b> (1)	<b>71</b> (-)
Gastropoda	<i>Crepidula fornicata</i>	9(988)	25(10,757)	long disperser	yes	<b>2</b> (-)	<b>68</b> (50)
	<i>Crepidatella dilatata</i>	1(3)	N/A	direct developer	no	<b>2</b> (2)	<b>35</b> (3)
Gymnolaemata	<i>Bugula neritina</i>	2(129)	2(46)	short disperser	yes	<b>1</b> (1)	<b>11</b> (1)
	<i>Watersipora subatra</i>	1(3)	0	short disperser	yes	<b>1</b> (1)	<b>3</b> (3)

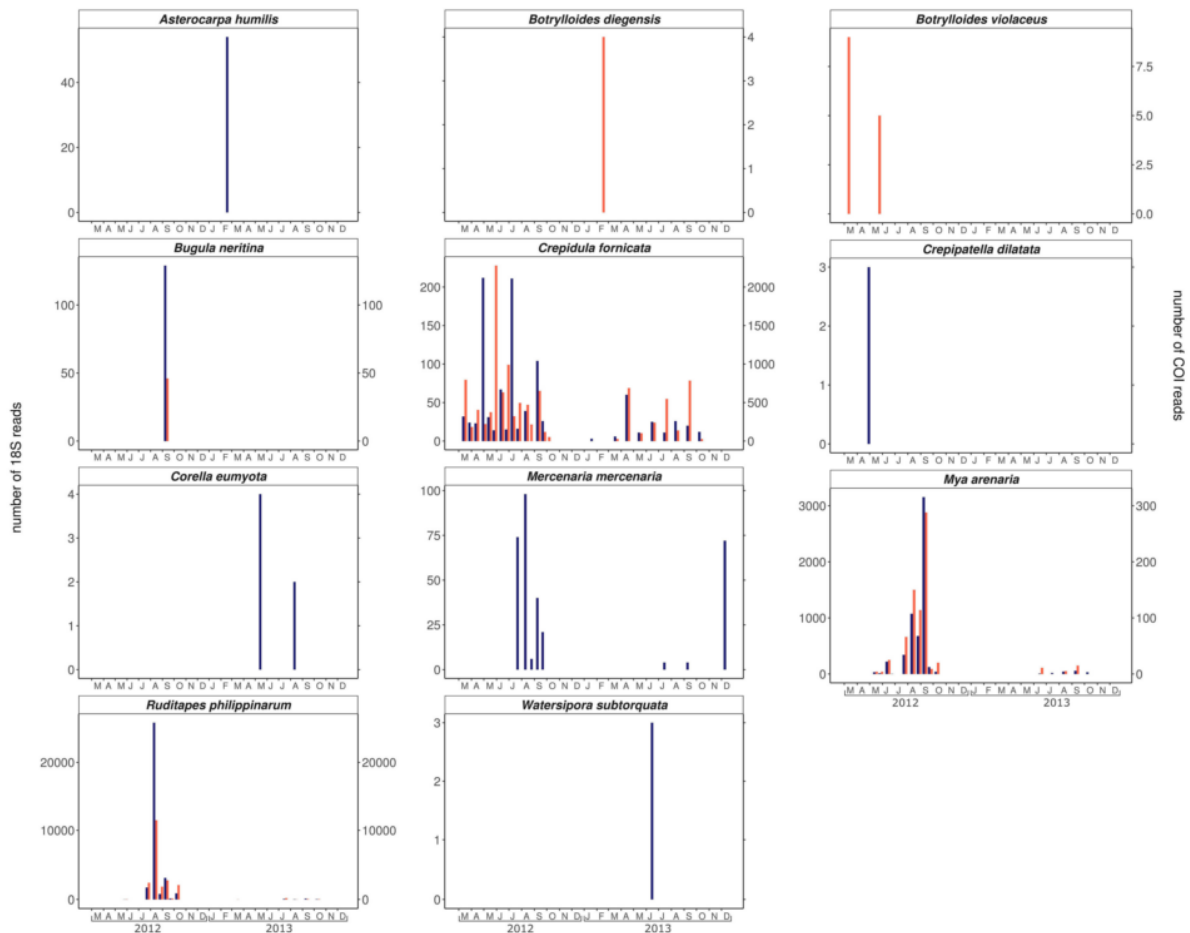
The Manila clam *Ruditapes philippinarum* was the most represented NIS with 32,677 and 20,936 assigned reads for 18S and COI, respectively (Table 2). It was also the most abundant among all bivalves (native and non-indigenous) representing 13% (18S) and 63% of reads assigned to this class. A large number of COI reads (10,757) were also assigned to the slipper limpet *Crepidula fornicata*, whereas only 203 reads were assigned to this species with 18S. Except for the Pacific oyster *Crassostrea gigas*, the species with long-lived larvae (spending on average 2–5 weeks in the water column) were associated with a large number of reads, ranging from 319 to 32,677 over all samples. Conversely, the six species with short-lived larvae and the species with direct development were represented by a small number of reads (less than 130 per species over all samples; Table 2).

With the two alternative assignment methods, the number of detected NIS was 25% to 50% lower than the one achieved with our custom-designed database

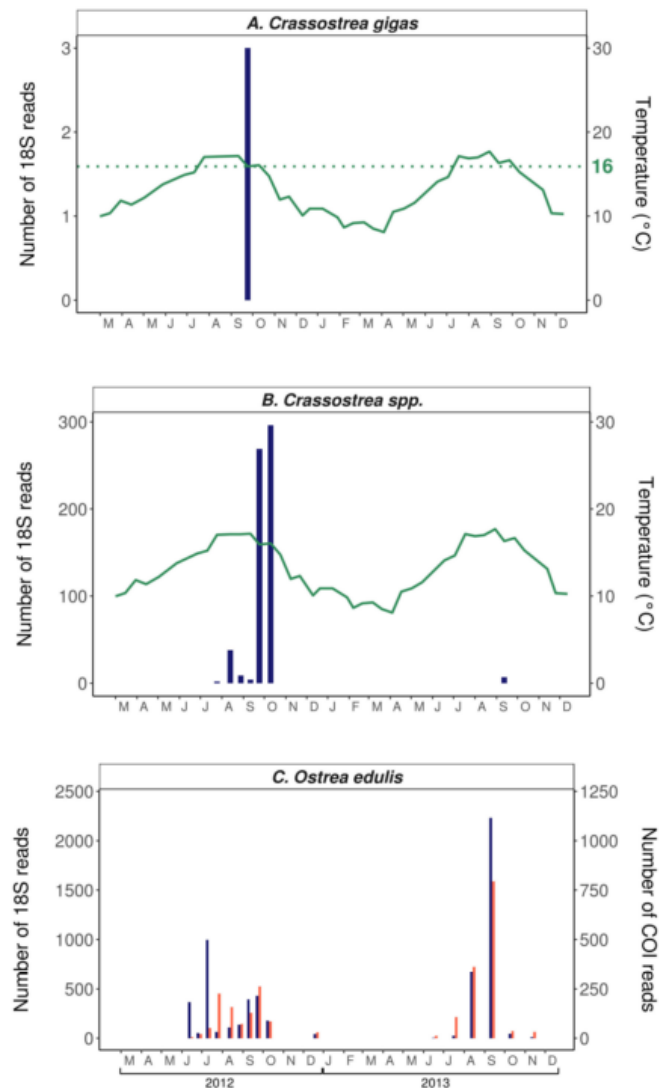


(Figure 1B), and no extra NIS were identified. The use of *ecotag* brought the lowest number of NIS (6) with five species assigned with 18S (*A. humilis*, *C. eumyota*, *C. fornicata*, *Mya arenaria* Linnaeus, 1758, and *R. philippinarum*) and four with COI (*Bugula neritina* (Linnaeus, 1758), *C. fornicata*, *M. arenaria*, and *R. philippinarum*). As compared to *ecotag*, the BLAST® approach against the GenBank nt database brought the same six NIS as well as *C. gigas* and *Mercenaria mercenaria* with 18S and *B. violaceus* with COI.

### c. Temporal variations



**Figure 2** Distribution of reads across all sampling dates for 11 of the 12 non-indigenous species identified with either 18S (blue, left axis), or COI (red, right axis). The results for *Crassostrea gigas* are presented in Figure 3.



**Figure 3** Distribution of reads across all sampling dates for two oyster species detected in the dataset, the non-indigenous Pacific oyster *Crassostrea gigas* (A) and the native European flat oyster *Ostrea edulis* (C). Distribution of reads assigned to *Crassostrea* spp. (and presumably *C. gigas*) is also indicated (B). The number of 18S reads (blue) for each sampling date is shown on the left axis while the number of COI reads (red) is represented on the right axis in C. The green curve in A and B represents the variations in sea surface temperature measured with a CTD probe (Seabird SBE19) at each zooplankton sampling event. Data for mid-May, mid-June and August 2012 were not available.

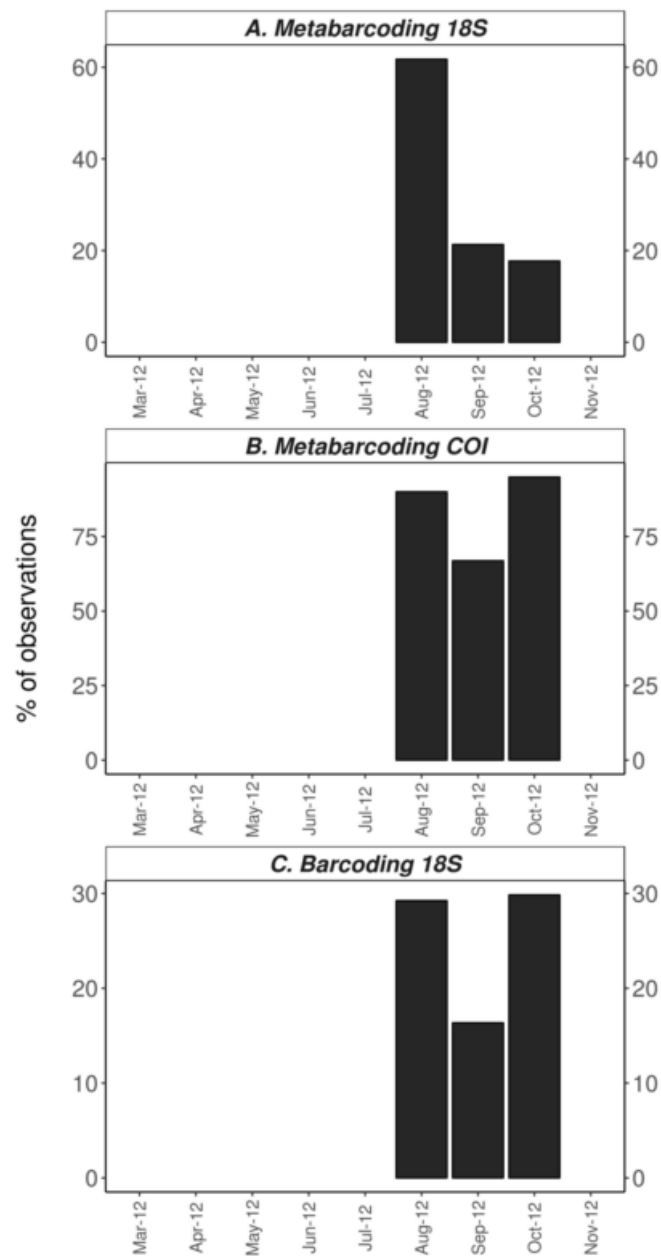
For the soft-shell clam *M. arenaria*, the Manila clam *R. philippinarum* and the slipper limpet *C. fornicata*, the temporal window over which the species were detected was the same for 18S and COI (Figure 2). In addition, for these taxa as well as the hard-shell clam *M. mercenaria*, identified by 18S only, the read distribution exhibited a seasonal pattern. They occurred in the plankton mainly during summer and autumn (e.g. July to October for *R. philippinarum*) and were almost absent in

winter except for *C. fornicata* which was detected from March to October each year. These results were in contrast to those obtained from direct developers and short dispersers which were observed in a single or few samples (e.g. February 2013 for the ascidians *A. humilis* and *B. diegensis*, Figure 2). The only long-disperser species with no clear seasonal pattern was the Pacific oyster *C. gigas*, which was identified with only twelve 18S reads in a single sample in September 2012 (Figure 3A), and no COI reads. Note however that 628 18S reads assigned to the genus *Crassostrea*, which could presumably be assigned to *C. gigas*, were observed from July to October 2012, and in September 2013 (Figure 3B). As a comparison, its native counterpart, the European flat oyster *Ostrea edulis* Linnaeus, 1758, was identified with thousands of reads (18S: 5,751; COI: 2,316) from June to October each year (Figure 3C).

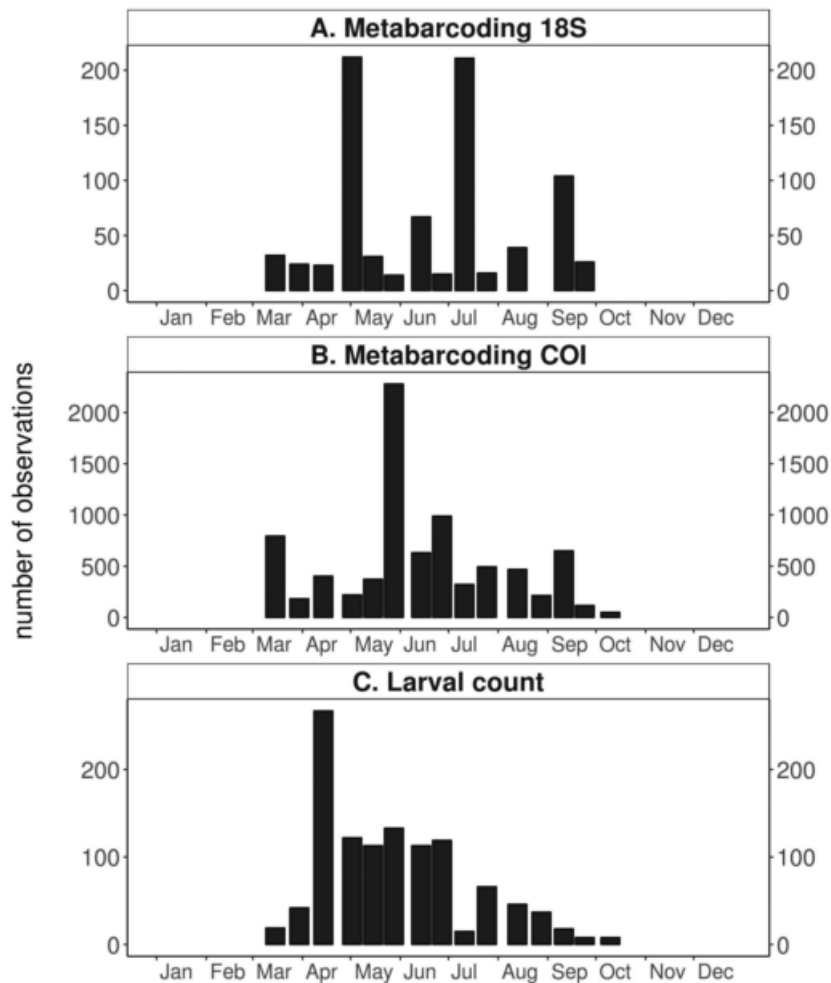
#### d. Comparison of the metabarcoding results with other datasets

The metabarcoding results were compared with data obtained through either barcoding of individual larvae (Manila clam *R. philippinarum*) or morphological identification of larvae (slipper limpet *C. fornicata*). Whatever the method considered, *R. philippinarum* was observed in three samples only, namely August, September and October 2012 (Figure 4). The 18S reads assigned to the Manila clam *R. philippinarum* accounted for between 18% and 62% of the number of reads assigned to bivalves in the samples where the species occurred (Figure 4A). This proportion was even higher when considering COI, ranging from 67% to 95% across samples (Figure 4B). These temporal variations matched well those displayed by the number of larvae identified using traditional barcoding, even if in this case the proportion of clam larvae was lower, ranging from 16% to 30% (Figure 4C).

In the metabarcoding datasets *C. fornicata* was identified in all samples from March to October 2012 with both markers (except the 28.08.12 and 10.10.12 with 18S; Figure 5A, B), which was congruent with the observation of larvae from this species based on morphological identification over the same year (Figure 5C). No correlation between larval counts and the number of reads was observed, neither for 18S ( $r = -0.07$ ,  $p = 0.81$ ) nor COI ( $r = 0.29$ ,  $p = 0.29$ ).



**Figure 4** Relative abundance of Manila clam (*Ruditapes philippinarum*) reads ('Metabarcoding' plots, A, 18S, and B, COI) or larvae ('Barcoding' plot, C) within bivalves, for nine samples collected in 2012. For the metabarcoding data (A and B), the number of reads assigned to *R. philippinarum* was divided by the number of reads assigned to the class Bivalvia. For individual barcoding data (C), the number of larvae identified as *R. philippinarum* based on the amplification of the 18S marker was divided by the number of identified bivalve larvae.



**Figure 5** Distribution of reads assigned to the slipper limpet (*Crepidula fornicata*) for 18S (A) and COI (B) within samples from the year 2012. **C.** Temporal variations in the number of larvae of this species counted in samples collected at the same sampling dates in 2012.

## Discussion

DNA metabarcoding is predicted as a promising approach to identify NIS at all steps of the introduction process (e.g. Comtet et al. 2015; Darling et al. 2017; Zaiko et al. 2018), and this study provides evidence supporting this statement. We examined zooplankton samples collected over 22 consecutive months, targeting larvae of benthic NIS. By using two DNA markers, 18S (nuclear DNA) and COI (mitochondrial DNA), and a custom-designed database, we identified 12 NIS among the four targeted taxonomic groups (Table 2). Our method proved to be efficient in detecting both long and short dispersers. Moreover, three of the identified species had never been recorded in the study area and could be novel introductions,

previously unnoticed ancient introductions, or false positives. Finally, seasonal variations inferred from read distributions of several species were consistent with their known reproductive periods.

### a. Zooplankton metabarcoding: an efficient tool to detect non-indigenous benthic species

Among the four classes studied, 42 NIS were targeted, present either in the study bay or in neighbouring areas, notably along the French and English Western English Channel coasts. Twelve of them (29%) were identified using our approach. However, only 17 species out of 42 were previously reported in the study bay, and one of them (*Ciona robusta* Hoshino and Tokioka, 1967) was observed only once in 2012 and never observed since then despite regular surveys (Bouchemousse et al. 2017). Thus, 75% of these 16 expected NIS were actually detected. Zooplankton metabarcoding thus appears as a powerful tool for NIS detection provided that multiple markers and a custom-designed database are used.

Only four of the 12 NIS were identified with both COI and 18S. This result agrees with previous studies which highlighted the usefulness of combining multiple markers in metabarcoding approaches to detect NIS (Borrell et al. 2017; Zhang et al. 2018). The detection success was higher with 18S (10 species) as compared to COI (6 species). COI is commonly used for (meta)barcoding studies (Bucklin et al. 2011; Comtet et al. 2015), and has been suggested as the barcode of choice for metazoans (Andújar et al. 2018; Hebert et al. 2003). This is in part due to its high polymorphism allowing the identification at the species level, but also to the availability of a large number of reference sequences, representing species from all over the world (Porter and Hajibabaei 2018). However, the lack of conserved primer binding sites can cause amplification failure and taxonomic bias that may prevent its use for broad biodiversity surveys (Clarke et al. 2014; Deagle et al. 2014). Out of the 24 NIS tested for amplification success, one fourth could not be amplified ( $n = 3$ ) or had a weak signal ( $n = 3$ ; Figure S3), which might have prevented their detection. Furthermore, PCR failure did not apply consistently across all species of a family and did not appear to be correlated with phylogeny (Table S1, Figure S3). This makes predicting the lack of detection of a given taxon very challenging. Moreover, additional PCR biases can arise when DNAs from several species are competing with one another, so obtaining amplicons from individual DNA does not ensure the correct detection of a species in

a mixture. Proposed ways of preventing these issues are the design of specific primers, targeting a particular taxonomic group, which could be used in multiplexes to increase taxonomic coverage (Kelly et al. 2017; Zhang et al. 2018), or the combination of COI with more conserved markers. The 18S gene has been suggested as another barcode for high-throughput sequencing-based assessment of biodiversity (e.g. Abad et al. 2017; Zhan et al. 2014). It is highly conserved across species, allowing the design of universal sets of primers targeting different regions and a broad range of taxa, as illustrated in Figure S3. However, this lack of variability may impair taxonomic resolution. In our study, several closely-related species could not be differentiated with this marker (e.g. species from the genus *Botrylloides*). Again, taxonomic groups are not affected in the same way. Out of the six NIS identified solely with 18S (Table 2), three (*Corella eumyota*, *Asterocarpa humilis* and *Watersipora subatra* (Ortmann, 1890)) are efficiently discriminated by this marker (Figures S4 and S5). As they have all been reported in the study area, we assumed that they were actually present in our samples. Thus 18S can be a useful complement to COI, but it is critical to verify its discriminating power to avoid drawing erroneous conclusions about the presence of a NIS.

Our ability to identify a species not only depends on the amplification efficiency and discrimination power of the chosen markers, but also on the quality of the reference sequences. Building a manually-checked custom-designed reference database (with good quality sequences and reliable identification) comprising hundreds of species is time-consuming, and one might wonder if the reliability of the results obtained is worth the effort. Taxonomic assignment is commonly based on public databases, like the nucleotide collection of GenBank (nt), which are recognized to include mistakes (e.g. Harris 2003), with risks of false assignments or failure to detect target NIS. One common example of such risks is the description of a new species within a species complex. This is the case of the tunicates *Ciona intestinalis* (Linnaeus, 1767) and *C. robusta*, which were reported in our study area (Bouchemousse et al. 2016b). Formerly known as *C. intestinalis* type B and type A, respectively (Zhan et al. 2015), they were recently reclassified as *C. intestinalis* and *C. robusta* based on molecular and morphological evidences (Gissi et al. 2017, and references therein). No reliable assignment to the species level would have been possible since some references obtained before the taxonomic revision are still falsely attributed to *C. intestinalis* (e.g. AK173399.1). More specialized and curated databases (e.g. BOLD, Silva) must be preferred for species identification, if references are available for the targeted taxa. Besides, the completion of the reference database is a key condition for an accurate detection of NIS (Briski et al. 2016; Comtet et al. 2015)

and producing new reference sequences for known or potential NIS is worth the effort. In our results, three NIS were not identified when using BLAST® with GenBank nt (*Crepidatella dilatata*, *Watersipora subtorquata* (d'Orbigny 1852), and *Botrylloides diegensis*). No public reference was available for our targeted markers, and they were identified based on our locally-produced references. Another species, *Bugula neritina*, was also not detected with the 18S marker when using the nt public database. Divergent lineages often exist within accepted species (Pante et al. 2015), and species complexes have been uncovered in many invasive species, such as *Bugula neritina* (Fehlauer-Ale et al. 2014) or *Watersipora* spp. (Mackie et al. 2012). Obtaining local references is therefore important to ensure correct assignment, particularly when using a threshold value. For example, the sequences we produced locally for *B. neritina* diverged by 3% from the one available in GenBank (AF499749.1), which most likely originated from a Chinese population. This difference, higher than the threshold chosen for assignment with 18S, prevented the identification of this NIS when using only public references. Finally, the alignment method used for taxonomic assignment matters, as shown here by using two alignment tools (BLAST® versus *ecotag* command from OBITools). *ecotag* required building a composite reference database composed solely of full-length sequences, which lead to a reduction in the number of available references and the number of species with references. By using this approach, six NIS were lost as compared to the BLAST® procedure against the custom-designed database. Although particularly adapted to biodiversity studies as it allows assignment at higher taxonomic levels (e.g. family) and not only at the species level, *ecotag* thus appeared less efficient here because the available references (including some of our locally-produced reference sequences) comprised incomplete (i.e. shorter) sequences.

## **b. An efficient method to detect both long and short dispersers**

Zooplankton metabarcoding has proved efficient in identifying species with both a long and a short pelagic larval duration. Amongst the 16 expected NIS, only three (*Crepidula fornicata*, *Crassostrea gigas*, and *Mya arenaria*) release long-lived planktotrophic larvae. These species are expected to be more easily detected in low-frequency sampling strategies because of their extended pelagic duration. All three were detected. The 13 other NIS with a previous record either have short-lived lecithotrophic larvae (larvae that do not spend more than 48h in the water column), or are direct-developers (one species, *Tritia neritea* (Linnaeus, 1758)). Six of these, all



short dispersers, were identified (two bryozoans and four ascidians; Table 2). This demonstrated the efficiency of the method to detect species releasing short-lived larvae, as also observed by Stefanni et al. (2018). Note however that their detection does not necessarily imply the presence of larvae in our samples, but may result from remains of benthic individuals (e.g. fragments of branching bryozoans), rafting colonies (e.g. Worcester 1994 for colonial ascidians), or post-settlement life stages resuspended from the bottom (e.g. Hamel et al. 2019; Valanko et al. 2010).

Besides amplification failures (Figure S3), our inability to detect the remaining seven already reported short-dispersive or direct-developing NIS might, in part, be explained by the mismatch between their biological features and our sampling strategy. Most of the undetected NIS are commonly reported and/or preferentially found on artificial substrates of the nearby marinas or on ropes and cages in aquaculture facilities (e.g. Airoidi et al. 2015; Bishop et al. 2015a, b; Bouchemousse et al. 2016a; Glasby et al. 2007; Simkanin et al. 2012), including in the study bay (Authors, *personal observations*, Laurent Lévêque, Station Biologique of Roscoff, *personal communication*). The scarcity of this type of artificial substrate at our sampling site might explain their absence in this part of the bay. Moreover, short-lived larvae of these species might not disperse far enough to reach the sampling site. An alternative explanation is provided by the example of the tunicate *Styela clava* Herdman, 1881. We expected to detect this species since it has often been observed in natural rocky reefs nearby the sampling site. Its weak amplification for COI might have precluded its detection with this marker but some sequences were expected to be assigned to *Styela clava* with 18S. One possible explanation could be the discrepancy between sampling time and the reproductive behaviour of the species, since it has been shown to spawn at the end of the day, with a maximum abundance of tadpole larvae in the middle of the following day (Bourque et al. 2007). This suggests that adapting the sampling time to existing knowledge of the target NIS reproductive features, especially for those with short-lived larvae, and periodic and short spawning events, should be considered to improve the detection capacity of the approach.

### c. Detection of previously unreported species

Three species, two with long-lived larvae and one direct-developer, with no previous local record were identified in this study. Among them, the Manila clam

*Ruditapes philippinarum*, was detected in 12 samples with either one of the markers. Since data from individual barcoding confirmed the presence of its larvae in additional samples collected at the same time as our metabarcoding data (August to October 2012; Figure 4), we are confident that the detection of this species is not the result of a false positive. Reads assigned to *R. philippinarum* were numerous, and dominated those assigned to bivalves (60–80%, depending on the marker; Figure 4), which questioned the origin of these larvae. This species was imported to France for aquaculture purposes (Flassch and Leborgne 1994), but is not cultivated in the study bay, although some trials have been conducted in the 1970's (Flassch 1988). Some individuals have been observed but no substantial population has been reported until now (authors' personal observations). The larvae could have originated from an unknown farmed population in the bay or from an overlooked local established population. They could also come from neighbouring populations (e.g. Caill-Milly et al. 2014) or from a more distant source. The study bay has a ferry terminal with regular connections with ports from northern Spain and southern England, which may favor larval transport and release *via* ballast water. In particular the Spanish ports of Bilbao and Santander, connected to Roscoff in the bay, are known to harbour *R. philippinarum* (Bidegain et al. 2015; Zorita et al. 2013).

For the two other unreported species that have been detected (*Crepidatella dilatata* and *Mercenaria mercenaria*), doubt may be raised. *M. mercenaria* was only detected with 18S despite the lack of PCR failure with COI (Figure S3). *M. mercenaria* is a bivalve of commercial interest, imported from North America to South Brittany in the 20<sup>th</sup> century for aquaculture purposes (Marteil 1956). It has then been reported in several parts of Europe, the closest to our study bay being in Southampton (Ansell 1963) and in the Gulf of Morbihan where it appears to have successfully established (Gouilletquer et al. 2002). It may also have reached the nearer bay of Brest (C. Paillard, University of Brest, *personal communication*). It is thus possible that this species has arrived in the bay of Morlaix, at least as dispersing larvae, especially since the periods of detection are in agreement with its known reproductive period (Ansell et al. 1964; Ansell and Lander 1967). However, the phylogenetic tree of the family Veneridae (Figure S6) showed that the 18S marker used does not allow reliable discrimination between *M. mercenaria* and closely-related species (e.g. *Dosinia corrugata* (Reeve, 1850) or *Clausinella fasciata* (da Costa, 1778)), some being present in the bay of Morlaix. Focusing on the direct-developer *Crepidatella dilatata*, the results are even more challenging. Only three reads corresponding to the same unique variant were assigned to this species with 99.45% identity to the closest 367 bp 18S reference sequence (produced locally). However, the identity percentage with *Crepidula*

*fornicata*, a NIS already reported and particularly abundant in the study bay (Rigal et al. 2010) reached 98.9%, a high value just below our selected threshold for 18S (99%). In addition, *C. dilatata* is native to the SE Pacific and has been reported as an introduced species only along the Atlantic coasts of northern Spain (Collin et al. 2009), which makes the presence of *C. dilatata* in our samples unlikely. Additional markers would be needed to ascertain the presence *vs.* absence of the two NIS cited above. A candidate for further study is the mitochondrial gene 16S, which seems to provide a good balance between taxonomic resolution and detection breadth for the study of marine invertebrates (Kelly et al. 2016, 2017).

#### d. Read counts as a proxy for larval abundance and NIS reproductive success

Detecting reads of a specific benthic NIS in a zooplankton metabarcoding dataset most likely indicates the presence of its larvae, although an alternative origin (shed cells, mucus, see above) could not be discarded. When resulting from the presence of larvae, read occurrence would be indicative of the reproductive status of populations that have most likely established in the vicinity of the sampling area. In such cases, the temporal distribution of reads can further allow to investigate the species reproductive dynamics. In the three benthic NIS with long-lived larvae and large number of reads (i.e. *M. arenaria*, *R. philippinarum*, *C. fornicata*), such distribution suggested a seasonal pattern congruent with their known reproductive cycle. For example, our results, with both markers, suggested that larvae of the soft-shell clam *Mya arenaria* occurred in the plankton from May to October 2012, with two distinct periods, one in May–June, and the other in July–September (Figure 2). This result is in agreement with its known reproductive periods in various parts of its distribution range (either native or introduced; e.g. Brousseau 1987; Cross et al. 2012; Warwick and Price 2009; Winther and Gray 1985). Results from 2013 showed a similar pattern, with lower read numbers. This might reflect interannual variability in the reproduction of *Mya arenaria*, as observed in other populations (e.g. Strasser and Günther 2001).

The match between the reproductive periods inferred from our metabarcoding data and those known for the targeted species is ascertained when looking at the well-studied slipper limpet (*Crepidula fornicata*). This species has been established at our sampling point for decades (Dupont et al. 2003; Le Cam and Viard 2011; Rigal et al. 2010). Mean concentrations of *C. fornicata* larvae at our sampling

location averaged monthly between 2005 and 2011 indicated that larvae occurred from March to October (Figure S7; Leroy 2011; Rigal 2009). These observations are congruent with our metabarcoding results (Figure 5; Figure S7). Further, the mean distribution of COI reads associated with *C. fornicata* was similar to the mean distribution of larval concentrations (Figure S7). This is in agreement with the results obtained in the meta-analysis of Lamb et al. (2019) who found that read counts loosely correspond to relative occurrence of species in the samples. However, some discrepancies are visible. In both datasets an abundance peak is observed in May, whereas a second peak is observed in September or in August, based on the metabarcoding and morphological datasets, respectively. In addition, when comparing larval abundances of the slipper limpet *C. fornicata* based on morphological identification to the number of reads assigned to this species in 2012 (Figure 5), no correlation was found with either marker or even between both markers. This result is congruent with what has been found in other studies (e.g. Elbrecht and Leese 2015; Pochon et al. 2013; Sun et al. 2015) and compels us to nuance our data interpretation. Many elements can explain why we did not observe a correlation between read and larval counts for *C. fornicata*. For instance, the comparison between read abundances and individual counts can be arguable since the size of individuals will influence the quantity of DNA available after extraction, and thus their relative abundance in metabarcoding assays (Elbrecht and Leese 2015; Elbrecht et al. 2017; Harvey et al. 2017). Also, the presence of potential primer biases in the DNA mixture would add even more discrepancies in the read count *vs.* species abundance correlation, as demonstrated by Elbrecht and Leese (2015) or Piñol et al. (2019). Since 18S did not exhibit any amplification failure (Figure S3), it is expected to display a better correlation with biomass as demonstrated by Clarke et al. (2017) in zooplankton samples. However, our results did not show a better signal with this marker.

The mechanisms cited above might also explain the lack of correlation between data collected for both 18S and COI in the case of *C. fornicata*, or the discrepancies observed between metabarcoding and single-larva barcoding for the Manila clam *R. philippinarum* (Figure 4). Our data illustrated a clear seasonal distribution (Figure 2) in the range of the known reproductive periods over this species' introduced distribution range (Laruelle et al. 1994). However, single-larva barcoding showed that 30% of bivalve larvae were assigned to this species, whereas its relative proportion with metabarcoding reached 60–80%. The experimental design might also contribute to this lack of correlation, such as different extraction efficiencies, the normalisation of DNA concentrations prior to sequencing, or the type

of sequencing platform. These points should be taken into account in further studies (Kelly et al. 2014; Lamb et al. 2019).

It is important to point that all the above statements rely on the assumption that traditional methods are devoid of any biases and give accurate estimates of species presence and abundance, which metabarcoding should match with. However, they have their own limitations and biases that may contribute to the discrepancies reported (e.g. Kelly et al. 2017). For instance, in the case of *C. fornicata*, the samples used for the two approaches were collected with different nets with a different mesh size. Because of the size of *C. fornicata* larvae (from 400 µm; Pechenik and Lima 1984), we did not expect a major effect of the mesh size on larval abundance, although we cannot reject this hypothesis to explain the observed discrepancy. In addition, because samples with the two nets were usually collected with a fifteen-minute interval, we cannot discard the possibility of short-term variability in larval abundance between the samples.

As the only long disperser for which no seasonal pattern was recovered, the case of the Pacific oyster *Crassostrea gigas* is interesting. This species was identified by a low number of reads, in September 2012 (Figure 3). This result was confirmed by the single-larva barcoding approach (only five larvae of *C. gigas* were observed, one in August, and four in October; data not shown), which suggested that almost no larvae were present. This finding was unexpected since numerous oyster farms are located in the bay of Morlaix, producing more than 5,000 tons of *C. gigas* per year (Comité Régional Conchylicole – Bretagne Nord, *personal communication*). Furthermore, a previous work reported the presence of Pacific oyster larvae at a station located 6 km from our sampling site in August and September 2007 (Philippart et al. 2012). Although *C. gigas* has become an invader in natural habitats in many places along the North and South Brittany coasts, it is absent from natural habitats in the Bay of Morlaix (Le Berre et al. 2009). A likely explanation is that the temperatures prevailing in the bay prevent either spawning or larval survival for this species. Several studies have shown that the proliferation of *C. gigas* occurred when summer temperatures were high (Diederich et al. 2005; Dutertre et al. 2010), and that spawning requires a threshold temperature of 18–20 °C (Dutertre et al. 2009, 2010; Enríquez-Díaz et al. 2008). Further, Rico-Villa et al. (2009) showed that larval development is optimal above 22 °C and is slowed at 17 °C, even if larvae can survive at 15 °C (His et al. 1989). The temperatures observed in the bay of Morlaix at the time of sampling did not exceed 18 °C (Figure 3), and could have limited the spawning of farmed Pacific oysters and the survival and development of produced larvae. Another

possible explanation is that the larvae are exported off the bay, a hypothesis suggested to explain the low proliferation of *C. fornicata* at this same location (Rigal et al. 2010). Finally, the increase in the proportion of infertile triploid Pacific oyster in French production might have contributed to the low number of larvae produced by local farming. In contrast, the native flat oyster (*Ostrea edulis*), which only accounts for a low percentage of the produced oysters in the bay, has been identified with a high number of reads in both datasets (Figure 3). The seasonal pattern displayed by metabarcoding results is in agreement with the known reproductive period of this species (Eagling et al. 2018), which is known to spawn at temperatures lower than those required by *C. gigas* (Mann 1979). Compared to its Pacific counterpart, the flat oyster (*Ostrea edulis*) seems to be able to reproduce in the bay, and the larvae might either be coming from nearby oyster farms or from a local established natural population.

## Conclusion

DNA metabarcoding from bulk zooplankton samples is effective to detect benthic NIS. Overall our results suggested that the use of a custom-designed database combined to a BLAST®-based alignment was the more efficient approach for NIS detection and to avoid false assignments. This pinpoints the need for NIS-dedicated databases as advocated by Darling et al. (2017) or Dias et al. (2017), especially since using metabarcoding as a tool to detect NIS is becoming more popular. Focusing on the larval stage allowed us to go beyond the simple presence of a given NIS and helped us to, firstly, evaluate the reproductive success of the identified species and, secondly, to show that one notorious and conspicuous NIS (*C. gigas*) apparently do not produce larvae in the study area. In addition, for the long dispersers where enough reads were obtained, a reproductive window could be defined. Interestingly it was always consistent with the known breeding period of the species concerned. In this context, the use of DNA metabarcoding on zooplankton surveys would be a valuable tool to support surveillance programmes. Despite these findings, our study also highlighted important points to take into account in future studies, including sampling frequency (to increase the likelihood of detecting short dispersers) or the importance of using multiple and complementary markers.

## Acknowledgements

The authors would like to thank John Bishop, Charlotte Roby, François Rigal, Fanny Leroy, and Sophie Delerue-Ricard, for making their samples, sequences and/or data available. We are also thankful to the crew of the Neomysis and to the Diving and Marine core service, at Station Biologique of Roscoff, for their help during sampling. We also would like to thank the ABIMS Platform for providing access calculation resources needed to perform the data analyses. This project was supported jointly by the Interreg IVa Marinexus programme and the Fondation TOTAL (project Aquanis2.0). The sequencing of individual larvae was conducted at the Centre National de Séquençage (Genoscope, Evry, France) in the framework of the programme Bibliothèque du Vivant, jointly supported by the CNRS, the MNHN and the INRA. MC acknowledges a PhD grant by Région Bretagne (ENIGME ARED project) and Sorbonne Université.

## References

- Abad D, Albaina A, Aguirre M, Laza-Martínez A, Uriarte I, Iriarte A, Villate F, Estonba A (2016) Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. *Marine Biology* 163(7): 149
- Abad D, Albaina A, Aguirre M, Estonba A (2017) 18S V9 metabarcoding correctly depicts plankton estuarine community drivers. *Marine Ecology Progress Series* 584: 31-43
- Airoldi L, Turon X, Perkol-Finkel S, Rius M (2015) Corridors for aliens but not for natives: effects of marine urban sprawl at a regional scale. *Diversity and Distributions* 21(7): 755-768
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215(3): 403-410
- Andújar C, Arribas P, Yu DW, Vogler AP, Emerson BC (2018) Why the COI barcode should be the community DNA metabarcode for the metazoa. *Molecular Ecology* 27(20): 3968-3975
- Ansell AD (1963) *Venus mercenaria* (L) In Southampton Water. *Ecology* 44(2): 396-397
- Ansell AD, Lander KF (1967) Studies on the Hard-Shell Clam, *Venus mercenaria*, in British Waters. III. Further Observations on the Seasonal Biochemical Cycle and on Spawning. *Journal of Applied Ecology* 4(2): 425-435
- Ansell AD, Lander KF, Coughlan J, Loosmore FA (1964) Studies on the Hard-Shell Clam, *Venus mercenaria*, in British Waters. I. Growth and Reproduction in Natural and Experimental Colonies. *Journal of Applied Ecology* 1(1): 63-82
- Appeltans W, Ah Yong Shane T, Anderson G, Angel Martin V, Artois T, Bailly N, Bamber R, Barber A, Bartsch I, Berta A, Błażewicz-Paszkowycz M, Bock P, Boxshall G, Boyko Christopher B, Brandão Simone N, Bray Rod A, Bruce Niel L, Cairns Stephen D, Chan T-Y, Cheng L, Collins Allen G, Cribb T, Curini-Galletti M, Dahdouh-Guebas F, Davie Peter JF, Dawson Michael N, De Clerck O, Decock W, De Grave S, de Voogd Nicole J, Domning Daryl P, Emig Christian C, Erséus C, Eschmeyer W, Fauchald K, Fautin Daphne G, Feist Stephen W, Franssen Charles HJM, Furuya H, Garcia-Alvarez O, Gerken S, Gibson D, Gittenberger A, Gofas S, Gómez-Daglio L, Gordon Dennis P, Guiry Michael D, Hernandez F, Hoeksema Bert W, Hopcroft Russell R, Jaume D, Kirk P, Koedam N, Koenemann S, Kolb Jürgen B, Kristensen Reinhardt M, Kroh A, Lambert G, Lazarus David B, Lemaitre R, Longshaw M, Lowry J, Macpherson E, Madin Laurence P, Mah C, Mapstone G, McLaughlin Patsy A, Mees J, Meland K, Messing Charles G, Mills Claudia E, Molodtsova Tina N, Mooi R, Neuhaus B, Ng Peter KL, Nielsen C, Norenburg J, Opresko Dennis M, Osawa M, Paulay G, Perrin W, Pilger John F, Poore Gary CB, Pugh P, Read Geoffrey B, Reimer James D, Rius M, Rocha Rosana M, Saiz-Salinas José I, Scarabino V, Schierwater B,

- Schmidt-Rhaesa A, Schnabel Kareen E, Schotte M, Schuchert P, Schwabe E, Segers H, Self-Sullivan C, Shenkar N, Siegel V, Sterrer W, Stöhr S, Swalla B, Tasker Mark L, Thuesen Erik V, Timm T, Todaro MA, Turon X, Tyler S, Uetz P, van der Land J, Vanhoorne B, van Ofwegen Leen P, van Soest Rob WM, Vanaverbeke J, Walker-Smith G, Walter TC, Warren A, Williams Gary C, Wilson Simon P, Costello Mark J (2012) The magnitude of global marine species diversity. *Current Biology* 22(23): 2189-2202
- Ardura A, Zaiko A, Martinez JL, Samulioviene A, Semenova A, Garcia-Vazquez E (2015) eDNA and specific primers for early detection of invasive species – A case study on the bivalve *Rangia cuneata*, currently spreading in Europe. *Marine Environmental Research* 112: 48-55
- Bayne BL, Ahrens M, Allen SK, D'auriac MA, Backeljau T, Beninger P, Bohn R, Boudry P, Davis J, Green T, Guo X, Hedgecock D, Ibarra A, Kingsley-Smith P, Krause M, Langdon C, Lapègue S, Li C, Manahan D, Mann R, Perez-Paralle L, Powell EN, Rawson PD, Speiser D, Sanchez J-L, Shumway S, Wang H (2017) The proposed dropping of the genus *Crassostrea* for all pacific cupped oysters and its replacement by a new genus *Magallana*: a dissenting view. *Journal of shellfish research* 36(3): 545-547
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013) GenBank. *Nucleic Acids Research* 41(Database issue): D36-42
- Bidegain G, Bárcena JF, García A, Juanes JA (2015) Predicting coexistence and predominance patterns between the introduced Manila clam (*Ruditapes philippinarum*) and the European native clam (*Ruditapes decussatus*). *Estuarine, Coastal and Shelf Science* 152: 162-172
- Bishop J, Wood C, Yunnice A, Griffiths C (2015a) Unheralded arrivals: non-native sessile invertebrates in marinas on the English coast. *Aquatic Invasions* 10: 249-264
- Bishop JDD, Wood CA, Lévêque L, Yunnice ALE, Viard F (2015b) Repeated rapid assessment surveys reveal contrasting trends in occupancy of marinas by non-indigenous species on opposite sides of the western English Channel. *Marine Pollution Bulletin* 95(2): 699-706
- Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU, Richardson DM (2011) A proposed unified framework for biological invasions. *Trends in Ecology & Evolution* 26(7): 333-339
- Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E (2005) Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360(1462): 1935-43
- Borrell YJ, Miralles L, Do Huu H, Mohammed-Geba K, Garcia-Vazquez E (2017) DNA in a bottle—Rapid metabarcoding survey for early alerts of invasive species in ports. *PLoS ONE* 12(9): e0183347
- Bouchemousse S, Bishop JDD, Viard F (2016a) Contrasting global genetic patterns in two biologically similar, widespread and invasive *Ciona* species (Tunicata, Ascidiacea). *Scientific Reports* 6: 24875
- Bouchemousse S, Lévêque L, Dubois G, Viard F (2016b) Co-occurrence and reproductive synchrony do not ensure hybridization between an alien tunicate and its interfertile native congener. *Evolutionary Ecology* 30(1): 69-87
- Bouchemousse S, Lévêque L, Viard F (2017) Do settlement dynamics influence competitive interactions between an alien tunicate and its native congener? *Ecology and Evolution* 7(1): 200-213
- Bourque D, Davidson J, MacNair NG, Arsenault G, LeBlanc AR, Landry T, Miron G (2007) Reproduction and early life history of an invasive ascidian *Styela clava* Herdman in Prince Edward Island, Canada. *Journal of Experimental Marine Biology and Ecology* 342(1): 78-84
- 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
- Briski E, Ghabooli S, Bailey SA, MacIsaac HJ (2016) Are genetic databases sufficiently populated to detect non-indigenous species? *Biological Invasions* 18(7): 1911-1922



- Brousseau DJ (1987) A comparative study of the reproductive cycle of the soft-shell clam, *Mya arenaria* in Long Island sound. *Journal of Shellfish Research* 6(1): 7-15
- Brown EA, Chain FJJ, Zhan A, MacIsaac HJ, Cristescu ME (2016) Early detection of aquatic invaders using metabarcoding reveals a high number of non-indigenous species in Canadian ports. *Diversity and Distributions* 22(10): 1045-1059
- Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA Barcoding of Marine Metazoa. *Annual Review of Marine Science* 3(1): 471-508
- Bucklin A, Lindeque PK, Rodriguez-Ezpeleta N, Albaina A, Lehtiniemi M (2016) Metabarcoding of marine zooplankton: prospects, progress and pitfalls. *Journal of Plankton Research* 38(3): 393-400
- Caill-Milly N, Bru N, Barranger M, Gallon L, D'amico F (2014) Morphological trends of four Manila clam populations (*Venerupis philippinarum*) on the French Atlantic coast: identified spatial patterns and their relationship to environmental variability. *Journal of Shellfish Research* 33(2): 355-372
- Carlton JT, Geller JB (1993) Ecological roulette: the global transport of nonindigenous marine organisms. *Science* 261(5117): 78-82
- Chain FJJ, Brown EA, MacIsaac HJ, Cristescu ME (2016) Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. *Diversity and Distributions* 22(5): 493-504
- Clarke LJ, Soubrier J, Weyrich LS, Cooper A (2014) Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources* 14(6): 1160-1170
- Clarke LJ, Beard JM, Swadling KM, Deagle BE (2017) Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. *Ecology and Evolution* 7(3): 873-883
- Collin R, Farrell P, Cragg S (2009) Confirmation of the identification and establishment of the South American slipper limpet *Crepidatella dilatata* (Lamarck 1822) (Caenogastropoda: Calyptraeidae) in Northern Spain. *Aquatic Invasions* 4(2): 377-380
- Comtet T, Sandionigi A, Viard F, Casiraghi M (2015) DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions* 17(3): 905-922
- Connell SD (2001) Urban structures as marine habitats: an experimental comparison of the composition and abundance of subtidal epibiota among pilings, pontoons and rocky reefs. *Marine Environmental Research* 52(2): 115-125
- Cowen RK, Gawarkiewicz G, Pineda J, Thorrold SR, Werner FE (2007) Population connectivity in marine systems an overview. *Oceanography* 20(3): 14-21
- Cristescu ME (2014) From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution* 29(10): 566-571
- Cross ME, Lynch S, Whitaker A, O'Riordan RM, Culloty SC (2012) The reproductive biology of the softshell clam, *Mya arenaria*, in Ireland, and the possible impacts of climate variability. *Journal of Marine Biology* 2012: 9
- Darling JA, Tepolt CK (2008) Highly sensitive detection of invasive shore crab (*Carcinus maenas* and *Carcinus aestuarii*) larvae in mixed plankton samples using polymerase chain reaction and restriction fragment length polymorphisms (PCR-RFLP). *Aquatic Invasions* 3(2): 141-152
- Darling JA, Galil BS, Carvalho GR, Rius M, Viard F, Piraino S (2017) Recommendations for developing and applying genetic tools to assess and manage biological invasions in marine ecosystems. *Marine Policy* 85: 54-64

- Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters* 10(9): 20140562
- Deagle BE, Clarke LJ, Kitchener JA, Polanowski AM, Davidson AT (2017) Genetic monitoring of open ocean biodiversity: an evaluation of DNA metabarcoding for processing continuous plankton recorder samples. *Molecular Ecology Resources* 18: 391-406, doi:10.1111/1755-0998.12740.
- Dias J, Fotadar S, Munoz J, Hewitt M, Lukehurst S, Hourston M, Wellington C, Duggan R, Bridgwood S, Massam M, Aitken V, De Lestang P, McKirdy S, Willan R, Kirkendale L, Giannetta J, Corsini-Foka M, Pothoven S, Gower F, Viard F, Buschbaum C, Scarcella G, Strafella P, Bishop MJ, Sullivan T, Buttino I, Madduppa H, Huhn M, Zabin CJ, Bacela-Spychalska C, Wójcik-Fudalewska D, Markert A, Maximov A, Kautsky L, Jaspers C, Kotta J, Pärnoja M, Robledo D, Tsiamis K, Küpper FC, Žuljević A, McDonald JI, Snow M (2017) Establishment of a taxonomic and molecular reference collection to support the identification of species regulated by the Western Australian Prevention List for Introduced Marine Pests. *Management of Biological Invasions* 8(2): 215-225
- Diederich S, Nehls G, van Beusekom JEE, Reise K (2005) Introduced Pacific oysters (*Crassostrea gigas*) in the northern Wadden Sea: invasion accelerated by warm summers? *Helgoland Marine Research* 59(2): 97-106
- Dunstan PK, Bax NJ (2007) How far can marine species go? Influence of population biology and larval movement on future range limits. *Marine Ecology Progress Series* 344: 15-28
- Dupont L, Jollivet D, Viard F (2003) High genetic diversity and ephemeral drift effects in a successful introduced mollusc (*Crepidula fornicata*: Gastropoda). *Marine Ecology Progress Series* 253: 183-195
- Dutertre M, Beninger PG, Barillé L, Papin M, Rosa P, Barillé A-L, Haure J (2009) Temperature and seston quantity and quality effects on field reproduction of farmed oysters, *Crassostrea gigas*, in Bourgneuf Bay, France. *Aquatic Living Resources* 22(3): 319-329
- Dutertre M, Beninger PG, Barillé L, Papin M, Haure J (2010) Rising water temperatures, reproduction and recruitment of an invasive oyster, *Crassostrea gigas*, on the French Atlantic coast. *Marine Environmental Research* 69(1): 1-9
- Eagling LE, Ashton EC, Jensen AC, Sigwart JD, Murray D, Roberts D (2018) Spatial and temporal differences in gonad development, sex ratios and reproductive output, influence the sustainability of exploited populations of the European oyster, *Ostrea edulis*. *Aquatic Conservation: Marine and Freshwater Ecosystems* 28(2): 270-281
- Elbrecht V, Leese F (2015) Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PLoS ONE* 10(7): e0130324
- Elbrecht V, Peinert B, Leese F (2017) Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecology and Evolution* 7(17): 6918-6926
- Enríquez-Díaz M, Pouvreau S, Chávez-Villalba J, Le Pennec M (2008) Gametogenesis, reproductive investment, and spawning behavior of the Pacific giant oyster *Crassostrea gigas*: evidence of an environment-dependent strategy. *Aquaculture International* 17(5): 491
- Espiñeira M, González-Lavín N, Vieites JM, Santaclara FJ (2009) Development of a method for the genetic identification of commercial bivalve species based on mitochondrial 18S rRNA sequences. *Journal of Agricultural and Food Chemistry* 57(2): 495-502
- Fehlauer-Ale KH, Mackie JA, Lim-Fong GE, Ale E, Pie MR, Waeschenbach A (2014) Cryptic species in the cosmopolitan *Bugula neritina* complex (Bryozoa, Cheilostomata). *Zoologica Scripta* 43(2): 193-205

- Firth LB, Knights AM, Bridger D, Evans AJ, Mieszkowska N, Moore PJ, O'Connor NE, Sheehan EV, Thompson RC, Hawkins SJ (2016) Ocean sprawl: challenges and opportunities for biodiversity management in a changing world. *Oceanography and Marine Biology: An Annual Review* 54: 193-269
- Flassch J-P (1988) La palourde. Dossier d'élevage. Ifremer, Brest, France, 106 pp.
- Flassch J-P, Leborgne Y (1992) Introduction in Europe, from 1972 to 1980, of the Japanese Manila clam (*Tapes philippinarum*) and the effects on aquaculture production and natural settlement. *ICES marine Sciences Symposia* 194: 92-96
- Fonseca V, Carvalho G, Sung W, Johnson HF, Power D, Neill S, Packer M, Blaxter ML, Lamshead PJ, Thomas W, Creer S (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications* 1: 98
- Garland ED, Zimmer CA (2002) Techniques for the identification of bivalve larvae. *Marine Ecology Progress Series* 225: 299-310
- Geller J, Meyer C, Parker M, Hawk H (2013) Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources* 13(5): 851-861
- Gissi C, Hastings KEM, Gasparini F, Stach T, Pennati R, Manni L (2017) An unprecedented taxonomic revision of a model organism: the paradigmatic case of *Ciona robusta* and *Ciona intestinalis*. *Zoologica Scripta* 46(5): 521-522
- Glasby TM, Connell SD, Holloway MG, Hewitt CL (2007) Nonindigenous biota on artificial structures: could habitat creation facilitate biological invasions? *Marine Biology* 151(3): 887-895
- Gouletquer P, Bachelet G, Sauriau PG, Noel P (2002) Open Atlantic coast of Europe — A century of introduced species into French waters. In: Leppäkoski E, Gollasch S, Olenin S (eds), *Invasive Aquatic Species of Europe. Distribution, Impacts and Management* Springer Netherlands, Dordrecht, pp 276-290
- Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra W, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet AL, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research* 41(Database issue): D597-604
- Hamel J-F, Sun J, Gianasi BL, Montgomery EM, Kenchington EL, Burel B, Rowe S, Winger PD, Mercier A (2019) Active buoyancy adjustment increases dispersal potential in benthic marine animals. *Journal of Animal Ecology* 88(6): 820-832
- Harris DJ (2003) Can you bank on GenBank? *Trends in Ecology & Evolution* 18(7): 317-319
- Harvey JBJ, Hoy MS, Rodriguez RJ (2009) Molecular detection of native and invasive marine invertebrate larvae present in ballast and open water environmental samples collected in Puget Sound. *Journal of Experimental Marine Biology and Ecology* 369(2): 93-99
- Harvey JBJ, Johnson SB, Fisher JL, Peterson WT, Vrijenhoek RC (2017) Comparison of morphological and next generation DNA sequencing methods for assessing zooplankton assemblages. *Journal of Experimental Marine Biology and Ecology* 487: 113-126
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences* 270(1512): 313-21
- His E, Robert R, Dinet A (1989) Combined effects of temperature and salinity on fed and starved larvae of the mediterranean mussel *Mytilus galloprovincialis* and the Japanese oyster *Crassostrea gigas*. *Marine Biology* 100(4): 455-463

- Jung SW, Kim HJ, Park JS, Lee T-K, Shin K, Jeong S-Y, Hwang JY, Yoo J-W (2018) Planktonic bivalve larvae identification and quantification in Gomsu Bay, South Korea, using next-generation sequencing analysis and microscopic observations. *Aquaculture* 490: 297-302
- Kelly RP, Port JA, Yamahara KM, Crowder LB (2014) Using environmental DNA to census marine fishes in a large mesocosm. *PLoS ONE* 9(1): e86175
- Kelly RP, O'Donnell JL, Lowell NC, Shelton AO, Samhuri JF, Hennessey SM, Feist BE, Williams GD (2016) Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ* 4: e2444
- Kelly RP, Closek CJ, O'Donnell JL, Kralj JE, Shelton AO, Samhuri JF (2017) Genetic and manual survey methods yield different and complementary views of an ecosystem. *Frontiers in Marine Science* 3: 283
- Lamb PD, Hunter E, Pinnegar JK, Creer S, Davies RG, Taylor MI (2019) How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology* 28(2): 420-430
- Laruelle F, Guillou J, Paulet YM (1994) Reproductive pattern of the clams, *Ruditapes decussatus* and *R. philippinarum* on intertidal flats in Brittany. *Journal of the Marine Biological Association of the United Kingdom* 74(2): 351-366
- Lasota R, Piłczyńska J, Williams ST, Wołowicz M (2013) Fast and easy method for total DNA extraction and gene amplification from larvae, spat and adult mussels *Mytilus trossulus* from the Baltic Sea. *Oceanological and Hydrobiological Studies* 42(4): 486-489
- Le Berre I, Hily C, Lejart M, Gouill R (2009) Analyse spatiale de la prolifération de *C. gigas* en Bretagne. *Cybergeo: European Journal of Geography*, doi:10.4000/cybergeo.22818
- Le Cam S, Viard F (2011) Infestation of the invasive mollusc *Crepidula fornicata* by the native shell borer *Cliona celata*: a case of high parasite load without detrimental effects. *Biological Invasions* 13(5): 1087-1098
- Le Goff-Vitry MC, Chipman AD, Comtet T (2007) In situ hybridization on whole larvae: a novel method for monitoring bivalve larvae. *Marine Ecology Progress Series* 343: 161-172
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology* 10(1): 34
- Leroy F (2011) Influence des conditions trophiques sur le développement larvaire de l'espèce invasive *Crepidula fornicata*: conséquences sur ses capacités de dispersion. PhD Thesis, Université Pierre et Marie Curie, Paris, France
- Lindeque PK, Parry HE, Harmer RA, Somerfield PJ, Atkinson A (2013) Next generation sequencing reveals the hidden diversity of zooplankton assemblages. *PLoS ONE* 8(11): e81327
- Lopez-Escardo D, Paps J, de Vargas C, Massana R, Ruiz-Trillo I, del Campo J (2018) Metabarcoding analysis on European coastal samples reveals new molecular metazoan diversity. *Scientific Reports* 8(1): 9106-9106
- Mackie JA, Darling JA, Geller JB (2012) Ecology of cryptic invasions: latitudinal segregation among *Watersipora* (Bryozoa) species. *Scientific Reports* 2: 871
- Mann R (1979) Some biochemical and physiological aspects of growth and gametogenesis in *Crassostrea gigas* and *Ostrea edulis* grown at sustained elevated temperatures. *Journal of the Marine Biological Association of the United Kingdom* 59(1): 95-110
- Marteil L (1956) Acclimatation du clam (*Venus mercenaria* L.) en Bretagne. *Revue des Travaux de l'Institut des Pêches Maritimes* 20(2): 157-160
- Mileikovsky SA (1971) Types of larval development in marine bottom invertebrates, their distribution and ecological significance: a re-evaluation. *Marine Biology* 10(3): 193-213

- Mohrbeck I, Raupach MJ, Martínez Arbizu P, Knebelsberger T, Laakmann S (2015) High-throughput sequencing—the key to rapid biodiversity assessment of marine metazoa? *PLoS ONE* 10(10): e0140342
- Molnar JL, Gamboa RL, Revenga C, Spalding MD (2008) Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment* 6(9): 485-492
- Nuñez AL, Katsanevakis S, Zenetos A, Cardoso AC (2014) Gateways to alien invasions in the European seas. *Aquatic Invasions* 9(2): 133-144
- Ojaveer H, Galil BS, Campbell ML, Carlton JT, Canning-Clode J, Cook EJ, Davidson AD, Hewitt CL, Jelmert A, Marchini A, McKenzie CH, Minchin D, Occhipinti-Ambrogi A, Olenin S, Ruiz G (2015) Classification of non-indigenous species based on their impacts: considerations for application in marine management. *PLoS Biology* 13(4): e1002130
- Pante E, Puillandre N, Viricel A, Arnaud-Haond S, Aurelle D, Castelin M, Chenuil A, Destombe C, Forcioli D, Valero M, Viard F, Samadi S (2015) Species are hypotheses: avoid connectivity assessments based on pillars of sand. *Molecular Ecology* 24(3): 525-544
- Pechenik JA, Lima GM (1984) Relationship between growth, differentiation, and length of larval life for individually reared larvae of the marine gastropod, *Crepidula fornicata*. *The Biological Bulletin* 166(3): 537-549
- Philippart CJM, Amaral A, Asmus R, van Bleijswijk J, Bremner J, Buchholz F, Cabanellas-Reboredo M, Catarino D, Cattrijsse A, Charles F, Comtet T, Cunha A, Deudero S, Duchêne J-C, Fraschetti S, Gentil F, Gittenberger A, Guizien K, Gonçalves JM, Guarneri G, Hendriks I, Hussel B, Vieira RP, Reijnen BT, Sampaio I, Serrao E, Pinto IS, Thiebaut E, Viard F, Zuur AF (2012) Spatial synchronies in the seasonal occurrence of larvae of oysters (*Crassostrea gigas*) and mussels (*Mytilus edulis/galloprovincialis*) in European coastal waters. *Estuarine, Coastal and Shelf Science* 108: 52-63
- Piñol J, Senar MA, Symondson WOC (2019) The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology* 28(2): 407-419
- Pochon X, Bott NJ, Smith KF, Wood SA (2013) Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. *PLoS ONE* 8(9): e73935
- Porter TM, Hajibabaei M (2018) Over 2.5 million COI sequences in GenBank and growing. *PLoS ONE* 13(9): e0200177
- Pringle JM, Blakeslee AMH, Byers JE, Roman J (2011) Asymmetric dispersal allows an upstream region to control population structure throughout a species' range. *Proceedings of the National Academy of Sciences* 108(37): 15288-15293
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41(Database issue): D590-D596
- R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular Ecology Notes* 7(3): 355-64
- Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics* 16(6): 276-277
- Rico-Villa B, Pouvreau S, Robert R (2009) Influence of food density and temperature on ingestion, growth and settlement of Pacific oyster larvae, *Crassostrea gigas*. *Aquaculture* 287(3): 395-401

- Rigal F (2009) Dynamique spatio-temporelle du nuage larvaire du gastéropode introduit *Crepidula fornicata* au sein d'une baie mégatidale, la baie de Morlaix (France). PhD Thesis, Université Pierre et Marie Curie, Paris, France, 159 pp
- Rigal F, Viard F, Ayata S-D, Comtet T (2010) Does larval supply explain the low proliferation of the invasive gastropod *Crepidula fornicata* in a tidal estuary? *Biological Invasions* 12(9): 3171-3186
- Riquet F, Comtet T, Broquet T, Viard F (2017) Unexpected collective larval dispersal but little support for sweepstakes reproductive success in the highly dispersive brooding mollusc *Crepidula fornicata*. *Molecular Ecology* 26(20): 5467-5483
- Sánchez A, Quinteiro J, Rey-Méndez M, Perez-Martín RI, Sotelo CG (2015) Identification and quantification of two species of oyster larvae using real-time PCR. *Aquatic Living Resources* 27(3-4): 135-145
- Sardain A, Sardain E, Leung B (2019) Global forecasts of shipping traffic and biological invasions to 2050. *Nature Sustainability* 2(4): 274-282
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, Winter M, Arianoutsou M, Bacher S, Blasius B, Brundu G, Capinha C, Celesti-Grapow L, Dawson W, Dullinger S, Fuentes N, Jäger H, Kartesz J, Kenis M, Kreft H, Kühn I, Lenzner B, Liebhold A, Mosena A, Moser D, Nishino M, Pearman D, Pergl J, Rabitsch W, Rojas-Sandoval J, Roques A, Rorke S, Rossinelli S, Roy HE, Scalera R, Schindler S, Štajerová K, Tokarska-Guzik B, van Kleunen M, Walker K, Weigelt P, Yamanaka T, Essl F (2017) No saturation in the accumulation of alien species worldwide. *Nature Communications* 8: 14435
- Shanks AL (2009) Pelagic larval duration and dispersal distance revisited. *The Biological Bulletin* 216(3): 373-385
- Simberloff D, Martin J-L, Genovesi P, Maris V, Wardle DA, Aronson J, Courchamp F, Galil B, García-Berthou E, Pascal M, Pyšek P, Sousa R, Tabacchi E, Vilà M (2013) Impacts of biological invasions: what's what and the way forward. *Trends in Ecology & Evolution* 28(1): 58-66
- Simkanin C, Davidson IC, Dower JF, Jamieson G, Therriault TW (2012) Anthropogenic structures and the infiltration of natural benthos by invasive ascidians. *Marine Ecology* 33(4): 499-511
- Southward AJ, Langmead O, Hardman-Mountford NJ, Aiken J, Boalch GT, Dando PR, Genner MJ, Joint I, Kendall MA, Halliday NC, Harris RP, Leaper R, Mieszkowska N, Pingree RD, Richardson AJ, Sims DW, Smith T, Walne AW, Hawkins SJ (2005) Long-term oceanographic and ecological research in the Western English Channel. *Advances in Marine Biology* 47: 1-105
- Stefanni S, Stanković D, Borme D, de Olazabal A, Juretić T, Pallavicini A, Tirelli V (2018) Multi-marker metabarcoding approach to study mesozooplankton at basin scale. *Scientific Reports* 8(1): 12085
- Strasser M, Günther C-P (2001) Larval supply of predator and prey: temporal mismatch between crabs and bivalves after a severe winter in the Wadden Sea. *Journal of Sea Research* 46(1): 57-67
- Sun C, Zhao Y, Li H, Dong Y, MacIsaac HJ, Zhan A (2015) Unreliable quantitation of species abundance based on high-throughput sequencing data of zooplankton communities. *Aquatic Biology* 24(1): 9-15
- Taris N, Comtet T, Stolba R, Lasbleiz R, Pechenik JA, Viard F (2010) Experimental induction of larval metamorphosis by a naturally-produced halogenated compound (dibromomethane) in the invasive mollusc *Crepidula fornicata* (L.). *Journal of Experimental Marine Biology and Ecology* 393(1): 71-77
- Tepolt CK, Darling JA, Bagley MJ, Geller JB, Blum MJ, Grosholz ED (2009) European green crabs (*Carcinus maenas*) in the northeastern Pacific: genetic evidence for high population connectivity and current-mediated expansion from a single introduced source population. *Diversity and Distributions* 15(6): 997-1009

- Thiriot-Quévieux C, Scheltema RS (1982) Planktonic larvae of New England gastropods. V. *Bittium alternatum*, *Triphora nigrocincta*, *Cerithiopsis emersoni*, *Lunatia heros*, and *Crepidula plana*. *Malacologia* 23: 37-46
- UNESCO (1968) Zooplankton sampling. Monographs on oceanographic methodology. Vol. 2. UNESCO, Paris, 174 pp.
- Valanko S, Norkko A, Norkko J (2010) Strategies of post-larval dispersal in non-tidal soft-sediment communities. *Journal of Experimental Marine Biology and Ecology* 384(1): 51-60
- Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp GH, Geniez P, Pont D, Argillier C, Baudoin J-M, Peroux T, Crivelli AJ, Olivier A, Acqueberge M, Le Brun M, Møller PR, Willerslev E, Dejean T (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology* 25(4): 929-942
- Viard F, Comtet T (2015) 18. Applications of DNA-based methods for the study of biological invasions. In: Canning-Clode J (ed), *Biological Invasions in Changing Ecosystems*. Sciendo Migration, pp 411-435
- Warwick RM, Price R (2009) Macrofauna production in an estuarine mud-flat. *Journal of the Marine Biological Association of the United Kingdom* 55(1): 1-18
- Werner B (1955) Über die anatomie, die entwicklung und biologie des veligers und der veliconcha von *Crepidula fornicata* L. (Gastropoda Prosobranchia). *Helgoländer Wissenschaftliche Meeresuntersuchungen* 5(2): 169-217
- Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York
- Winther U, Gray JS (1985) The biology of *Mya arenaria* (Bivalvia) in the eutrophic inner Oslofjord. *Sarsia* 70(1): 1-9
- Worcester SE (1994) Adult rafting versus larval swimming: dispersal and recruitment of a botryllid ascidian on eelgrass. *Marine Biology* 121(2): 309-317
- WoRMS Editorial Board (2019) World Register of Marine Species (WoRMS). <http://www.marinespecies.org> (accessed 4 May 2018)
- Zaiko A, Samuiloviene A, Ardura A, Garcia-Vazquez E (2015) Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. *Marine Pollution Bulletin* 100(1): 53-59
- Zaiko A, Pochon X, Garcia-Vazquez E, Olenin S, Wood SA (2018) Advantages and limitations of environmental DNA/RNA tools for marine biosecurity: management and surveillance of non-indigenous species. *Frontiers in Marine Science* 5(322)
- Zhan A, Hulák M, Sylvester F, Huang X, Adebayo AA, Abbott CL, Adamowicz SJ, Heath DD, Cristescu ME, MacIsaac HJ (2013) High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution* 4(6): 558-565
- Zhan A, Bailey SA, Heath DD, Macisaac HJ (2014) Performance comparison of genetic markers for high-throughput sequencing-based biodiversity assessment in complex communities. *Molecular Ecology Resources* 14(5): 1049-1059
- Zhan A, Briski E, Bock DG, Ghabooli S, MacIsaac HJ (2015) Ascidiaceans as models for studying invasion success. *Marine Biology* 162(12): 2449-2470
- Zhang GK, Chain FJJ, Abbott CL, Cristescu ME (2018) Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. *Evolutionary Applications* 11(10): 1901-1914
- Zorita I, Solaun O, Borja A, Franco J, Muxika I, Pascual M (2013) Spatial distribution and temporal trends of soft-bottom marine benthic alien species collected during the period 1989–2008 in the Nervión estuary (southeastern Bay of Biscay). *Journal of Sea Research* 83: 104-110

# Conclusion and perspectives



© Wilfried Thomas – SBR





As the number of marine biological introductions is constantly growing, there is a pressing need for effective tools to monitor non-indigenous species (NIS) in both anthropogenic and natural environments. NIS monitoring fundamentally requires methods allowing the rapid and accurate identification of known NIS, and the rapid and accurate detection of potential new arrivals. Ideally, monitoring tools should allow processing numerous samples of various types and origins to cover as many locations, seasons and habitats as possible, in a standardized way. Traditional methods based on the identification of species according to morphological criteria are not always sufficient to detect NIS and have been routinely complemented by molecular techniques such as DNA barcoding. The fast development of HTS approaches now renders their application to biodiversity surveys easier. As such, HTS-based methods might be a good addition to traditional methods for non-indigenous species detection and survey. In this work, I tested this hypothesis, with marine coastal species as case studies, working mainly in marinas and adjacent habitats, which are entry-points of NIS.

The key outcomes of this thesis work concern the potential for NIS detection but also the basic knowledge that can be gained on invasion ecology. Across the different studies, we showed that:

- ❖ HTS can be used on a diversity of marine sample types, aiming to detect NIS: water samples ([Chapters II.1, II.2, and III.1](#)), organisms settled on experimental plates ([Chapter III.1](#)), plankton samples ([Chapter III.2](#)), and preservative ethanol ([Chapter I](#)). The use of preservative ethanol had previously been validated in terrestrial and freshwater samples, but not for marine samples. Preservation time however matters, and samples should be kept no longer than 6 months before being processed.

- ❖ When coupled with taxonomic assignment, HTS-based approaches can detect NIS in the previously cited types of samples. Two thirds of the NIS observed in 10 marinas over two seasons with traditional methods (quadrat surveys in [Chapter II](#)) were recovered by metabarcoding applied to environmental DNA (eDNA) from water samples, an encouraging result for routine NIS surveys. On the other hand, one-third were false-negatives (i.e. species present but not detected by the method), and included important species considering their invasive status, such as the Pacific oyster *Crassostrea gigas* or the colonial tunicate *Didemnum vexillum*. This important caveat should be accounted for in surveillance programs. Metabarcoding approaches also produced false positives (i.e. species thought to be here but actually absent), which is also a serious limitation for NIS surveys, particularly when targeting early detection of novel NIS ([Chapter II](#)).

❖ Thanks to HTS-based studies, NIS were also detected in natural habitats close to a marina (Chapter III), including a pristine (or supposed to be) habitat (e.g. Site MEL in Chapter III.1). Interestingly and unexpectedly, the proportion of NIS detected in natural habitats of the study bay (where numerous human activities are existing, such as leisure boating, aquaculture, fishing etc.) was not much lower than the one from the marina.

❖ HTS without taxonomic assignment (i.e. OTUs or ASVs-based analyses) also provided new insights into several aspects regarding marina community diversity and introduction processes. This approach allowed describing the diversity and structure of the studied communities (Chapter II.2) or populations (Chapter I), with results similar to traditional methods such as *in situ* morphology-based identification (on scrapped quadrats in Chapter II.2), laboratory morphology-based identification (panels in Chapter III.1) or single individual Sanger sequencing (on single zooids in Chapter I). These methods can thus provide effective biomonitoring tools, for instance to rapidly detect shifts in local communities. When using ASVs, these techniques are also promising as population genetics tools, and may feed new indicators related to the whole genetic diversity present in a given area or site.

❖ Applied to plankton samples, metabarcoding allowed identifying NIS at the larval stage (including short dispersers) (Chapter III.2). Not only may this approach help early detection of NIS spreading from neighboring areas (larval stages being the primary actors of natural dispersal for many marine invertebrates), but it may also give insights into the reproductive patterns of established NIS (e.g. no reproduction of the Pacific oyster *C. gigas* in our study area, extended reproduction period of the gastropod *Crepidula fornicata*).

❖ Whatever the objectives (biomonitoring at the community level, genetic diversity assessment, NIS detection), and whatever the sampling strategy (Chapter III.1), the results of this thesis showed that the bioinformatics pipeline and parameters are critical, and their choices are a matter of compromise regarding the objectives.

### 1. Is metabarcoding ready for marine NIS detection?

The use of molecular techniques has greatly improved our ability to detect introduced species. Any scientist without taxonomic expertise, but with the required laboratory skills, can now identify specimens, even when morphological criteria are lacking. It requires, however, that reference sequences are available from voucher specimens identified by expert taxonomists. The added value of molecular techniques

dedicated to active surveillance (i.e. targeting one or a few species; Darling, et al. 2017), such as DNA barcoding (Briski, et al. 2011), PCR-RFLP (Darling and Tepolt 2008), endpoint PCR (Ardura, et al. 2015), qPCR or ddPCR (Dysthe, et al. 2018; Harper, et al. 2018; Wood, et al. 2019) has already been demonstrated. All these techniques, however, require targeting a few numbers of species which are known to be present or expected to arrive in the study area. Contrary to these targeted approaches, DNA metabarcoding can detect a wide range of species, without the need of anticipating their potential presence, and thus might be of great value for investigating trends in NIS distribution (Darling, et al. 2017) and early detection (Trebitz, et al. 2017). This is a reason why metabarcoding studies applied to introduced species detection are flourishing, and extensive testing continues to be performed in order to evaluate their potential use in routine biosurveillance (Duarte, et al. 2021).

In this thesis, and in agreement with previous studies (see Table 2 in the Introduction), DNA metabarcoding has shown promising results for NIS early detection. In addition to potentially pinpointing some misidentifications in previous inventories, as was most likely the case for the two species *Polycarpa tenera* and *Cerithiopsis petanii* detected in several marinas around Brittany in [Chapter II](#), the metabarcoding approach is able to detect truly novel introductions, as might be the case for the gastropod *Haminoea orteai* detected in [Chapter II.1](#). More analyses, based on dedicated field surveys and targeted molecular approaches, are required to validate this hypothesis, and eliminate the possibility of false positives due to the assignment method used in this work (see below). If validated, however, it will make an excellent example in favor of using metabarcoding for species detection. Similarly, the detection of the bivalve *Ruditapes philippinarum* in plankton samples from [Chapter III](#), although less unexpected because specimens of this species were already observed in Brittany, revealed the presence of an established population. The high proportion of *R. philippinarum* larvae in our samples suggested that a high number of reproductive individuals must have been present in the bay of Morlaix although no report of a consequent population has ever been reported in this bay. This finding might not have been possible with traditional methods because bivalve larvae are very difficult or impossible to discriminate based on morphological criteria (Garland and Zimmer 2002). Barcoding of individual larvae allowed confirmation of this result ([Chapter III.2](#)), but this approach is much more time-consuming, and its use routinely might be at the expense of the number of samples to be processed. These few examples are good advocates for the use of DNA metabarcoding for NIS detection but this thesis, as well as other works, also raised many issues which still need to be addressed before rushing into the HTS era.

One of the main concerns when using metabarcoding data is the precision of the assignments, which partly depends on the accuracy of the reference databases. This is a well-known issue which has been raised by many authors (e.g. Harris 2003; Briski, et al. 2016; Viard, et al. 2019) but no satisfying methodology is available to date for accurate species detection. Several “curated” databases have been created, such as BOLD (Ratnasingham and Hebert 2007) or SILVA (Quast, et al. 2013) in the hope of achieving a better reference quality for particular markers or taxa. The number of verifications required to increase their accuracy, however, came at the expense of their completeness (Hestetun, et al. 2020). In this thesis, all taxonomic assignments have been performed by comparison to a restricted database composed of sequences extracted from GenBank in order to get the representation of as many taxa as possible. We chose not to use other databases because most of their public references are also archived in GenBank and because their somewhat better accuracy does not offset their lower completeness (Meiklejohn, et al. 2019; Hestetun, et al. 2020). Despite our choice of favoring completeness over accuracy, some of the species detected in this work can clearly be imputed to a lack of reference. This is the case of *Clavelina meridionalis*, for example, which was found in the dataset produced and analyzed in chapter II. This ascidian was only assigned with 18S, for which no reference sequence was available for the closely related species *C. lepadiformis*, which is native from our study area and conspicuous in the surveyed marinas (found in quadrats). Moreover, this native congener, was, assigned with the two other markers for which references were available. To avoid such problems, a greater effort should be devoted to the generation of local reference sequences which will, in turn, allow a better completeness of public databases. For this thesis, references for more than 50 species were produced across all markers, most of which were not available for at least one of the three targeted genes. The combined efforts of every science team working with metabarcoding will progressively increase the number of taxa represented in public databases and we can hope that metabarcoding assignments will become more and more accurate with time.

In a perfect world, where all possible haplotypes for all existing species would possess a reference sequence in public databases, taxonomic assignment would be very straightforward and would only require the selection of references identical to our query. Sadly, we are far from achieving this ideal and several tools are available to help assigning ASVs or OTUs to the most likely taxon. Among the many choices, three of them were tested during this thesis: ecotag from the OBITOOLS, RDP CLASSIFIER, and a BLAST-based custom method. All these did not perform evenly, and produced different results, associated to particular strengths and biases. The first two were

found to be very conservative and usually had a lower rate of assignment to the species-level, which is particularly appropriate when dealing with markers with low discriminating power, such as 18S, to avoid false positives. They are, however, expected to produce a higher number of false negatives which might be especially problematic when used for NIS detection. Moreover, they are not free of false positives, especially when faced with a lack of reference sequences. For these reasons, we chose to use a BLAST-based approach with percent cover and identity thresholds. We are well aware that this approach is prone to identifying species which are not truly present in our dataset (i.e. false positives) but it was selected to minimize the number of false negatives as much as possible, especially when references are available. With such an approach, all suspicious assignments need to be carefully verified before concluding that a new introduced species was detected. The check list should include 1) verifying the reliability of the reference sequences (associated publication and sources), 2) confirming the presence of references for closely related native species, and 3) evaluating the resolution of the markers used for this taxonomic group (phylogenetic tree). An approach of taxonomic assignment by phylogenetic placement, such as EPA-NG (Barbera, et al. 2018), might be a good alternative to the methods cited above. It requires, however, a robust phylogenetic tree for all taxa included in the study, and this might be particularly difficult to obtain when targeting taxa with a broad taxonomic coverage. The ambition of the project EukRef was to create a reliable tree for all eukaryotes, mostly with ribosomal RNA markers, for its use in metabarcoding with phylogenetic placement tools (del Campo, et al. 2018). Mostly devoted to protists for now, we can hope that its implementation will further increase the accuracy of species identification using metabarcoding.

All issues discussed above as well as problems related to the discriminating power of markers discussed in the second part of [Chapter III](#), and bioinformatics related issues discussed in the first part of [Chapter II](#), are proofs that DNA metabarcoding is not fully ready for NIS detection, and especially for early detection of novel introduced species. In this context, Darling, et al. (2020) have called for being particularly cautious regarding 'incidental detection', when using species list produced by metabarcoding studies that were not specifically designed to detect NIS. This is particularly true for the numerous and increasing metabarcoding studies carried out as global biodiversity surveys. The identification of DNA assigned to a previously unreported species is not yet sufficient for inferring that this taxon is actually present in the study area, but it must serve as a signal to actively look for it by repeating the experiments and using other approaches. This step-by-step process has been implemented in a decision-support tree by Sepulveda, et al. (2020), to help

practitioners and stakeholders take the full benefit of metabarcoding approaches in support of NIS management.

This thesis work illustrated that both false negatives and false positives can be obtained following a metabarcoding study, creating uncertainties, as could be the case with any experimental methods. These uncertainties should however not be used as a disclaimer for the use of this approach in NIS surveys. The limitations and sources of errors are now quite well understood, and may be corrected or reduced in a near future. The use of eDNA and metabarcoding thus seems to be mature enough to help in NIS detection and monitoring, when all the necessary precautions are taken (Darling, et al. 2017; Trebitz, et al. 2017; Sepulveda, et al. 2020).

## 2. Added value of amplification-based HTS approaches in studying the introduction process

When focusing on biological introductions, amplicon high-throughput sequencing approaches, such as DNA metabarcoding, have been mostly tested for NIS detection. As mentioned above, this particular application is still challenging but many other questions can be addressed by these techniques. One of their most popular utilization is the uncovering of diversity patterns that can be done with or without prior taxonomic assignment. Many studies have successfully studied spatial and temporal diversity patterns with HTS-based approaches or metabarcoding data (Deiner, et al. 2017), including in marine environments (e.g. Chain, et al. 2016; Kelly, et al. 2016; Port, et al. 2016; Lacoursière-Roussel, et al. 2018; Bakker, et al. 2019; Jeunen, et al. 2019). In [Chapters II.2](#) and [III.1](#), we were also able to gain insights into spatial and temporal variations in marine assemblages using both assigned and unassigned HTS datasets. For instance, we showed that the marina communities are more spatially structured than we originally expected, and that the extent of biotic homogenization might be less important than previously hypothesized at the scale of our study ([Chapter II.2](#)).

Interestingly, the conclusions drawn from HTS and metabarcoding data were quite similar to those drawn from morphology-based assessments, confirming the potential of HTS techniques for diversity analyses, and a new generation of biomonitoring (Baird and Hajibabaei 2012). The use of DNA metabarcoding offered the possibility to examine a much wider taxonomic range that could have been possible with habitat-targeted and morphology-based observations alone. This is

especially the case when working with eDNA, and more particularly water, because the same sample can contain DNA from multiple habitats, with both benthic (hard- and soft-bottom) and pelagic organisms. Since all components of an ecosystem are connected through complex interactions, their joint assessment and analysis will allow recovery of the global dynamics at the ecosystem level. In the case of biological introductions, NIS can indirectly impact a whole ecosystem, and monitoring changes in diversity for every component is particularly valuable. In this context, this thesis work illustrated some particular issues. Firstly, when trying to uncover the place and role of NIS in species assemblages, a prerequisite is to perform a taxonomic assignment, which still faces important challenges (as described in the precedent section). It further requires obtaining reliable data on the status of the detected species which is not an easy task when it concerns hundreds of taxa. More generally, and besides the specific study of NIS contribution to diversity patterns, the choice of the bioinformatics pipeline and data processing can have an effect on diversity indices calculations, as shown by Mächler, et al. (2020) with Hill numbers. This can be particularly problematic for comparison of data not processed in the same manner, further illustrating the need of a standardized approach. The most challenging issues that we encountered in this thesis, however, were related to the primers used for producing the molecular data. We showed that both 16S and COI (except those designed for the target approach in [Chapter I](#)) displayed amplification biases, in particular for ascidian species, which were important taxa to include in this work regarding their substantial contribution to the marina communities. The use of a multi-marker approach somewhat compensated the amplification biases but some species were still not detected. DNA from both ascidians *Ascidella aspersa* and *Ascidella scabra*, for example, could not be discriminated because 18S was the only marker with primers amplifying most tunicate species, including *Ascidella* spp., but the sequences were identical for the two species. Moreover, despite the use of primers designed to target metazoan species, the markers also showed a high amount of non-metazoan amplifications when applied to eDNA. For instance, most of the eDNA datasets for COI and 18S were composed of Chromista and Plantae DNA (see [Chapter II.2](#)). Only approximately one fourth of reads were left for metazoans, and even less for the targeted benthic species. To circumvent this problem, different 18S primers were used to produce the dataset in [Chapter III.1](#), which are targeting the same fragment but supposed to be more specific of metazoan taxa (Sinniger, et al. 2016). The results, however, were not significantly improved with the recovery of only 31% of metazoans. This problem has already been observed for the primers used in this thesis and especially for the ones designed by Leray, et al. (2013) for COI (Stat, et al. 2017; Macher, et al. 2018; Collins, et al. 2019). In addition to eukaryotic non-



specific amplifications, many reads could not be assigned and could correspond to prokaryotic sequences. Siddall, et al. (2009) demonstrated that the primers designed by Folmer, et al. (1994), often used in barcoding studies, and targeting a fragment comprising the one amplified by the primers designed by Leray, et al. (2013), was prone to amplify also marine gamma-Proteobacteria. All of these amplification issues can lead to detection failure in rare target species, and advocate for the use of more specific, less degenerated primers. Since designing such primers for large taxonomic groups can be tricky or even impossible, Corse, et al. (2019) proposed the use of multiplexed primers targeting the same fragment, each primer set being specific of a taxonomic group. In this work, the use of primers specifically designed for amplifying the *Botrylloides* genus (Chapter I) allowed to discriminate the two expected species; one of them (*B. violaceus*) was not recovered with more generalist primers for COI (Chapter II.1), and finally was not identified because it was indistinguishable from other species with 18S and 16S. The design of specific primers might be time-consuming but this step is crucial to get more reliable results and more efforts should be devoted to the development of such approaches in the future.

A further promising application of HTS-based methods in the study of introduction processes is its use for intraspecific diversity assessments. The approach tested in the first chapter of this thesis gave promising results, and allowed us to get an accurate picture of the haplotypic diversity observed in the mock communities created. Here tested on DNA extracted from preservative ethanol, this technique could also be applied to eDNA, from water samples for example, to gain insights into the genetic diversity of a whole population (e.g. Sigsgaard, et al. 2016; Tsuji, et al. 2020), and reconstruct phylogeography patterns for multiple species at once (i.e. comparative phylogeography; Turon, et al. 2020). Such data might be very valuable to identify source populations of introduced species or pathways of introduction. With the combination of different and multiple markers, it might also improve our understanding of the genetic impact of NIS on native populations *via* hybridization for example (Stewart and Taylor 2020). Finally, the relative low cost and ease of water sampling makes it more favorable to regular sampling in order to get time-series data. This would allow detection in time shifts in genetic diversity that could inform us on the selection occurring on NIS during their establishment in a new environment. Even if intraspecific diversity assessments with HTS-based techniques are still in their infancy, their development is quickly rising and they should soon be part of the biomonitoring tools available for introduced species.

In complement of all added values cited above and exemplified in this manuscript, several other applications of HTS-based methods, and more particularly DNA metabarcoding, can be foreseen, and may follow this thesis work. One of them is the study of co-occurring species in marine assemblages. Such information could easily be extracted from metabarcoding data (Laroche, et al. 2018) and could be related to positive or negative interactions between NIS and native species. It could, for example, allow the investigation of biotic parameters responsible for the establishment success or failure of introduced species such as biotic resistance (deRivera, et al. 2005; Leclerc, et al. 2020) or invasional meltdown (Simberloff and Von Holle 1999; O'Loughlin and Green 2017). It could also be used for food web interactions analyses (Zamora-Terol, et al.), and could serve for evaluating the impact of NIS on ecosystem functioning. This approach is, however, very dependent on accurate species detection and identification, which can be difficult to obtain. Nevertheless, detection probability could be improved by the use of multi-species occupancy modelling as suggested by McClenaghan, et al. (2020). We showed in [Chapter II.1](#) that false negatives can be a big problem when applying metabarcoding to eDNA, and occupancy modelling could help take into account detection failures in community-wide species occurrence estimates. Models have already been adapted to take into account the stochasticity of DNA detection in water (Schmidt, et al. 2013) and this approach should be more frequently used in future analyses.

### **3. The future of molecular methods for the study of biological introductions**

All current molecular methods for targeting and monitoring particular species are PCR-based. Their need for a thermocycling process requires that all samples must be treated in a dedicated laboratory by trained personnel. It would be interesting, however, to be able to detect species in the field, by directly analyzing environmental samples, such as water. Several isothermal techniques for DNA amplification are available such as the Recombinase Polymerase Amplification (RPA) which amplifies DNA in approximately 30 min at a constant temperature between 37°C and 42°C (Lobato and O'Sullivan 2018). This approach, coupled with a fluorescent probe for detection, could be, for example, used for the rapid detection, on site, of NIS in ballast water of international ships or in ports. A new protocol using RPA for rapid species detection on site has been developed by coupling fluorescence detection with the CRISPR/Cas technology. It is supposed to enhance the differential detection

of closely related species and can be easily adapted to detect any species. Its use has been successfully tested on eDNA for the detection of *Salmo salar* in Irish rivers (Williams, et al. 2019) and it could become a valuable tool as “early warning” system for management of NIS.

Most current metabarcoding studies use the Illumina® sequencing technology because of its low cost, high throughput, and relatively low error rate. This method, however, does not allow the use of markers longer than 500 bp. This can complicate taxonomic assignment because short markers have less taxonomic and phylogenetic resolution, especially when targeting fragments with a low evolution rate such as 18S. Other sequencing platforms are available such as the Pacific Biosciences Platform (PacBio) or the Oxford Nanopore Technologies (ONT), more dedicated to long-read sequencing. Despite their appealing capacity of sequencing fragments of over 10 kb long, for a fraction of the cost of Sanger sequencing, their use for metabarcoding purposes have been impeded by their high error rate (Quail, et al. 2012). A few studies, however, are developing protocols with a very stringent quality sequence selection to improve the reliability of these new sequencing techniques (e.g. Jamy, et al. 2020) and they could become the new standard for metabarcoding strategies in a near future. It should be noted, however, that increasing fragment length might decrease the detection probability when applied to eDNA because of the presence of many short degraded DNA fragment. Additionally to the fragment length advantage of these technologies, ONT released the MinION platform in 2014, which also offers a low price, portability, and fast sequencing chemistry (Jain, et al. 2015). Several studies are developing bioinformatics pipelines to analyze full-length DNA barcodes with this approach (e.g. Baloğlu, et al. 2020; Santos, et al. 2020) and MinION represents a promising alternative for future biomonitoring programs.

All above techniques are focusing on DNA but other aspects of molecular biology could allow us to learn more about the introduction process. Epigenetic variations, for example, such as DNA methylation, can impact the phenotype produced by the same DNA fragment. In that regard, epigenetic mechanisms could facilitate the establishment and spread of NIS by increasing phenotypic plasticity. Several studies have already evaluated the role of DNA methylation in NIS acclimation to their new habitats (Hawes, et al. 2018 and references therein). They were able to link some methylation signals to environmental conditions (salinity, temperature) or to particular habitats. These results are encouraging, and similar experiments will probably arise in the future to complete our knowledge on NIS

establishment success. Another type of molecular technique which could be of significant help in NIS biomonitoring is the environmental RNA (eRNA). Since it is supposed to degrade more rapidly than DNA in marine water, its detection should be a better proxy for living organisms. This could be of particular interest for NIS detection in ballast water, for example, because only living organisms are a real threat when discharged into a new location (Pochon, et al. 2017). Several studies have evaluated its potential for invasion studies (Zaiko, et al. 2018) and references therein) both for detection and for gaining insights in functional diversity within an ecosystem (Laroche, et al. 2018).

As a conclusion, the rapid evolution in molecular biology practices and technologies is offering a wide range of possibilities for NIS monitoring. All of these methods are very promising but, both currently used approaches and those still in development have their shares of biases and issues that need to be carefully assessed before using their results to establish management practices. Most of these techniques still require heavy machinery and extensive laboratory work, but advances in sequencing technologies will progressively transform the way we use these molecular tools. In a few years, it will most likely be possible to detect any species of interest from water samples, directly on site, by using a portable device which could be plugged into our phones. They will also become more affordable and easy to use which will increase their use for routine monitoring. We can hope that the multiplicity of data collected over the world will allow us to increase our knowledge on biological introductions and lead to remarkable discoveries.

## References from the general introduction and conclusion

- Abad D, Albaina A, Aguirre M, Laza-Martínez A, Uriarte I, Iriarte A, Villate F, Estonba A. (Abad2016 co-authors). 2016. Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. *Marine Biology* 163:149.
- Alberdi A, Aizpurua O, Gilbert MTP, Bohmann K. 2018. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution* 9:134-147.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403-410.
- Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, KjÆR KH, Orlando L, Gilbert MTP, Willerslev E. 2012. Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology* 21:1966-1979.
- Andújar C, Arribas P, Yu DW, Vogler AP, Emerson BC. 2018. Why the COI barcode should be the community DNA metabarcode for the metazoa. *Molecular Ecology* 27:3968-3975.
- Ardura A, Borrell YJ, Fernández S, González Arenales M, Martínez JL, Garcia-Vazquez E. 2020. Nuisance Algae in Ballast Water Facing International Conventions. Insights from DNA Metabarcoding in Ships Arriving in Bay of Biscay. *Water* 12:2168.
- Ardura A, Zaiko A. 2018. PCR-based assay for *Mya arenaria* detection from marine environmental samples and tracking its invasion in coastal ecosystems. *Journal for Nature Conservation* 43:1-7.
- Ardura A, Zaiko A, Borrell YJ, Samuiloviene A, Garcia-Vazquez E. 2017. Novel tools for early detection of a global aquatic invasive, the zebra mussel *Dreissena polymorpha*. *Aquatic Conservation: Marine and Freshwater Ecosystems* 27:165-176.
- Ardura A, Zaiko A, Martinez JL, Samuiloviene A, Semenova A, Garcia-Vazquez E. 2015. eDNA and specific primers for early detection of invasive species – A case study on the bivalve *Rangia cuneata*, currently spreading in Europe. *Marine Environmental Research* 112:48-55.
- Aronson RB, Thatje S, Clarke A, Peck LS, Blake DB, Wilga CD, Seibel BA. 2007. Climate Change and Invasibility of the Antarctic Benthos. *Annual Review of Ecology, Evolution, and Systematics* 38:129-154.
- Azevedo J, Antunes JT, Machado AM, Vasconcelos V, Leão PN, Froufe E. 2020. Monitoring of biofouling communities in a Portuguese port using a combined morphological and metabarcoding approach. *Sci Rep* 10:1-15.
- Azmi F, Hewitt CL, Campbell ML. 2014. A hub and spoke network model to analyse the secondary dispersal of introduced marine species in Indonesia. *ICES Journal of Marine Science* 72:1069-1077.
- Bacher S, Blackburn TM, Essl F, Genovesi P, Heikkilä J, Jeschke JM, Jones G, Keller R, Kenis M, Kueffer C, et al. 2018. Socio-economic impact classification of alien taxa (SEICAT). *Methods in Ecology and Evolution* 9:159-168.
- Baird DJ, Hajibabaei M. 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology* 21:2039-2044.
- Bakker J, Wangensteen OS, Baillie C, Buddo D, Chapman DD, Gallagher AJ, Guttridge TL, Hertler H, Mariani S. 2019. Biodiversity assessment of tropical shelf eukaryotic communities via pelagic eDNA metabarcoding. *Ecology and Evolution* 9:14341-14355.
- Baloğlu B, Chen Z, Elbrecht V, Braukmann T, MacDonald S, Steinke D. 2020. A workflow for accurate metabarcoding using nanopore MinION sequencing. *bioRxiv:2020.2005.2021.108852*.
- Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T, Stamatakis A. 2018. EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Systematic Biology* 68:365-369.
- Bishop JDD, Wood C, Yunnice A, Griffiths C. 2015a. Unheralded arrivals: non-native sessile invertebrates in marinas on the English coast. *Aquatic Invasions* 10:249-264.
- Bishop JDD, Wood CA, Lévêque L, Yunnice ALE, Viard F. 2015b. Repeated rapid assessment surveys reveal contrasting trends in occupancy of marinas by non-indigenous species on opposite sides of the western English Channel. *Marine Pollution Bulletin* 95:699-706.

- Blackburn TM, Bellard C, Ricciardi A. 2019. Alien versus native species as drivers of recent extinctions. *Frontiers in Ecology and the Environment* 17:203-207.
- Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JR, Richardson DM. 2011. A proposed unified framework for biological invasions. *Trends in Ecology & Evolution* 26:333-339.
- Bock DG, Caseys C, Cousens RD, Hahn MA, Heredia SM, Hübner S, Turner KG, Whitney KD, Rieseberg LH. 2015. What we still don't know about invasion genetics. *Molecular Ecology* 24:2277-2297.
- Boettner GH, Elkinton JS, Boettner CJ. 2000. Effects of a Biological Control Introduction on Three Nontarget Native Species of Saturniid Moths. *Conservation Biology* 14:1798-1806.
- Borrell YJ, Miralles L, Do Huu H, Mohammed-Geba K, Garcia-Vazquez E. 2017. DNA in a bottle—Rapid metabarcoding survey for early alerts of invasive species in ports. *PLOS ONE* 12:e0183347.
- Borrell YJ, Miralles L, Martínez-Marqués A, Semeraro A, Arias A, Carleos CE, García-Vázquez E. 2018. Metabarcoding and post-sampling strategies to discover non-indigenous species: A case study in the estuaries of the central south Bay of Biscay. *Journal for Nature Conservation* 42:67-74.
- Bortolus A. 2008. Error Cascades in the Biological Sciences: The Unwanted Consequences of Using Bad Taxonomy in Ecology. *AMBIO: A Journal of the Human Environment* 37:114-118, 115.
- 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* 16:176-182.
- Briski E, Cristescu ME, Bailey SA, MacIsaac HJ. 2011. Use of DNA barcoding to detect invertebrate invasive species from diapausing eggs. *Biological Invasions* 13:1325-1340.
- Briski E, Ghabooli S, Bailey SA, MacIsaac HJ. 2016. Are genetic databases sufficiently populated to detect non-indigenous species? *Biological Invasions* 18:1911-1922.
- Brown EA, Chain FJJ, Zhan A, MacIsaac HJ, Cristescu ME. 2016. Early detection of aquatic invaders using metabarcoding reveals a high number of non-indigenous species in Canadian ports. *Diversity and Distributions* 22:1045-1059.
- Bucklin A, Steinke D, Blanco-Bercial L. 2011. DNA barcoding of marine metazoa. *Annual Review of Marine Science* 3:471-508.
- Bugnot AB, Mayer-Pinto M, Airoldi L, Heery EC, Johnston EL, Critchley LP, Strain EMA, Morris RL, Loke LHL, Bishop MJ, et al. 2020. Current and projected global extent of marine built structures. *Nature Sustainability*.
- Bulleri F, Chapman MG. 2004. Intertidal assemblages on artificial and natural habitats in marinas on the north-west coast of Italy. *Marine Biology* 145:381-391.
- Cahill AE, Pearman JK, Borja A, Carugati L, Carvalho S, Danovaro R, Dashfield S, David R, Féral J-P, Olenin S, et al. 2018. A comparative analysis of metabarcoding and morphology-based identification of benthic communities across different regional seas. *Ecology and Evolution* 8:8908-8920.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581.
- Camatti E, Pansera M, Bergamasco A. 2019. The Copepod *Acartia tonsa* Dana in a Microtidal Mediterranean Lagoon: History of a Successful Invasion. *Water* 11:1200.
- Canning-Clode J, Fofonoff P, Riedel G, Torchin M, Ruiz G. 2011. The Effects of Copper Pollution on Fouling Assemblage Diversity: A Tropical-Temperate Comparison. *PLOS ONE* 6:e18026.
- Capinha C, Essl F, Seebens H, Moser D, Pereira HM. 2015. The dispersal of alien species redefines biogeography in the Anthropocene. *Science* 348:1248-1251.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, et al. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335-336.
- Carlton J. 2001. *Introduced Species in US Coastal Waters: Environmental Impacts and Management Priorities*. Arlington, TX: Pew Oceans Commission.
- Carlton JT. 1996. Biological Invasions and Cryptogenic Species. *Ecology* 77:1653-1655.
- Carlton JT, Geller JB. 1993. Ecological roulette: the global transport of nonindigenous marine organisms. *Science* 261:78-82.
- Cassey P, Delean S, Lockwood JL, Sadowski JS, Blackburn TM. 2018. Dissecting the null model for biological invasions: A meta-analysis of the propagule pressure effect. *PLOS Biology* 16:e2005987.

- Catford JA, Vesk PA, Richardson DM, Pyšek P. 2012. Quantifying levels of biological invasion: towards the objective classification of invaded and invisable ecosystems. *Global Change Biology* 18:44-62.
- Chain FJJ, Brown EA, MacIsaac HJ, Cristescu ME. 2016. Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. *Diversity and Distributions* 22:493-504.
- Clarke Murray C, Pakhomov EA, Therriault TW. 2011. Recreational boating: a large unregulated vector transporting marine invasive species. *Diversity and Distributions* 17:1161-1172.
- Cohen AN, Harris LH, Bingham BL, Carlton JT, Chapman JW, Lambert CC, Lambert G, Ljubenkov JC, Murray SN, Rao LC, et al. 2005. Rapid Assessment Survey for Exotic Organisms in Southern California Bays and Harbors, and Abundance in Port and Non-port Areas. *Biological Invasions* 7:995-1002.
- Collins RA, Bakker J, Wangensteen OS, Soto AZ, Corrigan L, Sims DW, Genner MJ, Mariani S. 2019. Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution* 10:1985-2001.
- Comtet T, Sandionigi A, Viard F, Casiraghi M. (Comtet2015 co-authors). 2015. DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions* 17:905-922.
- Connell SD. 2001. Urban structures as marine habitats: an experimental comparison of the composition and abundance of subtidal epibiota among pilings, pontoons and rocky reefs. *Marine Environmental Research* 52:115-125.
- Cooling M, Hoffmann BD. 2015. Here today, gone tomorrow: declines and local extinctions of invasive ant populations in the absence of intervention. *Biological Invasions* 17:3351-3357.
- Cordier T, Frontalini F, Cermakova K, Apothéoz-Perret-Gentil L, Treglia M, Scantamburlo E, Bonamin V, Pawlowski J. 2019. Multi-marker eDNA metabarcoding survey to assess the environmental impact of three offshore gas platforms in the North Adriatic Sea (Italy). *Marine Environmental Research* 146:24-34.
- Corse E, Meglécz E, Archambaud G, Ardisson M, Martin J-F, Tougard C, Chappaz R, Dubut V. 2017. A from-benchtop-to-desktop workflow for validating HTS data and for taxonomic identification in diet metabarcoding studies. *Molecular Ecology Resources* 17:e146-e159.
- Corse E, Tougard C, Archambaud-Suard G, Agnèse J-F, Messu Mandeng FD, Bilong Bilong CF, Duneau D, Zinger L, Chappaz R, Xu CCY, et al. 2019. One-locus-several-primers: A strategy to improve the taxonomic and haplotypic coverage in diet metabarcoding studies. *Ecology and Evolution* 9:4603-4620.
- Couton M, Comtet T, Le Cam S, Corre E, Viard F. 2019. Metabarcoding on planktonic larval stages: an efficient approach for detecting and investigating life cycle dynamics of benthic aliens. *Management of Biological Invasions* 10:657-689.
- Coutts ADM, Dodgshun TJ. 2007. The nature and extent of organisms in vessel sea-chests: A protected mechanism for marine bioinvasions. *Marine Pollution Bulletin* 54:875-886.
- Darling JA, Carlton JT. 2018. A Framework for Understanding Marine Cosmopolitanism in the Anthropocene. *Frontiers in Marine Science* 5.
- Darling JA, Galil BS, Carvalho GR, Rius M, Viard F, Piraino S. 2017. Recommendations for developing and applying genetic tools to assess and manage biological invasions in marine ecosystems. *Marine Policy* 85:54-64.
- Darling JA, Martinson J, Gong Y, Okum S, Pilgrim E, Lohan KMP, Carney KJ, Ruiz GM. 2018. Ballast water exchange and invasion risk posed by intracoastal vessel traffic: an evaluation using high throughput sequencing. *Environmental Science & Technology* 52:9926-9936.
- Darling JA, Martinson J, Pagenkopp-Lohan K, Carney KJ, Pilgrim E, Banerji A, Holzer KK, Ruiz GM. 2020a. Metabarcoding quantifies differences in accumulation of ballast water borne biodiversity among three port systems in the United States. *Science of The Total Environment*:141456.
- Darling JA, Pochon X, Abbott CL, Inglis GJ, Zaiko A. 2020b. The risks of using molecular biodiversity data for incidental detection of species of concern. *Diversity and Distributions* 26:1116-1121.

- Darling JA, Tepolt CK. 2008. Highly sensitive detection of invasive shore crab (*Carcinus maenas* and *Carcinus aestuarii*) larvae in mixed plankton samples using polymerase chain reaction and restriction fragment length polymorphisms (PCR-RFLP). *Aquatic Invasions* 3:141-152.
- Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, de Vere N, et al. 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology* 26:5872-5895.
- Deiner K, Lopez J, Bourne S, Holman L, Seymour M, Grey EK, Lacoursière A, Li Y, Renshaw MA, Pfrender ME. 2018. Optimising the detection of marine taxonomic richness using environmental DNA metabarcoding: the effects of filter material, pore size and extraction method. *Metabarcoding and Metagenomics* 2:e28963.
- del Campo J, Kolisko M, Boscaro V, Santoferrara LF, Nenarokov S, Massana R, Guillou L, Simpson A, Berney C, de Vargas C, et al. 2018. EukRef: Phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. *PLOS Biology* 16:e2005849.
- deRivera CE, Ruiz GM, Hines AH, Jivoff P. 2005. Biotic resistance to invasion: Native predator limits abundance and distribution of an introduced crab. *Ecology* 86:3364-3376.
- deWaard JR, Landry J-F, Schmidt BC, Derhousoff J, McLean JA, Humble LM. 2009. In the dark in a large urban park: DNA barcodes illuminate cryptic and introduced moth species. *Biodiversity and Conservation* 18:3825-3839.
- Duarte CM, Pitt KA, Lucas CH, Purcell JE, Uye S-i, Robinson K, Brotz L, Decker MB, Sutherland KR, Malej A, et al. 2013. Is global ocean sprawl a cause of jellyfish blooms? *Frontiers in Ecology and the Environment* 11:91-97.
- Duarte S, Vieira PE, Lavrador AS, Costa FO. 2020. Status and prospects of marine NIS detection and monitoring through (e)DNA metabarcoding. *bioRxiv:2020.2005.2025.114280*.
- Duarte S, Vieira PE, Lavrador AS, Costa FO. 2021. Status and prospects of marine NIS detection and monitoring through (e)DNA metabarcoding. *Science of The Total Environment* 751:141729.
- Dupont L, Viard F, David P, Bishop JDD. 2007. Combined effects of bottlenecks and selfing in populations of *Corella eumyota*, a recently introduced sea squirt in the English Channel. *Diversity and Distributions* 13:808-817.
- Dysthe JC, Rodgers T, Franklin TW, Carim KJ, Young MK, McKelvey KS, Mock KE, Schwartz MK. 2018. Repurposing environmental DNA samples—detecting the western pearlshell (*Margaritifera falcata*) as a proof of concept. *Ecology and Evolution* 8:2659-2670.
- Eichmiller JJ, Miller LM, Sorensen PW. 2016. Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Molecular Ecology Resources* 16:56-68.
- Elbrecht V, Vamos EE, Steinke D, Leese F. 2018. Estimating intraspecific genetic diversity from community DNA metabarcoding data. *PeerJ* 6:e4644.
- Ficetola GF, Pansu J, Bonin A, Coissac E, Giguët-Covex C, De Barba M, Gielly L, Lopes CM, Boyer F, Pompanon F, et al. 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources* 15:543-556.
- Firth LB, Knights AM, Bridger D, Evans AJ, Mieszkowska N, Moore PJ, O'Connor NE, Sheehan EV, Thompson RC, Hawkins SJ. 2016. Ocean sprawl: challenges and opportunities for biodiversity management in a changing world. *Oceanography and Marine Biology: An Annual Review* 54:193-269.
- Fletcher LM, Zaiko A, Atalah J, Richter I, Dufour CM, Pochon X, Wood SA, Hopkins GA. (Fletcher2017 co-authors). 2017. Bilge water as a vector for the spread of marine pests: a morphological, metabarcoding and experimental assessment. *Biological Invasions* 19:2851-2867.
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology* 3:294-299.
- Fournier A, Penone C, Pennino MG, Courchamp F. 2019. Predicting future invaders and future invasions. *Proceedings of the National Academy of Sciences* 116:7905-7910.
- Galil BS, Marchini A, Occhipinti-Ambrogi A. 2018. East is east and West is west? Management of marine bioinvasions in the Mediterranean Sea. *Estuarine, Coastal and Shelf Science* 201:7-16.



- Garland ED, Zimmer CA. 2002. Techniques for the identification of bivalve larvae. *Marine Ecology Progress Series* 225:299-310.
- Geller JB, Darling JA, Carlton JT. 2010. Genetic Perspectives on Marine Biological Invasions. *Annual Review of Marine Science* 2:367-393.
- Ghabooli S, Zhan A, Paolucci E, Hernandez MR, Briski E, Cristescu ME, MacIsaac HJ. 2016. Population attenuation in zooplankton communities during transoceanic transfer in ballast water. *Ecology and Evolution* 6:6170-6177.
- Glasby TM, Connell SD, Holloway MG, Hewitt CL. (Glasby2007 co-authors). 2007. Nonindigenous biota on artificial structures: could habitat creation facilitate biological invasions? *Marine Biology* 151:887-895.
- Gollasch S. 2007. Is Ballast Water a Major Dispersal Mechanism for Marine Organisms? In: Nentwig W, editor. *Biological Invasions*. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 49-57.
- Gollasch S. 2006. Overview on introduced aquatic species in European navigational and adjacent waters. *Helgoland Marine Research* 60:84-89.
- Gollasch S, Macdonald E, Belson S, Botnen H, Christensen JT, Hamer JP, Houvenaghel G, Jelmert A, Lucas I, Masson D, et al. 2002. Life in Ballast Tanks. In: Leppäkoski E, Gollasch S, Olenin S, editors. *Invasive Aquatic Species of Europe. Distribution, Impacts and Management*. Dordrecht: Springer Netherlands. p. 217-231.
- Grey EK, Bernatchez L, Cassey P, Deiner K, Deveney M, Howland KL, Lacoursière-Roussel A, Leong SCY, Li Y, Olds B, et al. 2018. Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Sci Rep* 8:8843.
- Günther B, Knebelberger T, Neumann H, Laakmann S, Martínez Arbizu P. 2018. Metabarcoding of marine environmental DNA based on mitochondrial and nuclear genes. *Sci Rep* 8:14822.
- Gurevitch J, Fox GA, Wardle GM, Inderjit, Taub D. 2011. Emergent insights from the synthesis of conceptual frameworks for biological invasions. *Ecology Letters* 14:407-418.
- Gurevitch J, Padilla DK. 2004. Are invasive species a major cause of extinctions? *Trends in Ecology & Evolution* 19:470-474.
- Guzinski J, Ballenghien M, Daguin-Thiébaud C, Lévêque L, Viard F. 2018. Population genomics of the introduced and cultivated Pacific kelp *Undaria pinnatifida*: Marinas—not farms—drive regional connectivity and establishment in natural rocky reefs. *Evolutionary Applications* 11:1582-1597.
- Hajibabaei M, Spall JL, Shokralla S, van Konynenburg S. (Hajibabaei2012 co-authors). 2012. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecology* 12:28.
- Harper LR, Lawson Handley L, Hahn C, Boonham N, Rees HC, Gough KC, Lewis E, Adams IP, Brotherton P, Phillips S, et al. 2018. Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*). *bioRxiv*.
- Harris DJ. 2003. Can you bank on GenBank? *Trends in Ecology & Evolution* 18:317-319.
- Hawes NA, Fidler AE, Tremblay LA, Pochon X, Dunphy BJ, Smith KF. 2018. Understanding the role of DNA methylation in successful biological invasions: a review. *Biological Invasions* 20:2285-2300.
- Hawley W, Reiter P, Copeland R, Pumpuni C, Craig G. 1987. *Aedes albopictus* in North America: probable introduction in used tires from northern Asia. *Science* 236:1114-1116.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences* 270:313-321.
- Hellmann JJ, Byers JE, Bierwagen BG, Dukes JS. 2008. Five Potential Consequences of Climate Change for Invasive Species. *Conservation Biology* 22:534-543.
- Hestetun JT, Bye-Ingebrigtsen E, Nilsson RH, Glover AG, Johansen P-O, Dahlgren TG. 2020. Significant taxon sampling gaps in DNA databases limit the operational use of marine macrofauna metabarcoding. *Marine Biodiversity* 50:70.
- Holman LE, de Bruyn M, Creer S, Carvalho G, Robidart J, Rius M. 2019. Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. *Sci Rep* 9:11559.

- Hudson J, Viard F, Roby C, Rius M. 2016. Anthropogenic transport of species across native ranges: unpredictable genetic and evolutionary consequences. *Biology Letters* 12:20160620.
- Hufbauer RA, Facon B, Ravigné V, Turgeon J, Foucaud J, Lee CE, Rey O, Estoup A. 2012. Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. *Evolutionary Applications* 5:89-101.
- Huhn M, Madduppa HH, Khair M, Sabrian A, Irawati Y, Anggraini NP, Wilkinson SP, Simpson T, Iwasaki K, Setiamarga DH. 2020. Keeping up with introduced marine species at a remote biodiversity hotspot: awareness, training and collaboration across different sectors is key. *Biological Invasions* 22:749-771.
- Ibabe A, Rayón F, Martínez JL, García-Vázquez E. 2020. Environmental DNA from plastic and textile marine litter detects exotic and nuisance species nearby ports. *PLOS ONE* 15:e0228811.
- Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. 2015. Improved data analysis for the MinION nanopore sequencer. *Nat Methods* 12:351-356.
- Jamy M, Foster R, Barbera P, Czech L, Kozlov A, Stamatakis A, Bending G, Hilton S, Bass D, Burki F. 2020. Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. *Molecular Ecology Resources* 20:429-443.
- Jeunen G-J, Knapp M, Spencer HG, Lamare MD, Taylor HR, Stat M, Bunce M, Gemmell NJ. 2019. Environmental DNA (eDNA) metabarcoding reveals strong discrimination among diverse marine habitats connected by water movement. *Molecular Ecology Resources* 19:426-438.
- Katsanevakis S, Gatto F, Zenetos A, Cardoso AC. 2013. How many marine aliens in Europe? *Management of Biological Invasions* 4:37-42.
- Katsanevakis S, Wallentinus I, Zenetos A, Leppäkoski E, Çinar ME, Öztürk B, Grabowski M, Golani D, Cardoso AC. 2014. Impacts of invasive alien marine species on ecosystem services and biodiversity: a pan-European review. *Aquatic Invasions* 9:391-423.
- Kelly RP, O'Donnell JL, Lowell NC, Shelton AO, Samhouri JF, Hennessey SM, Feist BE, Williams GD. 2016. Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ* 4:e2444.
- Klymus KE, Marshall NT, Stepien CA. 2017. Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. *PLOS ONE* 12:e0177643.
- Kozioł A, Stat M, Simpson T, Jarman S, DiBattista JD, Harvey ES, Marnane M, McDonald J, Bunce M. 2018. Environmental DNA metabarcoding studies are critically affected by substrate selection. *Molecular Ecology Resources* 19:366-376.
- Lacoursière-Roussel A, Howland K, Normandeau E, Grey EK, Archambault P, Deiner K, Lodge DM, Hernandez C, Leduc N, Bernatchez L. 2018. eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. *Ecology and Evolution* 8:7763-7777.
- Lamb PD, Hunter E, Pinnegar JK, Creer S, Davies RG, Taylor MI. 2019. How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology* 28:420-430.
- Laroche O, Pochon X, Tremblay LA, Ellis JI, Lear G, Wood SA. 2018. Incorporating molecular-based functional and co-occurrence network properties into benthic marine impact assessments. *FEMS Microbiology Ecology* 94.
- Lavesque N, Hutchings P, Abe H, Daffe G, Gunton LM, Glasby CJ. 2020. Confirmation of the exotic status of *Marphysa victori* Lavesque, Daffe, Bonifácio & Hutchings, 2017 (Annelida) in French waters and synonymy of *Marphysa bulla* Liu, Hutchings & Kupriyanova, 2018. *Aquatic Invasions* 15.
- Lawson Handley LJ, Estoup A, Evans DM, Thomas CE, Lombaert E, Facon B, Aebi A, Roy HE. 2011. Ecological genetics of invasive alien species. *BioControl* 56:409-428.
- Leclerc J-C, Viard F, Brante A. 2020. Experimental and survey-based evidences for effective biotic resistance by predators in ports. *Biological Invasions* 22:339-352.
- Leduc N, Lacoursière-Roussel A, Howland KL, Archambault P, Sevellec M, Normandeau E, Dispas A, Winkler G, McKindsey CW, Simard N, et al. 2019. Comparing eDNA metabarcoding and species collection for documenting Arctic metazoan biodiversity. *Environmental DNA* 1:342-358.
- Lehtiniemi M, Ojaveer H, David M, Galil B, Gollasch S, McKenzie C, Minchin D, Occhipinti-Ambrogi A, Olenin S, Pederson J. 2015. Dose of truth—Monitoring marine non-indigenous species to serve legislative requirements. *Marine Policy* 54:26-35.

- Lenz M, da Gama BAP, Gerner NV, Gobin J, Gröner F, Harry A, Jenkins SR, Kraufvelin P, Mummelthei C, Sareyka J, et al. 2011. Non-native marine invertebrates are more tolerant towards environmental stress than taxonomically related native species: Results from a globally replicated study. *Environmental Research* 111:943-952.
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ. (Leray2013 co-authors). 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology* 10:34.
- Lin Y, Zhan A, Hernandez MR, Paolucci E, MacIsaac HJ, Briski E. 2020. Can chlorination of ballast water reduce biological invasions? *Journal of Applied Ecology* 57:331-343.
- Lobato IM, O'Sullivan CK. 2018. Recombinase polymerase amplification: Basics, applications and recent advances. *TrAC Trends in Analytical Chemistry* 98:19-35.
- Lockwood JL, Cassey P, Blackburn TM. 2009. The more you introduce the more you get: the role of colonization pressure and propagule pressure in invasion ecology. *Diversity and Distributions* 15:904-910.
- López-Legentil S, Legentil ML, Erwin PM, Turon X. (López-Legentil2015 co-authors). 2015. Harbor networks as introduction gateways: contrasting distribution patterns of native and introduced ascidians. *Biological Invasions* 17:1623-1638.
- Macher J-N, Vivancos A, Piggott JJ, Centeno FC, Matthaei CD, Leese F. 2018. Comparison of environmental DNA and bulk-sample metabarcoding using highly degenerate cytochrome c oxidase I primers. *Molecular Ecology Resources* 18:1456-1468.
- Mächler E, Walser J-C, Altermatt F. 2020. Decision making and best practices for taxonomy-free eDNA metabarcoding in biomonitoring using Hill numbers. *bioRxiv:2020.2003.2031.017723*.
- Mack RN, Simberloff D, Mark Lonsdale W, Evans H, Clout M, Bazzaz FA. 2000. Biotic invasions: Causes, epidemiology, global consequences, and control. *Ecological Applications* 10:689-710.
- Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* 3:e1420.
- Mangelsdorf PC. 1983. The Mystery of Corn: New Perspectives. *Proceedings of the American Philosophical Society* 127:215-247.
- Markert A, Wehrmann A, Kröncke I. 2009. Recently established *Crassostrea*-reefs versus native *Mytilus*-beds: differences in ecosystem engineering affects the macrofaunal communities (Wadden Sea of Lower Saxony, southern German Bight). *Biological Invasions* 12:15.
- Marquina D, Esparza-Salas R, Roslin T, Ronquist F. 2019. Establishing arthropod community composition using metabarcoding: Surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. *Molecular Ecology Resources* 19:1516-1530.
- Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC. 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources* 15:28-41.
- McClenaghan B, Compson ZG, Hajibabaei M. 2020. Validating metabarcoding-based biodiversity assessments with multi-species occupancy models: A case study using coastal marine eDNA. *PLOS ONE* 15:e0224119.
- McGlashan DJ, Ponniah M, Cassey P, Viard F. 2008. Clarifying marine invasions with molecular markers: an illustration based on mtDNA from mistaken calyptraeid gastropod identifications. *Biological Invasions* 10:51-57.
- Meiklejohn K, Damaso N, Robertson J. 2019. Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. *PLOS ONE* 14:e0217084.
- Minchin D. 2006. The transport and spread of living aquatic species. In. p. 77-97.
- Minchin D, Gollasch S. 2002. Vectors — How Exotics Get Around. In: Leppäkoski E, Gollasch S, Olenin S, editors. *Invasive Aquatic Species of Europe. Distribution, Impacts and Management*. Dordrecht: Springer Netherlands. p. 183-192.
- Molnar JL, Gamboa RL, Revenga C, Spalding MD. 2008. Assessing the global threat of invasive species to marine biodiversity. *Frontiers in Ecology and the Environment* 6:485-492.

- Morgan EC, Overholt WA. 2005. Potential allelopathic effects of Brazilian pepper (*Schinus terebinthifolius* Raddi, Anacardiaceae) aqueous extract on germination and growth of selected Florida native plants. *The Journal of the Torrey Botanical Society* 132:11-15, 15.
- Naylor RL, Williams SL, Strong DR. 2001. Aquaculture - A Gateway for Exotic Species. *Science* 294:1655-1656.
- Newmaster SGNG, Fazekas AJFJ, Ragupathy SR. 2006. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. *Canadian Journal of Botany* 84:335-341.
- Núñez AL, Katsanevakis S, Zenetos A, Cardoso AC. 2014. Gateways to alien invasions in the European seas. *Aquatic Invasions* 9:133-144.
- Nyberg CD, Wallentinus I. 2005. Can species traits be used to predict marine macroalgal introductions? *Biological Invasions* 7:265-279.
- O'Loughlin LS, Green PT. 2017. Secondary invasion: When invasion success is contingent on other invaders altering the properties of recipient ecosystems. *Ecology and Evolution* 7:7628-7637.
- Ojaveer H, Galil BS, Campbell ML, Carlton JT, Canning-Clode J, Cook EJ, Davidson AD, Hewitt CL, Jelmert A, Marchini A, et al. 2015. Classification of non-indigenous species based on their impacts: considerations for application in marine management. *PLOS Biology* 13:e1002130.
- Ojaveer H, Galil BS, Minchin D, Olenin S, Amorim A, Canning-Clode J, Chainho P, Copp GH, Gollasch S, Jelmert A, et al. 2014. Ten recommendations for advancing the assessment and management of non-indigenous species in marine ecosystems. *Marine Policy* 44:160-165.
- Olden JD, Rooney TP. 2006. On defining and quantifying biotic homogenization. *Global Ecology and Biogeography* 15:113-120.
- Pagenkopp Lohan K, Campbell T, Guo J, Wheelock M, DiMaria R, Geller J. 2019. Intact vs. homogenized subsampling: testing impacts of pre-extraction processing of multi-species samples on invasive species detection. *Management of Biological Invasions* 10.
- Pagenkopp Lohan KM, Fleischer RC, Carney KJ, Holzer KK, Ruiz GM. 2016. Amplicon-Based Pyrosequencing Reveals High Diversity of Protistan Parasites in Ships' Ballast Water: Implications for Biogeography and Infectious Diseases. *Microbial Ecology* 71:530-542.
- Pagenkopp Lohan KM, Fleischer RC, Carney KJ, Holzer KK, Ruiz GM. 2017. Molecular characterisation of protistan species and communities in ships' ballast water across three U.S. coasts. *Diversity and Distributions* 23:680-691.
- Pauvert C, Buée M, Laval V, Edel-Hermann V, Fauchery L, Gautier A, Lesur I, Vallance J, Vacher C. 2019. Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal ecology* 41:23-33.
- Pérez-Portela R, Arranz V, Rius M, Turon X. 2013. Cryptic speciation or global spread? The case of a cosmopolitan marine invertebrate with limited dispersal capabilities. *Sci Rep* 3:3197-3197.
- Petersen KS, Rasmussen KL, Heinemeier J, Rud N. 1992. Clams before Columbus? *Nature* 359:679-679.
- Petri B, Chaganti SR, Chan P-S, Heath D. 2019. Phytoplankton growth characterization in short term MPN culture assays using 18S metabarcoding and qRT-PCR. *Water Research* 164:114941.
- Piola RF, Johnston EL. 2009. Comparing differential tolerance of native and non-indigenous marine species to metal pollution using novel assay techniques. *Environmental Pollution* 157:2853-2864.
- Pochon X, Bott NJ, Smith KF, Wood SA. 2013. Evaluating detection limits of next-generation sequencing for the surveillance and monitoring of international marine pests. *PLOS ONE* 8:e73935.
- Pochon X, Zaiko A, Fletcher LM, Laroche O, Wood SA. 2017. Wanted dead or alive? Using metabarcoding of environmental DNA and RNA to distinguish living assemblages for biosecurity applications. *PLOS ONE* 12:e0187636.
- Pochon X, Zaiko A, Hopkins GA, Banks JC, Wood SA. 2015. Early detection of eukaryotic communities from marine biofilm using high-throughput sequencing: an assessment of different sampling devices. *Biofouling* 31:241-251.
- Porco D, Decaëns T, Deharveng L, James SW, Skarżyński D, Erséus C, Butt KR, Richard B, Hebert PDN. 2013. Biological invasions in soil: DNA barcoding as a monitoring tool in a multiple taxa survey targeting European earthworms and springtails in North America. *Biological Invasions* 15:899-910.

- Port JA, O'Donnell JL, Romero-Maraccini OC, Leary PR, Litvin SY, Nickols KJ, Yamahara KM, Kelly RP. 2016. Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology* 25:527-541.
- Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13:341.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41:D590-D596.
- Ragueneau O, Chauvaud L, Leynaert A, Thouzeau G, Paulet Y-M, Bonnet S, Lorrain A, Grall J, Corvaisier R, Le Hir M, et al. 2002. Direct evidence of a biologically active coastal silicate pump: Ecological implications. *Limnology and Oceanography* 47:1849-1854.
- Ratnasingham S, Hebert PDN. 2007. BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Mol Ecol Notes* 7:355-364.
- Rey A, Basurko OC, Rodriguez-Ezpeleta N. 2020. Considerations for metabarcoding-based port biological baseline surveys aimed at marine nonindigenous species monitoring and risk assessments. *Ecology and Evolution* 10:2452-2465.
- Rey A, Carney KJ, Quinones LE, Pagenkopp Lohan KM, Ruiz GM, Basurko OC, Rodríguez-Ezpeleta N. 2019. Environmental DNA Metabarcoding: A Promising Tool for Ballast Water Monitoring. *Environmental Science & Technology* 53:11849-11859.
- Rius M, Turon X, Bernardi G, Volckaert FAM, Viard F. (Rius2015 co-authors). 2015. Marine invasion genetics: from spatio-temporal patterns to evolutionary outcomes. *Biological Invasions* 17:869-885.
- Rivero NK, Dafforn KA, Coleman MA, Johnston EL. 2013. Environmental and ecological changes associated with a marina. *Biofouling* 29:803-815.
- Robertson PA, Mill A, Novoa A, Jeschke JM, Essl F, Gallardo B, Geist J, Jarić I, Lambin X, Musseau C, et al. 2020. A proposed unified framework to describe the management of biological invasions. *Biological Invasions* 22:2633-2645.
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584.
- Roman J, Darling JA. 2007. Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology & Evolution* 22:454-464.
- Ruiz GM, Rawlings TK, Dobbs FC, Drake LA, Mullady T, Huq A, Colwell RR. 2000. Global spread of microorganisms by ships. *Nature* 408:49-50.
- Santos A, van Aerle R, Barrientos L, Martinez-Urtaza J. 2020. Computational methods for 16S metabarcoding studies using Nanopore sequencing data. *Computational and Structural Biotechnology Journal* 18:296-305.
- Sardain A, Sardain E, Leung B. 2019. Global forecasts of shipping traffic and biological invasions to 2050. *Nature Sustainability* 2:274-282.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537-7541.
- Schmidt BR, Kéry M, Ursenbacher S, Hyman OJ, Collins JP. 2013. Site occupancy models in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. *Methods in Ecology and Evolution* 4:646-653.
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. *Proceedings of the National Academy of Sciences* 109:6241-6246.
- Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, Winter M, Arianoutsou M, et al. 2017. No saturation in the accumulation of alien species worldwide. *Nature Communications* 8:14435.

- Sepulveda AJ, Nelson NM, Jerde CL, Luikart G. 2020. Are Environmental DNA Methods Ready for Aquatic Invasive Species Management? *Trends in Ecology & Evolution* 35:668-678.
- Shang L, Hu Z, Deng Y, Liu Y, Zhai X, Chai Z, Liu X, Zhan Z, Dobbs FC, Tang YZ. 2019. Metagenomic Sequencing Identifies Highly Diverse Assemblages of Dinoflagellate Cysts in Sediments from Ships' Ballast Tanks. *Microorganisms* 7:250.
- Shaw JLA, Weyrich LS, Hallegraeff G, Cooper A. 2019. Retrospective eDNA assessment of potentially harmful algae in historical ship ballast tank and marine port sediments. *Molecular Ecology* 28:2476-2485.
- Siddall ME, Fontanella FM, Watson SC, Kvist S, Erséus C. 2009. Barcoding Bamboozled by Bacteria: Convergence to Metazoan Mitochondrial Primer Targets by Marine Microbes. *Systematic Biology* 58:445-451.
- Sigsgaard EE, Nielsen IB, Bach SS, Lorenzen ED, Robinson DP, Knudsen SW, Pedersen MW, Jaidah MA, Orlando L, Willerslev E, et al. 2016. Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution* 1:0004.
- Simberloff D, Martin J-L, Genovesi P, Maris V, Wardle DA, Aronson J, Courchamp F, Galil B, García-Berthou E, Pascal M, et al. 2013. Impacts of biological invasions: what's what and the way forward. *Trends in Ecology & Evolution* 28:58-66.
- Simberloff D, Parker IM, Windle PN. 2005. Introduced species policy, management, and future research needs. *Frontiers in Ecology and the Environment* 3:12-20.
- Simberloff D, Von Holle B. (Simberloff1999 co-authors). 1999. Positive Interactions of Nonindigenous Species: Invasional Meltdown? *Biological Invasions* 1:21-32.
- Sinniger F, Pawlowski J, Harii S, Gooday AJ, Yamamoto H, Chevaldonné P, Cedhagen T, Carvalho G, Creer S. 2016. Worldwide analysis of sedimentary DNA reveals major gaps in taxonomic knowledge of deep-sea benthos. *Frontiers in Marine Science* 3.
- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences* 103:12115-12120.
- Stachowicz JJ, Whitlatch RB, Osman RW. 1999. Species Diversity and Invasion Resistance in a Marine Ecosystem. *Science* 286:1577-1579.
- Stat M, Huggett MJ, Bernasconi R, DiBattista JD, Berry TE, Newman SJ, Harvey ES, Bunce M. 2017. Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Sci Rep* 7:12240.
- Stefanni S, Stanković D, Borme D, de Olazabal A, Juretić T, Pallavicini A, Tirelli V. 2018. Multi-marker metabarcoding approach to study mesozooplankton at basin scale. *Sci Rep* 8:12085.
- Stewart KA, Taylor SA. 2020. Leveraging eDNA to expand the study of hybrid zones. *Molecular Ecology* 29:2768-2776.
- Strayer DL, D'Antonio CM, Essl F, Fowler MS, Geist J, Hilt S, Jarić I, Jöhnk K, Jones CG, Lambin X, et al. 2017. Boom-bust dynamics in biological invasions: towards an improved application of the concept. *Ecology Letters* 20:1337-1350.
- Streftaris N, Zenetos A, Papathanassiou E. 2005. Globalisation in marine ecosystems: The story of non-indigenous marine species across European seas. *Oceanography and Marine Biology* 43.
- Suarez-Menendez M, Planes S, Garcia-Vazquez E, Ardura A. Coastal Lagoon Through eDNA Metabarcoding.
- Sylvester F, Kalaci O, Leung B, Lacoursière-Roussel A, Murray CC, Choi FM, Bravo MA, Therriault TW, MacIsaac HJ. 2011. Hull fouling as an invasion vector: can simple models explain a complex problem? *Journal of Applied Ecology* 48:415-423.
- Taberlet P, Bonin A, Zinger L, Coissac E. 2018. *Environmental DNA: For Biodiversity Research and Monitoring*. Oxford: Oxford University Press.
- Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH. 2012. Environmental DNA. *Molecular Ecology* 21:1789-1793.
- Trebitz AS, Hoffman JC, Darling JA, Pilgrim EM, Kelly JR, Brown EA, Chadderton WL, Egan SP, Grey EK, Hashsham SA, et al. 2017. Early detection monitoring for aquatic non-indigenous species: Optimizing

- surveillance, incorporating advanced technologies, and identifying research needs. *Journal of Environmental Management* 202:299-310.
- Tsuji S, Maruyama A, Miya M, Ushio M, Sato H, Minamoto T, Yamanaka H. 2020. Environmental DNA analysis shows high potential as a tool for estimating intraspecific genetic diversity in a wild fish population. *Molecular Ecology Resources* 00:1-11.
- Turon X, Antich A, Palacín C, Præbel K, Wangensteen OS. 2020a. From metabarcoding to metaphylogeography: separating the wheat from the chaff. *Ecological Applications* 30:e02036.
- Turon X, Casso M, Pascual M, Viard F. 2020b. Looks can be deceiving: *Didemnum pseudovexillum* sp. nov. (Asciacea) in European harbours. *Marine Biodiversity* 50:48.
- Viard F, Comtet T. 2015. 18. Applications of DNA-based methods for the study of biological invasions. In: Canning-Clode J, editor. *Biological Invasions in Changing Ecosystems: Sciendo Migration*. p. 411-435.
- Viard F, David P, Darling JA. 2016. Marine invasions enter the genomic era: three lessons from the past, and the way forward. *Current zoology* 62:629-642.
- Viard F, Roby C, Turon X, Bouchemousse S, Bishop JDD. 2019. Cryptic diversity and database errors challenge non-indigenous species surveys: an illustration with *Botrylloides* spp. in the English Channel and Mediterranean Sea. *Frontiers in Marine Science* 6:615.
- Vieira L, Jones M, Taylor P. 2014. The identity of the invasive fouling bryozoan *Watersipora subtorquata* (d'Orbigny) and some other congeneric species.
- Vitousek PM, D'antonio CM, Loope LL, Rejmanek M, Westbrooks R. 1997. Introduced species: a significant component of human-caused global change. *New Zealand Journal of Ecology* 21:1-16.
- von Ammon U, Wood SA, Laroche O, Zaiko A, Lavery SD, Inglis GJ, Pochon X. 2019. Linking Environmental DNA and RNA for Improved Detection of the Marine Invasive Fanworm *Sabella spallanzanii*. *Frontiers in Marine Science* 6.
- von Ammon U, Wood SA, Laroche O, Zaiko A, Tait L, Lavery S, Inglis G, Pochon X. 2018a. The impact of artificial surfaces on marine bacterial and eukaryotic biofouling assemblages: A high-throughput sequencing analysis. *Marine Environmental Research* 133:57-66.
- von Ammon U, Wood SA, Laroche O, Zaiko A, Tait L, Lavery S, Inglis GJ, Pochon X. 2018b. Combining morpho-taxonomy and metabarcoding enhances the detection of non-indigenous marine pests in biofouling communities. *Sci Rep* 8:16290.
- Wangensteen OS, Palacín C, Guardiola M, Turon X. 2018. DNA metabarcoding of littoral hard-bottom communities: high diversity and database gaps revealed by two molecular markers. *PeerJ* 6:e4705.
- Williams M-A, O'Grady J, Ball B, Carlsson J, de Eyto E, McGinnity P, Jennings E, Regan F, Parle-McDermott A. 2019. The application of CRISPR-Cas for single species identification from environmental DNA. *Molecular Ecology Resources* 19:1106-1114.
- Williamson M, Fitter A. 1996. The Varying Success of Invaders. *Ecology* 77:1661-1666.
- Wolff WJ, Reise K. 2002. Oyster Imports as a Vector for the Introduction of Alien Species into Northern and Western European Coastal Waters. In: Leppäkoski E, Gollasch S, Olenin S, editors. *Invasive Aquatic Species of Europe. Distribution, Impacts and Management*. Dordrecht: Springer Netherlands. p. 193-205.
- Wood SA, Pochon X, Laroche O, von Ammon U, Adamson J, Zaiko A. 2019. A comparison of droplet digital polymerase chain reaction (PCR), quantitative PCR and metabarcoding for species-specific detection in environmental DNA. *Molecular Ecology Resources* 0:1-13.
- Wright D, Mitchelmore, Place A, Williams E, Orano D. 2019. Genomic and Microscopic Analysis of Ballast Water in the Great Lakes Region. *Applied Sciences* 9:2441.
- Zaiko A, Martinez JL, Ardura A, Clusa L, Borrell YJ, Samuiloviene A, Roca A, Garcia-Vazquez E. 2015a. Detecting nuisance species using NGST: Methodology shortcomings and possible application in ballast water monitoring. *Marine Environmental Research* 112:64-72.
- Zaiko A, Martinez JL, Schmidt-Petersen J, Ribicic D, Samuiloviene A, Garcia-Vazquez E. 2015b. Metabarcoding approach for the ballast water surveillance – An advantageous solution or an awkward challenge? *Marine Pollution Bulletin* 92:25-34.

- Zaiko A, Pochon X, Garcia-Vazquez E, Olenin S, Wood SA. 2018. Advantages and limitations of environmental DNA/RNA tools for marine biosecurity: management and surveillance of non-indigenous species. *Frontiers in Marine Science* 5.
- Zaiko A, Samuiloviene A, Ardura A, Garcia-Vazquez E. 2015c. Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. *Marine Pollution Bulletin* 100:53-59.
- Zaiko A, Schimanski K, Pochon X, Hopkins GA, Goldstien S, Floerl O, Wood SA. 2016. Metabarcoding improves detection of eukaryotes from early biofouling communities: implications for pest monitoring and pathway management. *Biofouling* 32:671-684.
- Zamora-Terol S, Novotny A, Winder M. 2020. Reconstructing marine plankton food web interactions using DNA metabarcoding. *Molecular Ecology*.





# List of tables

## Introduction

---

<b>Table 1</b> .....	<b>11</b>
Pathways and vectors (i.e. physical mechanisms) of introduction in marine systems. Table from Ojaveer, et al. (2014).	
<b>Table 2</b> .....	<b>18</b>
Overview of popular molecular techniques with some of their applications in the study of biological invasions, and NIS surveys. Modified from Viard and Comtet (2015), Darling, et al. (2017), and Zaiko, et al. (2018).	
<b>Table 3</b> .....	<b>30</b>
List of papers (sorted by publication year) reporting empirical data based on a metabarcoding approach to study non-indigenous species in the marine environment. The study region is provided with the sample type and marker(s) used. Data extracted and completed from Duarte, et al. (2020).	
<b>Table 4</b> .....	<b>33</b>
Research questions and challenges addressed in this thesis.	

## Chapter I

---

<b>Table 1</b> .....	<b>51</b>
Number of ASVs/OTUs retained with the six pipelines, and after post-treatment corrections (index-jumping, and selection on replicates). After comparison with SSIZ results, the number of expected haplotypes recovered, the names of missing haplotypes and the proportion of reads associated with unexpected sequences is indicated.	

## Chapter II.1

---

<b>Table 1</b> .....	<b>78</b>
Number and proportion (in parentheses) of ASVs and reads assigned to one of the reference sequence from the restricted database targeting ten metazoan phyla, for each marker (18S, COI, and 16S). The number indicated for "Taxa" includes assignments to species-, genus- and family-level. The last three columns indicate the number of species, the number of genera and the number of family detected for each marker.	

<b>Table 2</b> .....	<b>82</b>
List of non-indigenous (NIS) and cryptogenic (Crypto) species within the four target classes, known from the study area, and either observed in quadrats or identified in eDNA samples. Stars indicate that sequences were assigned to the genus or to another species of the same genus. The absence of reference sequence for each marker is indicated.	

## Chapter II.2

---

<b>Table 1</b> .....	<b>110</b>
Number of OTUs found in each of the eleven eDNA subsets for all three markers used in this study as well as the number of taxa assigned combining assignment across all markers.	
<b>Table 2</b> .....	<b>115</b>
PERMANOVA results comparing community composition between biogeographic regions, between marinas within regions, and between seasons, for the quadrat dataset and eDNA dataset based on taxonomic assignments across all three markers (9999 permutations). Non-significant values are in bold.	
<b>Table 3</b> .....	<b>115</b>
PERMANOVA pairwise comparisons for the three biogeographic regions. The test was computed on the quadrat dataset and eDNA dataset based on taxonomic assignments across all three markers. Due to the positive interaction between seasons and regions (see Table 2), permutations were constrained within seasons for each region (9999 permutations). WEC = Western English Channel, SB = Southern Brittany, and IS = Iroise Sea.	
<b>Table 4</b> .....	<b>118</b>
PERMANOVA results comparing community composition between biogeographic regions, between marinas within regions, and between seasons, for the quadrat dataset and eDNA dataset based on taxonomic assignments across all three markers (9999 permutations). Calculations were performed considering either native species or non-indigenous and cryptogenic species. Non-significant values are in bold.	
<b>Table 5</b> .....	<b>121</b>
List of the ten taxa showing the highest contribution to beta diversity for each season within either the quadrat dataset or the eDNA dataset based on OTU assignments across all markers. Species identified as non-indigenous or cryptogenic are in bold.	

## Chapter III.1

---

<b>Table 1</b> .....	<b>141</b>
Details of the six sampling sites within and outside the Bay of Morlaix. The first deployment period will further be referred to as "Summer" whereas the second period will be referred to as "Fall".	
<b>Table 2</b> .....	<b>147</b>
Number of hard-bottom sessile taxa identified among ten metazoan phyla whatever the taxonomic level considered. The number of species, genera, and families assigned are also detailed. The number and proportion (in parentheses) of ASVs and reads assigned to one of the reference sequence, for each marker and each dataset, are indicated.	
<b>Table 3</b> .....	<b>150</b>
List of taxa observed solely within the marina with at least one of the three methods. Species in bold are found only in the marina even when combining all three methods.	
<b>Table 4</b> .....	<b>153</b>
Summary statistics for the percentage of NIS and cryptogenic species recorded per method.	

## Chapter III.2

---

<b>Table 1</b> .....	<b>178</b>
Number of taxa (with the number of NIS among them indicated in bold and in parentheses), number of unique variants and number of reads identified at different taxonomic levels (species, genus, family or class) for each marker (18S and COI), within the four taxonomic groups of interest (Gymnolaemata, Gastropoda, Bivalvia, Ascidiacea) and using a BLAST <sup>®</sup> approach against a custom-designed database. The number of taxa (at the species, genus and family levels) with at least one reference sequence in the custom-designed database is given in the Ref column, with the number of reference for NIS in bold and in parentheses.	
<b>Table 2</b> .....	<b>181</b>
Non-indigenous species (NIS) in the four target classes detected by 18S, COI or both in at least one plankton sample collected in the bay of Morlaix. For each marker, the number of unique variants is given, with the number of reads in parentheses. For each NIS, the type of dispersal mode (Dispersal) is indicated with short and long disperser describing species with a life cycle including a pelagic larval stage lasting less or more than 2 days, respectively. 'Reported' indicates if the species has previously been reported in the study bay. For each marker, the total number (in bold) of reference sequences, retrieved from public databases or produced locally (in parenthesis), available in the custom-designed database is specified ( $N_{ref}$ ). Individual DNA was tested for amplification failure with the COI primers (see Figure S3 for amplification results). In case of COI amplification failure, N/A was added to the COI detection column.	

# List of figures

## Introduction

---

<b>Figure 1</b> .....	<b>3</b>
Global temporal trend in first record rate (dots) with the total number of established alien species during the time period considered given in parentheses. Data after 2000 (grey dots) are incomplete because of the delay between sampling and publication. Figure from Seebens, et al. (2017).	
<b>Figure 2</b> .....	<b>6</b>
Unified framework for biological invasions proposed by Blackburn, et al. (2011). An alphanumeric code (inside the white arrows) categorize the species along an invasion pathway.	
<b>Figure 3</b> .....	<b>8</b>
Dendrogram and map of compositional similarities among lists of introduced terrestrial gastropods before ( <b>A</b> and <b>B</b> ) and after ( <b>C</b> and <b>D</b> ) dispersal by humans. Colors indicate main clusters identified by the dendrogram and their corresponding locations in the world map. Figure from Capinha, et al. (2015). This figure highlights a redistribution of the species assemblages, and thus biogeographic boundaries after human-mediated transportation.	
<b>Figure 4</b> .....	<b>9</b>
Management actions specific to the stage of the introduction process, with coloured arrows highlighting expected changes in the species status following management actions. Figure from Robertson, et al. (2020).	
<b>Figure 5</b> .....	<b>13</b>
Number of NIS known or likely to be introduced by the most common human-assisted pathways. Percent of total number of species in assessment (n=329) is indicated. Figure from Molnar, et al. (2008).	
<b>Figure 6</b> .....	<b>14</b>
Pictures illustrating the species diversity (left) and population abundance (here the tunicate <i>Ciona intestinalis</i> , right) of fouling organisms attached to floating pontoons in French marinas. Photo credit: Wilfried Thomas – Station Biologique de Roscoff.	
<b>Figure 7</b> .....	<b>21</b>
A key issue for a reliable barcoding approach is the presence of a barcoding gap. Sequence databases are made of sequences produced with a given marker for one or several specimens that have been identified (usually from morphological criteria). The possibility to use this sequence as a barcode depends on the fact that the molecular variability is lower within species (i.e. between individuals of this species) than among species (i.e. between individuals belonging to different close species, such as within a genus). Figure modified from Viard and Comtet (2015).	

<b>Figure 8</b> .....	<b>23</b>
Chart of the different steps of the metabarcoding protocol, from sample collection to data analyses.	
<b>Figure 9</b> .....	<b>25</b>
Picture of marine water sampling in a marina for later extraction of environmental DNA. Photo credit: Yann Fontana – Station Biologique de Roscoff.	
<b>Figure 10</b> .....	<b>26</b>
Detailed steps of a dual-barcoded dual-indexed two-PCR library preparation for HTS sequencing used for most analyses done in this thesis. From the DNA sample, a first PCR is done with primers specific to the targeted gene region associated to one tag combination (specific to the PCR replicate) and one tail used in the following PCR. The second PCR is made to elongate the fragment with an index combination (specific to the sample) and primers to be used for the sequencing step.	
<b>Figure 11</b> .....	<b>27</b>
Illustration of the variation in percentage of accepted species (as listed in WORMS - World Register of Marine Species database; <a href="http://marinespecies.org/">http://marinespecies.org/</a> ) with reference in public databases (here in BOLD - Barcoding of Life Database; <a href="https://boldsystems.org/">https://boldsystems.org/</a> ) across different taxonomic groups. The number of accepted species for each class is given in parentheses. Data obtained on September 1, 2020.	
<b>Figure 12</b> .....	<b>28</b>
Number of publications using the term “metabarcoding” in Google Scholar by year of publication for the last decade. The number for 2020 only accounts papers published before September 2020.	

## Chapter I

---

<b>Figure 1</b> .....	<b>44</b>
(a) Collection sites of <i>Botrylloides</i> spp. colonies. SM = Saint-Malo, SQ = Saint-Quay-Portrieux, PG = Perros-Guirec, BLO = Bloiscon (Roscoff), AW = L’Aber Wrac’h, MB = Moulin Blanc (Brest), CAM = Camaret-sur-Mer, CON = Concarneau, ET = Étrel, TRI = La Trinité-sur-Mer. (b) <i>Botrylloides diegensis</i> . (c) <i>Botrylloides violaceus</i> . Photo credit: Yann Fontana.	
<b>Figure 2</b> .....	<b>46</b>
Overview of the experimental design from DNA extraction to data analyses. Dotted arrows represent the four different types of samples (3-, 6-, and 12-months ebDNA and bulkDNA). Data were processed with six bioinformatics pipelines. Extractions and amplifications protocols are detailed in the supporting information.	
<b>Figure 3</b> .....	<b>50</b>
Distribution patterns of <i>Botrylloides diegensis</i> (yellow) and <i>Botrylloides violaceus</i> (purple) as uncovered by SSIZ (scale pattern) or HTSA (results from DADA2 3-months ebDNA; plain color). See Fig. 1 for location codes.	

<b>Figure 4</b> .....	<b>53</b>
<p><b>(a)</b> Proportion of colonies or reads per haplotype in each jar (A and B) for each location (see Fig. 1 for location codes), as revealed by SSIZ (top panel) or HTSA using DADA2 for the four types of samples (four lower panels). ebDNA for ETA could not be amplified after 1 year. <b>(b)</b> Correlation between the proportion of reads (DADA2, 3 months) and the proportion of colonies (SSIZ) of a given haplotype in the same jar, with 95% confidence interval in grey. <b>(c)</b> Pearson correlation coefficient for each pipeline and sample type, as shown in B. All values were significant (<math>P &lt; 2.2 \times 10^{-16}</math>).</p>	
<b>Figure 5</b> .....	<b>54</b>
<p>Haplotype network built with ASVs produced by DADA2 on 3-months ebDNA data. Expected haplotypes are in colour, and unexpected sequences are in black. The size of the nodes represents the ASV abundance (fourth root of the number of reads) in the dataset. The number of crossing lines represents the number of mutations between two nodes. The dashed grey lines figure alternative links. The link between the two species has been shortened for visualization purposes and the 74-mutation step is written into brackets.</p>	
<b>Figure 6</b> .....	<b>55</b>
<p>Pairwise FST values computed from SSIZ (top left) or HTSA (DADA2 from 3-month ebDNA) (bottom right) data, and population clustering based on pairwise FST. The difference in clustering between the two datasets is highlighted in red. See Fig. 1 for location codes.</p>	

## Chapter II.1

---

<b>Figure 1</b> .....	<b>72</b>
<p>Location of the sampling sites. Abbreviations are as follows: SM = Saint-Malo, SQ = Saint-Quay-Portrieux, PG = Perros-Guirec, BLO = Bloscon (Roscoff), AW = L'Aber Wrac'h, MB = Moulin Blanc (Brest), CAM = Camaret-sur-Mer, CON = Concarneau, ET = Étrel, TRI = La Trinité-sur-Mer.</p>	
<b>Figure 2</b> .....	<b>79</b>
<p>Percentage of reads assigned to different kingdoms (left) or metazoan phyla (right), over all samples, for each of the three markers used in this study, as well as for Quadrats data (Quad).</p>	
<b>Figure 3</b> .....	<b>81</b>
<p>Comparison of eDNA metabarcoding and morphology-based detection. Only the taxa observed within quadrats and belonging to the four targeted classes of organisms are considered here. Taxa found with both methods within the same marina are displayed in green, those found only in quadrats are in red, and those only found in eDNA are in yellow. The first two panels correspond to samples from fall 2017 and from spring 2018, respectively. Metabarcoding detection is recorded across all markers. The column Tot represents the presence of each taxon in the whole dataset. The last three columns represent the presence of a taxon in the whole dataset (all marinas and dates combined) but for each marker separately. Black crosses indicate that no reference sequence was available for a particular taxon. The grey area represents species that were not isolated during morphological identification in</p>	

the fall 2017. They were all pooled in a unique category called "erected bryozoans". See figure 1 for location codes.

**Figure 4** .....83

Correlation between the number of individuals counted in all quadrats from a pontoon and the proportion of 18S reads within the sample of the same pontoon, for six tunicate species. Results of Pearson correlation tests are indicated within each plot. See figure 1 for location codes.

**Figure 5** .....85

Proportion of taxa recovered with eDNA metabarcoding across all markers or within quadrats, in samples from the two pontoons of a marina for a given season (shared) or in samples from only one pontoon (unique). See figure 1 for location codes.

**Figure 6** .....86

Distributions of read abundances (sum of all markers) for taxa found either in both 2-L replicates (Shared) or in only one replicate (Unique) for each pontoon. The number of taxa found in only one of the two 2L-replicates is indicated at the top of each panel (No) as well as the proportion of reads assigned to these taxa (%), for each pontoon in each marina. See figure 1 for location codes.

## Chapter II.2

---

**Figure 1** .....105

Map of the ten collection sites. Red squares indicate marinas located in the Western English Channel (Boreal province), blue circles indicate marinas on the Iroise Sea (transition zone), and green triangles are for marinas located in southern Brittany (Lusitanian province). Marina codes are as follows: SM = Saint-Malo, SQ = Saint-Quay-Portrieux, PG = Perros-Guirec, BLO = Bloscon (Roscoff), AW = L'Aber Wrac'h, MB = Moulin Blanc (Brest), CAM = Camaret-sur-Mer, CON = Concarneau, ET = Étrel, TRI = La Trinité-sur-Mer.

**Figure 2** .....111

Alpha-diversity between seasons (top panels) and regions (bottom panels), for 1) all OTUs, 2) OTUs of the 'benthic' subset, for the three markers, 3) All markers, benthic taxa assigned across the three markers combined (richness only), and 4) benthic taxa identified from quadrats. Significant differences (Wilcoxon test,  $P < 0.05$ ) are indicated with grey stars. A star between two bars indicate a significant difference between these two bars. A star on top of one bar indicate a significant difference between this bar and the two others.

**Figure 3** .....113

Ordination plots of principal component analysis results from Hellinger-transformed OTUs/taxa abundances (or occurrences) for pontoons from each locality according to their region (color) and the season of sampling (shape) on three datasets: All OTUs obtained with COI (A and B), benthic taxa assigned across all markers (C and D), and taxa from quadrat sampling (E and F). Sample scores are



displayed in scaling 1. See figure 1 for location codes. Left panel shows ordination with axes PCA1 and PCA2, right panel shows ordination with axes PCA1 and PCA3.

**Figure 4** ..... 114

Ordination plot of principal component analysis results from Hellinger-transformed OTU abundances for pontoons from each locality according to their region (colors) and the season of sampling (shape) on the three COI subsets: Chromista (A and B), Mobile (C and D), and Benthic (E and F). Sample scores are displayed in scaling 1. See figure 1 for location codes. Left panel shows ordination with axes PCA1 and PCA2, right panel shows ordination with axes PCA1 and PCA3.

**Figure 5** ..... 116

Correlation between the Bray-Curtis dissimilarity index for the quadrat dataset (top) or Jaccard index for the eDNA dataset ('All markers' subset) (bottom) and the "sailing" distance between marinas. Indices were calculated between pontoons for each season separately. Comparisons of pontoons from marinas from the same biogeographical region are coloured in dark blue whereas comparisons of pontoons from marinas from different biogeographical regions are coloured in orange.

**Figure 6** ..... 117

Proportion of non-indigenous and cryptogenic species (pink) and native species (seagreen) in every marina, in fall (left panels) and spring (right panels). Only data of the benthic taxa identified from assignment across all markers (upper panels) and data of the quadrat dataset (lower panels) are displayed. See figure 1 for location codes.

**Figure 7** ..... 119

Ordination plots of principal component analysis results from Hellinger-transformed taxa abundances between localities (left) and the ten species contributing the most to each axis (right) according to their region (colors) and the season of sampling (shape) on the quadrat dataset. Analyses were performed on either native species (A-B) or non-indigenous and cryptogenic species (C-D). Both sample scores and species scores are displayed in scaling 1. See figure 1 for location codes.

**Figure 8** ..... 120

Ordination plots of principal component analysis results from Hellinger-transformed taxa occurrences between localities (left) and the ten species contributing the most to each axis (right) according to their region (colors) and the season of sampling (shape) on the eDNA dataset based on taxonomic assignments across all markers. Analyses were performed on either native species (A-B) or non-indigenous and cryptogenic species (C-D). Both sample scores and species scores are displayed in scaling 1. See figure 1 for location codes.

## Chapter III.1

---

<b>Figure 1</b> .....	<b>140</b>
Location of each sampling point (left), and schematic and picture (from BBL site) of the experimental settlement structure (right). The site labelled BLO corresponds to Roscoff marina. Photo credit: Wilfried Thomas – Station biologique de Roscoff.	
<b>Figure 2</b> .....	<b>146</b>
Percentage of reads assigned to a kingdom (top) or a metazoan phylum (bottom), over all samples, from the plates bulkDNA or the water eDNA datasets, for each of the three markers used in this study. An extra column in the “plates” panels displays the abundance percentage of taxa assigned to each of the kingdoms and metazoan phyla listed for the morphology dataset (M).	
<b>Figure 3</b> .....	<b>148</b>
Number of species (A) and non-indigenous or cryptogenic species (B) recovered by the three methods used in this study. Note that “native” vs. “NIS and cryptogenic species” was established only for the species previously observed in marinas from Brittany (Chapter II.1, see Materials & Methods).	
<b>Figure 4</b> .....	<b>149</b>
Number of taxa observed in the marina (BLO) or in other parts of the bay (Other localities) for each of the three method used in this study.	
<b>Figure 5</b> .....	<b>151</b>
Benthic taxa richness distribution for each locality at the two seasons of retrieval of the settlement structures.	
<b>Figure 6</b> .....	<b>152</b>
Proportion of non-indigenous and cryptogenic species (pink) and native species (beige) as observed by the three methods used, Morphology (M), plates bulkDNA (B), and water eDNA (W). NIS that might be false positives (see Chapter 2 and Table 3) were excluded.	
<b>Figure 7</b> .....	<b>154</b>
Ordination plots of principal component analysis results from Hellinger-transformed taxa abundances collected from morphological identification on settlement plates. The two seasons of sampling have been treated separately (Summer: A-B; Fall C-D). Sample scores are displayed in scaling 1. Colours indicate the different sampled localities.	
<b>Figure 8</b> .....	<b>155</b>
Ordination plots of principal component analysis results from Hellinger-transformed taxa occurrences collected from metabarcoding of bulkDNA from settlement plates. The two seasons of sampling have been treated separately (Summer: A-B; Fall C-D). Sample scores are displayed in scaling 1. Colours indicate the different sampled localities.	
<b>Figure 9</b> .....	<b>156</b>
Ordination plots of principal component analysis results from Hellinger-transformed taxa occurrences collected from metabarcoding of eDNA from water samples. The two seasons of sampling have been	

treated separately (Summer: A-B; Fall C-D). Sample scores are displayed in scaling 1. Colours indicate the different sampled localities.

## Chapter III.2

---

**Figure 1** ..... 179

**A.** Proportion of reads assigned to each of the target classes with the three tested assignment methods, namely BLAST<sup>®</sup> against a custom-designed database (Bc), BLAST<sup>®</sup> against the GenBank nt database (Bnt), and the *ecotag* tool from the OBITools suite (E). The number of reads assigned to Ascidiacea and Gymnolaemata are too low to be noticeable. **B.** Number of species identified with each method for the four target classes (Ascidiacea: As, Bivalvia: Bi, Gastropoda: Ga, Gymnolaemata: Gy). The proportion of NIS (dark colour) and native species (light colour) is indicated.

**Figure 2** ..... 182

Distribution of reads across all sampling dates for 11 of the 12 non-indigenous species identified with either 18S (blue, left axis), or COI (red, right axis). The results for *Crassostrea gigas* are presented in Figure 3.

**Figure 3** ..... 183

Distribution of reads across all sampling dates for two oyster species detected in the dataset, the non-indigenous Pacific oyster *Crassostrea gigas* (A) and the native European flat oyster *Ostrea edulis* (C). Distribution of reads assigned to *Crassostrea* spp. (and presumably *C. gigas*) is also indicated (B). The number of 18S reads (blue) for each sampling date is shown on the left axis while the number of COI reads (red) is represented on the right axis in C. The green curve in A and B represents the variations in sea surface temperature measured with a CTD probe (Seabird SBE19) at each zooplankton sampling event. Data for mid-May, mid-June and August 2012 were not available.

**Figure 4** ..... 185

Relative abundance of Manila clam (*Ruditapes philippinarum*) reads ('Metabarcoding' plots, A, 18S, and B, COI) or larvae ('Barcoding' plot, C) within bivalves, for nine samples collected in 2012. For the metabarcoding data (A and B), the number of reads assigned to *R. philippinarum* was divided by the number of reads assigned to the class Bivalvia. For individual barcoding data (C), the number of larvae identified as *R. philippinarum* based on the amplification of the 18S marker was divided by the number of identified bivalve larvae.

**Figure 5** ..... 186

Distribution of reads assigned to the slipper limpet (*Crepidula fornicata*) for 18S (A) and COI (B) within samples from the year 2012. **C.** Temporal variations in the number of larvae of this species counted in samples collected at the same sampling dates in 2012.

## Appendix 1: Supplementary material from chapter I

---

### a. Protocol for single zooids DNA extraction and amplification

One zooid from each colony was isolated in ethanol under a dissecting microscope. DNA was extracted using the NucleoSpin® Tissue 96-well Kit (Macherey-Nagel), following the manufacturer's protocol with few modifications: just before lysis, each zooid was rinsed in PBS buffer (1X) to remove ethanol, dried on an absorbent paper, and placed in a 8-tube strip, containing buffer lysis T1 and proteinase K. After 2-3h at 56°C, an additional volume of 25µl of proteinase K (20 mg mL<sup>-1</sup>) was added to the lysis buffer and lysis was completed overnight. A two-step elution was performed with twice 60 µL of elution buffer pre-heated at 70°C. DNA extracts were stored at -20°C until amplification.

A 709-bp COI fragment was amplified for each zooid using the primers designed by Folmer, Black, Hoeh, Lutz, and Vrijenhoek (1994) LCO1490: 5'-GGTCAACAAATCATAAAGATATTGG-3' and HCO2198: 5'-TAAACTTCAGGGTGACCAAAAAATCA-3'. For each reaction the total volume (25 µL) was composed of 0.5 U GoTaq® DNA polymerase (Promega), 1X reaction buffer, 50 µM dNTPs, 2 mM MgCl<sub>2</sub>, 48 ng µL<sup>-1</sup> of bovine serum albumin (BSA), 0.3 µM of each primer and 5 µL of stock DNA. Amplification involved an initial denaturation step at 94 °C for 3 min, followed by 35 cycles at 94 °C for 50 s, 51.5 °C for 50 s and 72 °C for 1 min, and a final extension step at 72 °C for 5 min.

In order to improve amplification and sequence quality for 17 *Botrylloides violaceus* samples, and 59 *B. diegensis* samples, a second amplification and sequencing was done with specific primers, also targeting the Folmer region. For *B. violaceus*, we used primers designed by Callahan, Deibel, McKenzie, Hall, and Rise (2010) Violet\_Forward: 5'-TTAGGTTTTGGTCTAGGTTTATTG-3' and Violet\_Reverse: 5'-TAAATGTTGATAAAGTACAGGGTC-3', amplifying a 644-bp fragment. For *B. diegensis*, we used newly designed primers: Bdieg-COI-F: 5'-TGTCTACTAATCATAAAGATATTAG-3' and Bdieg-COI-R2: 5'-AATATACTTCAGGGTGCCAA-3', amplifying a fragment of 713 bp. For each reaction the total volume (25 µL) was composed of 1 U GoTaq® DNA polymerase (Promega), 1X reaction buffer, 0.2 mM dNTPs, 2 mM MgCl<sub>2</sub>, 12 ng µL<sup>-1</sup> of BSA, 0.8 µM

of each primer and 5 µl of DNA stock solution. For the two specific markers, amplification involved an initial denaturation step at 94 °C for 3 min, followed by 35 cycles at 94 °C for 50 s, 49.5 °C for 50 s and 72 °C for 1 min, and a final extension step at 72 °C for 5 min.

PCR products were Sanger sequenced in both directions by Eurofins Genomics (Germany, GmbH). Sequences were aligned using CodonCode Aligner v.5.0.1 (CodonCode Corporation, Dedham, MA), and trimmed to 607-bp for *B. diegensis* and 580-bp for *B. violaceus*.

## **b. Protocol for DNA extraction from preservative ethanol and communities**

During all extraction procedures, precautions were taken to avoid contamination of DNA extracts with external DNA: all consumables not sold as DNA-free were immersed in 12.5 % commercial bleach (0.65 % hypochlorite) for at least 30 min, rinsed and placed under UV light for at least 15 min. All equipment and bench surfaces were also DNA decontaminated before use.

After 3, 6, and 12 months of storage, each jar was shaken by several inversions, and three replicates of 1 mL preservative ethanol were pipetted and deposited in a deep-well plate. This plate was then placed in a dry block at 70 °C, under a fume hood overnight to allow the ethanol to evaporate. The DNA was then purified using the NucleoSpin® Tissue 96-well Kit (Macherey-Nagel). For lysis, 200 µL of T1 buffer and 25 µL of proteinase K (20 mg.mL<sup>-1</sup>) were added in each well and the plate was incubated at 56°C for 15 min. After that, the manufacturer's protocol was followed until the end. Elution was performed in 30 µL of elution buffer pre-heated at 70°C. The volume gathered after the first elution was placed again on the column for a second elution to maximize the yield. One extraction control was performed for each storage duration by adding lysis buffer on an empty tube. DNA concentrations were quantified by fluorescence using a PicoGreen® Quant-It™ dsDNA kit. DNA samples were stored at -20°C.

Soon after the collection of the last preservative ethanol sample (after one year of storage), all the material (i.e. colonies and remaining ethanol) from a jar was transferred to a 2-L glass beaker, mixed with an immersion blender (ErgoMixx

MSM66020, BOSCH®) until homogenization, and then filtered with a 41- $\mu\text{m}$  mesh size nylon filter. The solid fraction ( $> 41 \mu\text{m}$ ) was stored in a 50 mL tube at  $-20^\circ\text{C}$  until DNA extraction, which was performed no longer than four weeks later. DNA extraction was performed using the NucleoSpin® Soil kit (Macherey-Nagel). Previous to lysis, 300 mg of wet material of each sample were put in an oven at  $70^\circ\text{C}$  for 10 min to allow evaporation of residual ethanol. The bead beating step from the manufacturer's protocol was replaced by an overnight lysis step at  $56^\circ\text{C}$  where 1 mL of SL1 buffer, 150  $\mu\text{L}$  of SX buffer and 30  $\mu\text{L}$  of proteinase K ( $20 \text{ mg}\cdot\text{mL}^{-1}$ ) were added to each tube. The subsequent steps were performed following the manufacturer's protocol. Each sample was extracted in three replicates and two extraction controls were performed by adding lysis buffer to an empty tube. DNA was quantified by absorbance in a Spark TECAN reader using a NanoQuant Plate™. DNA samples were stored at  $-20^\circ\text{C}$ .

### c. Protocol for the two-step PCR COI library preparation

All library preparation steps were performed in the respect of strict rules preventing contamination of samples. They include, for example, the UV irradiation of all tips, plates and tubes before use, the mandatory use of filter tips, the strict separation of pre- and post-PCR steps in different labs.

Library preparation was performed using a dual-barcoded, dual-indexed two-step PCR procedure. First, a 455-bp COI portion was amplified for each sample using primers specifically designed to target species from the genus *Botrylloides*. BotrF2.2 - 5'-AGTGTTTTYATTCGTWTAGA-3' and BotrR7.1 - 5'-CAAAACARAGAYATRGARAAAYAT-3'. At the 5' end of each primer, a Nextera tail and an 8-bp tag were added to identify PCR and extraction replicates (see Table S1). Nine replicates (three tagged PCR pools for each of the three extraction replicates) were identified using a unique combination of five different tags.

**Table S1** Sequences of Nextera-tailed and tagged primers for amplifying COI in ebDNA and bulkDNA samples

NXTtag1_COIBotrF2.2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTATGCCAGTGTTTTYATTCGTWTA*G*A
NXTtag2_COIBotrF2.2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGGACTAAGTGTTTTYATTCGTWTA*G*A
NXTtag3_COIBotrF2.2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCACGTATAGTGTTTTYATTCGTWTA*G*A
NXTtag4_COIBotrF2.2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTACTCAGGAGTGTTTTYATTCGTWTA*G*A
NXTtag5_COIBotrF2.2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATCTTCAGAGTGTTTTYATTCGTWTA*G*A
NXTtag1_COIBotrR7.1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTATGCCCAAACARAGAYATRGARAAY*A*T
NXTtag2_COIBotrR7.1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGGACTACAAAACARAGAYATRGARAAY*A*T
NXTtag3_COIBotrR7.1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACGTATCAAACARAGAYATRGARAAY*A*T
NXTtag4_COIBotrR7.1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACTCAGGCAAACARAGAYATRGARAAY*A*T
NXTtag5_COIBotrR7.1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCTTCAGCAAACARAGAYATRGARAAY*A*T

Each PCR reaction was performed in a total volume of 10  $\mu$ L, composed of 0.3 U of Q5® Hot Start High-Fidelity DNA polymerase (New England Biolabs®, Inc.), 1X reaction buffer, 160  $\mu$ M dNTPs, 0.3  $\mu$ M of each tagged primer, 0.3  $\mu$ M of the same primers without the tag and the Illumina tail (to enhance amplification), and 2 ng DNA template. Amplification involved an initial denaturation step at 98 °C for 4 min, followed by 35 cycles at 94 °C for 1 min, 46 °C for 45 s and 72 °C for 1 min, and a final extension step at 72 °C for 10 min. PCR products were checked on a 1.5% agarose gel and visualised under UV light after ethidium bromide staining. For each extraction replicate, three tagged-primer combinations were used. To reduce any stochastic biases that could appear during amplification, three independent PCRs, using three different thermocyclers, were performed with each tagged-primer combination. The three PCR products amplified with the same tagged-primer combination were pooled. A total of nine technical replicates was thus obtained (i.e. three tagged-PCR replicates for each of the three extraction replicates per sample). After this first PCR step, the three tagged replicates from the same DNA extraction sample were pooled according to their intensity on the agarose gel visualisation. All pools were purified with paramagnetic beads in order to remove excess primers and putative primer dimers using the NucleoMag® NGS clean up and size select kit following the manufacturer's protocol (ratio of 1:1 PCR product vs. beads).

A second PCR was performed to complete Illumina® adapters and insert an index allowing sample identification (a sample being a combination of a jar and a type of sample; Fig. 2). We used the list of indexed adapters described in the Illumina® Nextera XT library preparation protocol (8 indexes i5 and 12 i7 allowing 96 combinations, see Table S2). PCR reactions were performed in a total volume of 10  $\mu$ L

composed of 0.4 U of Q5® Hotstart High-Fidelity DNA polymerase (New England Biolabs®, Inc.), 1X reaction buffer, 200 µM dNTPs, 0.13 µM of each primer and 1 µL DNA template. Amplification involved an initial denaturation step at 98 °C for 4 min, followed by 12 cycles at 98 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s, and a final extension step at 72 °C for 5 min.

**Table S2** Nextera XT i5 and i7 indices used in combination in the second PCR (i5 and i7 indices are in bold; stars indicate phosphorothiate bounds protecting primers from 3' exonuclease activity of the Q5 enzyme).

i7_N701	CAAGCAGAAGACGGCATAACGAGAT <b>TCGCCT</b> AGTCTCGTGGGCTC*G*G
i7_N702	CAAGCAGAAGACGGCATAACGAGAT <b>CTAGTACGG</b> TCTCGTGGGCTC*G*G
i7_N703	CAAGCAGAAGACGGCATAACGAGAT <b>TTCTGCCT</b> GTCTCGTGGGCTC*G*G
i7_N705	CAAGCAGAAGACGGCATAACGAGAT <b>AGGAGTCCG</b> TCTCGTGGGCTC*G*G
i7_N706	CAAGCAGAAGACGGCATAACGAGAT <b>CATGCCT</b> AGTCTCGTGGGCTC*G*G
i7_N707	CAAGCAGAAGACGGCATAACGAGAT <b>GTAGAGAGG</b> TCTCGTGGGCTC*G*G
i7_N710	CAAGCAGAAGACGGCATAACGAGAT <b>CAGCCTCGG</b> TCTCGTGGGCTC*G*G
i7_N711	CAAGCAGAAGACGGCATAACGAGAT <b>TGCCTCTT</b> GTCTCGTGGGCTC*G*G
i7_N712	CAAGCAGAAGACGGCATAACGAGAT <b>TCCTCTACG</b> TCTCGTGGGCTC*G*G
i7_N714	CAAGCAGAAGACGGCATAACGAGAT <b>TCATGAGCG</b> TCTCGTGGGCTC*G*G
i7_N720	CAAGCAGAAGACGGCATAACGAGAT <b>AGGCTCCGG</b> TCTCGTGGGCTC*G*G
i7_N723	CAAGCAGAAGACGGCATAACGAGAT <b>GAGCGCTAG</b> TCTCGTGGGCTC*G*G
i5_S503	AATGATACGGCGACCACCGAGATCTACAC <b>TATCCTCTT</b> CGTCGGCAGCG*T*C
i5_S505	AATGATACGGCGACCACCGAGATCTACAC <b>CGTAAGGAG</b> TCGTCGGCAGCG*T*C
i5_S506	AATGATACGGCGACCACCGAGATCTACAC <b>ACTGCATAT</b> CGTCGGCAGCG*T*C
i5_S510	AATGATACGGCGACCACCGAGATCTACAC <b>CGTCTAATT</b> CGTCGGCAGCG*T*C
i5_S513	AATGATACGGCGACCACCGAGATCTACAC <b>TCGACTAGT</b> CGTCGGCAGCG*T*C
i5_S516	AATGATACGGCGACCACCGAGATCTACAC <b>CCTAGAGTT</b> CGTCGGCAGCG*T*C
i5_S517	AATGATACGGCGACCACCGAGATCTACAC <b>GCGTAAGAT</b> CGTCGGCAGCG*T*C
i5_S522	AATGATACGGCGACCACCGAGATCTACAC <b>TTATGCGAT</b> CGTCGGCAGCG*T*C

After the second PCR, all PCR products were checked on a 1.5% agarose gel and visualized under UV light after ethidium bromide staining. The samples were then pooled according to their intensity on the agarose gel visualisation and the pool was purified by paramagnetic beads using the NucleoMag® NGS clean up and size select kit following manufacturer's protocol (ratio 1:1 PCR products vs. beads). Quantification of the library was performed with qPCR with the NEBNext® Library Quant Kit for Illumina® (New England Biolabs®, Inc.) and a DNA profile was performed using an DNA 1000 chip in a 1200 Bioanalyzer equipment (Agilent technologies, Inc.). Sequencing was performed on an Illumina® MiSeq sequencer with a 600 cycles v3 protocol with two index reads.



#### d. Community analysis based on 16S: protocol and results

In order to evaluate the overall diversity of metazoan species found within each jar, besides the targeted *Botrylloides* species, which includes the species associated to the sampled colonies (e.g. epibionts), a fragment of the 16S rDNA gene was amplified, for the 6-months ebDNA, and the bulkDNA extracts.

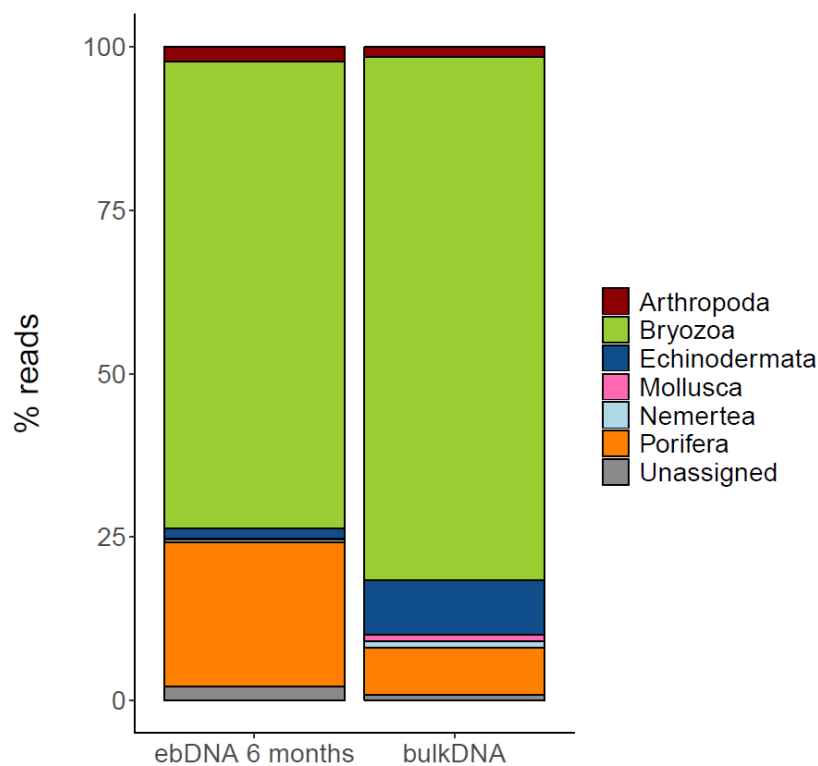
Library preparation and quantification were performed following the same two-steps PCR protocol as for COI, which is detailed in the section 3 above. Amplifications were done using the primer set designed by Kelly et al. (2016): 16S\_Metazoa\_fwd – 5'-AGTTACYYTAGGGATAACAGCG-3' and 16S\_Metazoa\_rev – 5'-CCGGTCTGAACTCAGATCAYGT-3'. The first PCR reaction (10 µL total volume) was composed of 0.6 U of Q5® High-Fidelity DNA polymerase (New England Biolabs®, Inc.), 1X reaction buffer, 200 µM dNTPs, 0.67 µM of each primer and 2 ng DNA template. Amplification involved an initial denaturation step at 98 °C for 4 min, followed by 40 cycles at 94 °C for 50 s, 61 °C for 45 s and 72 °C for 50 s, and a final extension step at 72 °C for 10 min. Sequencing was performed on an Illumina® MiSeq sequencer with a 300 cycles v2 micro cassette.

The 16S HTSA dataset was processed using DADA2 v-1.13.1 (Callahan et al., 2016), a denoising algorithm which removes PCR and sequencing errors and produces a set of amplicon sequence variants (ASVs). Index-jumping and replicate filters were applied as for the COI HTS dataset. ASVs resulting from the DADA2 analysis of the 16S dataset were then assigned against the nt GenBank database using the `ecotag` command from the `OBITOOLS` package, with no minimum identity threshold.

Over all samples, a total of 492,707 reads were obtained corresponding to 135 ASVs. Taxonomic assignment revealed that 99.2% of the reads corresponded to metazoans, as expected when using the 16S\_Metazoa primer set, designed by Kelly et al. (2016) to exclude non-metazoan taxa. The other 0.8% were unassigned eukaryotes. The most represented metazoan phyla were Bryozoa (76.5%), Porifera (13.4%), and Echinodermata (5.5%). The remaining 4.6% were assigned to Arthropoda, Nemertea, Mollusca, and unidentified metazoans (Fig. S1).

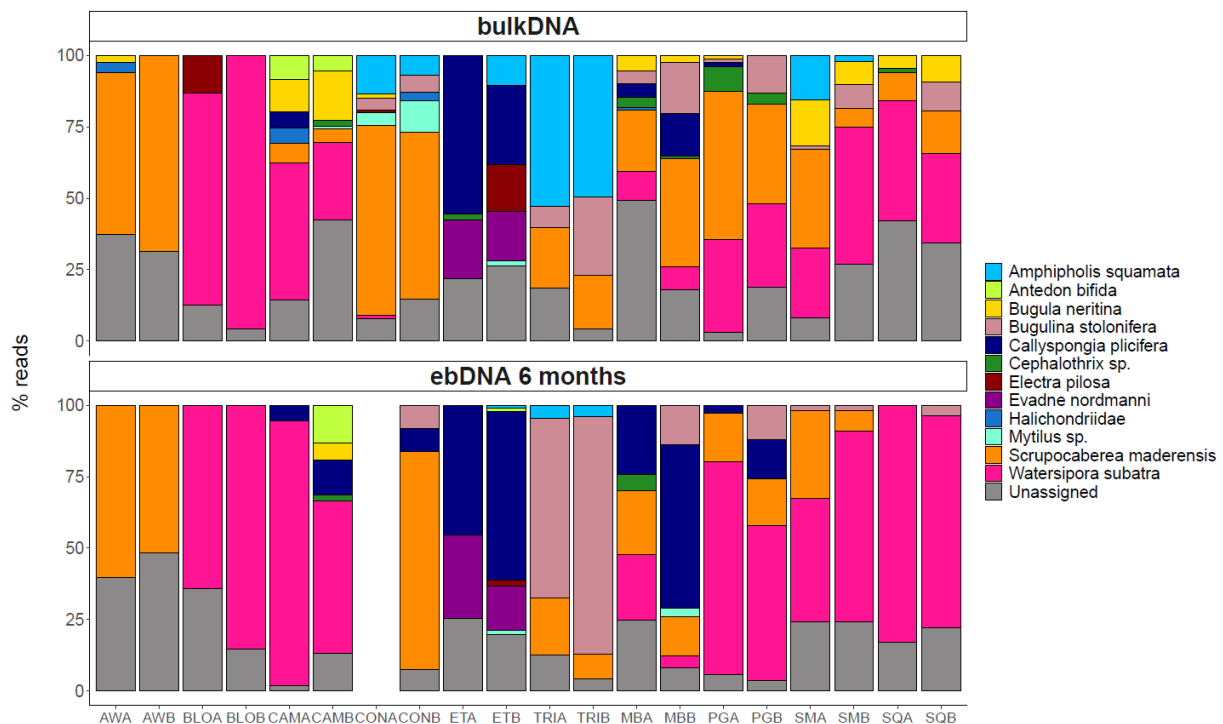
Importantly, no tunicates were found in our samples, which is surprising since we mostly expected DNA from tunicate species, notably *Botrylloides* spp. (in particular for bulkDNA) because they represented the majority of the biomass sampled. This result suggested important amplification biases for tunicates using

Kelly et al. (2016)'s primers. PCR amplifications of DNA from *B. violaceus* individual zooids always resulted in faint bands when using these primers. This supports the existence of 16S amplification biases for the target species, which could moreover be exacerbated under competition with other DNA. These biases could be explained when looking at the priming sites: two complete mitochondrial genomes are available in Genbank for *B. violaceus* (accession no. HF548552) and *B. diegensis* (accession no. NC 024103; registered under *B. leachii* but shown to be *B. diegensis*, Viard et al., 2019), showing two (*B. violaceus*) and three (for *B. diegensis*) mismatches on the forward primer and five on the reverse primer (for both species).



**Figure S3** Proportion of 16S reads assigned to each listed phylum for ethanol-based DNA (ebDNA 6 months) and bulkDNA samples using the ecotag tool.

Altogether 11 taxa were assigned down to the species or genus level (Fig. S2). The three most represented taxa, representing 65% of all reads, were *Watersipora subatra* (Ortmann, 1890) (34%), *Scrupocellaria maderensis* (now accepted as *Scrupocaberea maderensis* (Busk 1860)) (19%) and *Bugulina stolonifera* (Ryland, 1960) (12%).



**Figure S4** Proportion of 16S reads assigned to each listed species, genus or family for ethanol-based DNA (ebDNA 6 months) and bulkDNA samples using the ecotag tool. The sample of ebDNA at 6 months for the jar A in Concarneau (CON) could not be amplified.

Interestingly, *W. subatra*, an introduced encrusting bryozoan, was found in all study ports in quadrats scrapped under pontoons, during the same field campaign as for the collection of the *Botrylloides* samples. Similarly, the introduced erected bryozoan *B. stolonifera*, although not conspicuous, has been regularly reported in several of our study ports, during Rapid Assessment Surveys carried out by the Station Biologique of Roscoff (F. Viard & L. Lévêque, unpublished data). Regarding *S. maderensis*, we believe that this is an assignment error, as only two *Scrupocellaria* species (and none under the genus name *Scrupocaberea*) are present in Genbank, namely *S. madarensis* and *S. varians* – now accepted as *Pomocellaria varians* (Hincks, 1882)). This does not include the indigenous *S. scruposa*, which is found in our port surveys. The 16S marker used seems to amplify effectively bryozoans, which are important epibionts of our target species, and conspicuous in ports. This effectiveness is however balanced by the low number of references available in public databases for this marker, an issue frequently encountered particularly for NIS, and deserving further work between taxonomists and molecular biologists (Darling et al., 2017).

## e. Supplementary tables

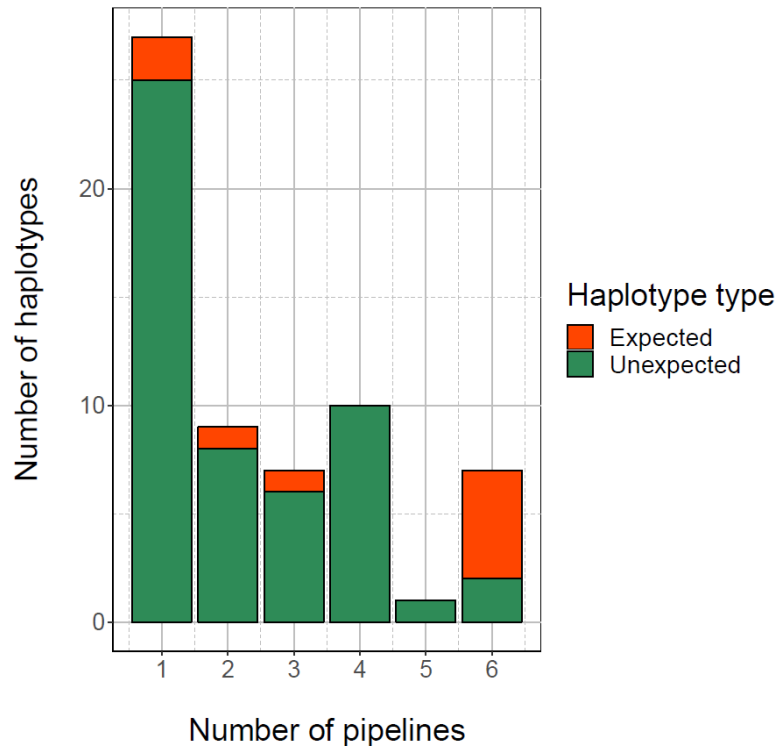
**Table S3** Chi<sup>2</sup> values calculated for the comparison between the proportions of colonies assigned to *Botrylloides violaceus* with Sanger sequencing on individual zooid (SSIZ) and the proportions of reads assigned to the same species with COI high-throughput sequencing on assemblages (HTSA). Comparisons have been performed for each pipeline and each type of sample (ethanol-based DNA after 3 months, 6 months or 1 year, and bulk DNA) in the three locations where the species has been detected. P-values are indicated in parentheses. Significant values are in red according to the cut-off which controls the false discovery rate (0.003) calculated with the brainwaver v-1.6 R package following Benjamini and Yekutieli (2001).

Marina	Date	DADA2	OBITOOLS	VSEARCH	OBI+SWARM	VS+SWARM	MOTHUR
<b>AW</b>	<b>3M</b>	1.45 (0.229)	2.02 (0.155)	0.42 (0.517)	0.008 (0.927)	0.02 (0.890)	0.09 (0.757)
	<b>6M</b>	2.87 (0.091)	3.32 (0.069)	1.27 (0.260)	0.37 (0.545)	0.36 (0.547)	0.54 (0.464)
	<b>1year</b>	2.20 (0.138)	3.04 (0.081)	0.94 (0.332)	0.31 (0.581)	0.40 (0.530)	0.52 (0.470)
	<b>bulk</b>	2.03 (0.154)	1.90 (0.168)	0.63 (0.429)	0.03 (0.861)	0.03 (0.856)	0.11 (0.738)
<b>PG</b>	<b>3M</b>	17.65 (2.6E-5)	15.16 (9.9E-5)	14.18 (1.7E-4)	13.22 (2.8E-4)	13.13 (2.9E-4)	15.44 (8.5E-5)
	<b>6M</b>	10.88 (9.7E-4)	10.94 (9.4E-4)	9.50 (0.002)	8.69 (0.003)	8.60 (0.003)	9.18 (0.002)
	<b>1year</b>	15.34 (9.0E-5)	15.30 (9.2E-5)	13.24 (2.7E-4)	36.18 (1.8E-9)	11.96 (5.4E-4)	12.90 (3.3E-4)
	<b>bulk</b>	4.30 (0.038)	4.17 (0.041)	3.59 (0.058)	3.37 (0.066)	3.31 (0.069)	3.30 (0.069)
<b>CON</b>	<b>3M</b>	6.90 (0.009)	6.28 (0.012)	6.69 (0.010)	7.05 (0.008)	6.95 (0.008)	6.51 (0.011)
	<b>6M</b>	7.45 (0.006)	6.58 (0.010)	6.62 (0.010)	7.02 (0.008)	6.74 (0.009)	6.30 (0.012)
	<b>1year</b>	8.13 (0.004)	7.65 (0.006)	7.61 (0.006)	8.05 (0.004)	7.78 (0.005)	7.37 (0.007)
	<b>bulk</b>	11.14 (8.4E-4)	26.27 (3.0E-7)	10.66 (0.001)	10.88 (9.7E-4)	11.14 (8.4E-4)	10.32 (0.001)

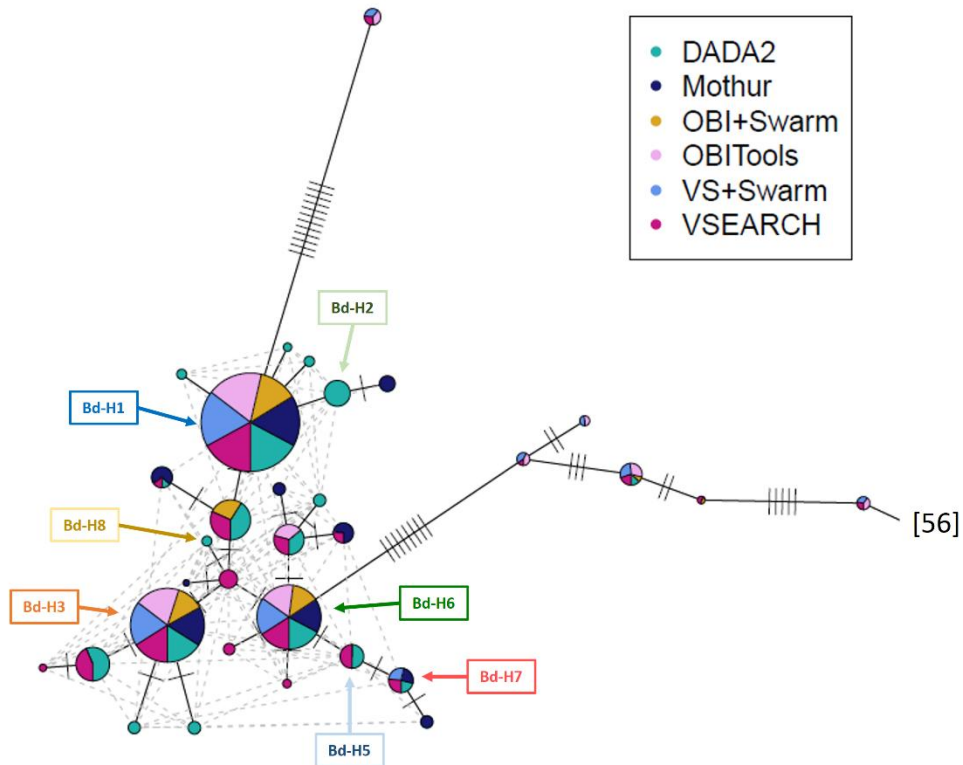
**Table S4** Values of average gene diversity per locus (Hs; as described by Nei (1973)) computed from the number of colonies (SSIZ) or the abundance of ASVs/OTUs (HTSA) for each marina, per pipeline and sample type (ebDNA after 3 months, 6 months or 1 year of storage and bulk DNA). See Figure 1 for the codes of the marinas.

Local	DADA2				OBITOOLS				VSEARCH				OBI+SWARM				VS+SWARM				MOTHUR				
	SSIZ	3mo	6mo	1yr	bulk	3mo	6mo	1yr	bulk	3mo	6mo	1yr	bulk	3mo	6mo	1yr	bulk	3mo	6mo	1yr	bulk	3mo	6mo	1yr	bulk
<b>AW</b>	0.530	0.528	0.562	0.484	0.544	0.476	0.505	0.434	0.492	0.507	0.539	0.459	0.527	0.446	0.477	0.415	0.459	0.418	0.442	0.395	0.430	0.414	0.440	0.390	0.426
<b>BLO</b>	0.444	0.543	0.510	0.495	0.578	0.472	0.437	0.426	0.505	0.542	0.512	0.498	0.578	0.505	0.475	0.463	0.531	0.451	0.421	0.415	0.471	0.456	0.424	0.410	0.478
<b>CAM</b>	0.057	0.005	0.015	0.001	0.034	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007
<b>CON</b>	0.363	0.525	0.456	0.430	0.524	0.445	0.386	0.353	0.442	0.495	0.424	0.396	0.490	0.415	0.369	0.340	0.409	0.415	0.370	0.341	0.409	0.432	0.362	0.331	0.419
<b>ET</b>	0.417	0.457	0.460	0.551	0.354	0.180	0.201	0.288	0.162	0.323	0.334	0.466	0.266	0.274	0.289	0.382	0.229	0.281	0.299	0.404	0.230	0.319	0.332	0.444	0.252
<b>TRI</b>	0.562	0.671	0.634	0.638	0.644	0.566	0.545	0.558	0.567	0.649	0.605	0.611	0.620	0.499	0.497	0.499	0.500	0.508	0.508	0.505	0.507	0.531	0.524	0.528	0.527
<b>MB</b>	0.208	0.126	0.132	0.092	0.244	0.070	0.083	0.069	0.195	0.080	0.092	0.063	0.222	0.077	0.087	0.074	0.187	0.074	0.085	0.073	0.184	0.072	0.089	0.071	0.195
<b>PG</b>	0.450	0.306	0.374	0.299	0.347	0.138	0.189	0.110	0.169	0.158	0.202	0.141	0.166	0.165	0.204	0.166	0.157	0.169	0.208	0.175	0.161	0.181	0.234	0.187	0.191
<b>SM</b>	0.398	0.519	0.496	0.381	0.362	0.451	0.426	0.335	0.303	0.509	0.489	0.373	0.355	0.421	0.409	0.337	0.306	0.422	0.406	0.331	0.299	0.432	0.422	0.336	0.308
<b>SQ</b>	0.257	0.366	0.376	0.536	0.312	0.326	0.326	0.491	0.262	0.367	0.369	0.534	0.307	0.323	0.332	0.474	0.262	0.323	0.323	0.475	0.261	0.315	0.324	0.474	0.257

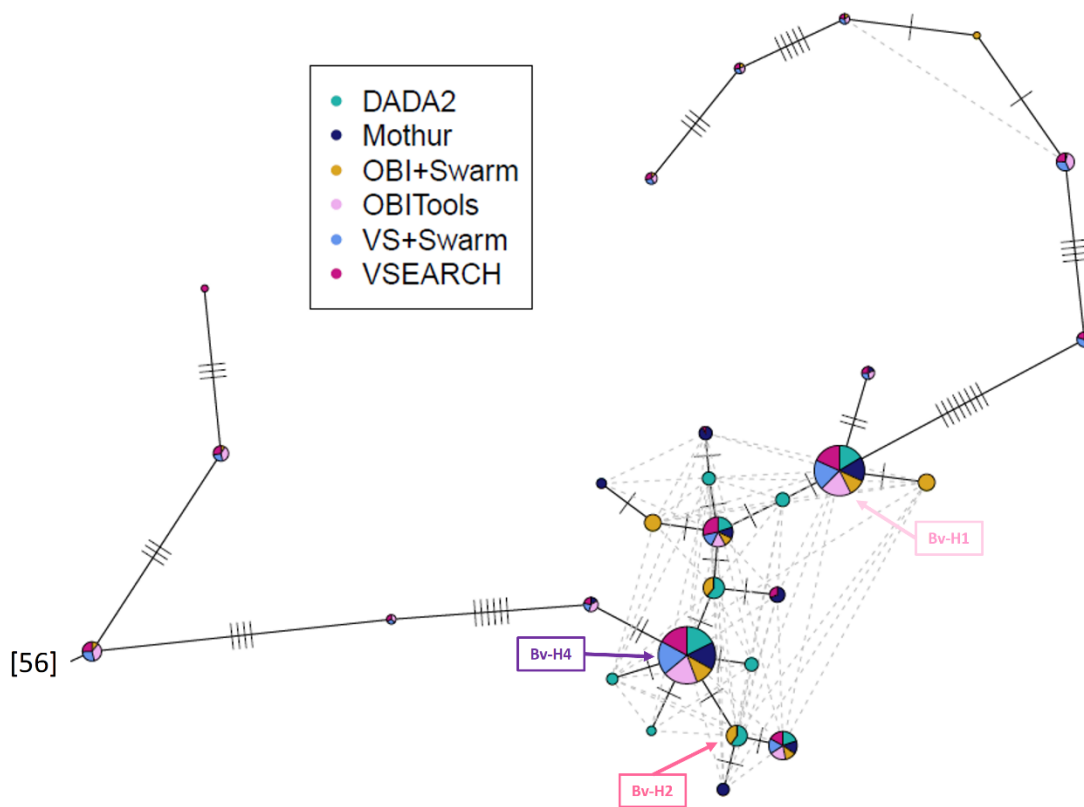
f. Supplementary figures



**Figure S5** Distribution of the number of expected (red) and unexpected (green) haplotypes identified by one, two, three, four, five or the six tested pipelines.

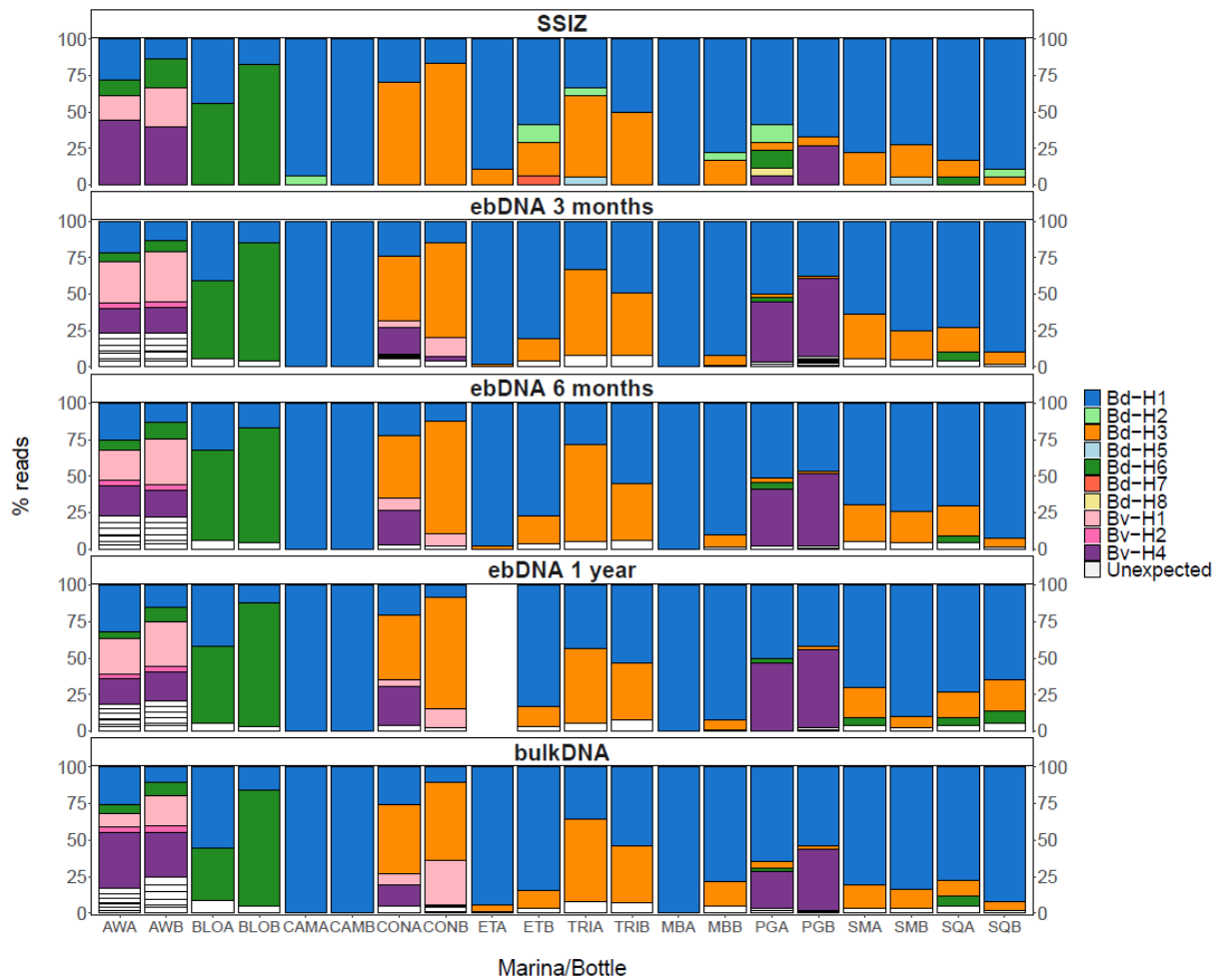


**Figure S6** Haplotype network for *Botrylloides diegensis* COI sequences performed with ASVs and OTUs produced by all six bioinformatics pipelines tested in this study. The size of a node represents the abundance (fourth root of the number of reads) of the corresponding ASV/OTU across all sample types. The contribution of each pipeline to the total amount of reads of each ASV/OTU is illustrated by the pie chart inside nodes. Expected haplotypes (i.e. haplotypes found with Sanger Sequencing of Individual Zooid, SSIZ) are labelled. The number of crossing lines represents the number of mutations between two nodes. The dashed grey lines figure alternative links. For a better visualization, only alternative links of one or two mutation steps are drawn. The link between *B. diegensis* and *B. violaceus* (presented in Figure S3) has been cut for visualization purposes and the 56 mutations separating the two species are written into brackets.

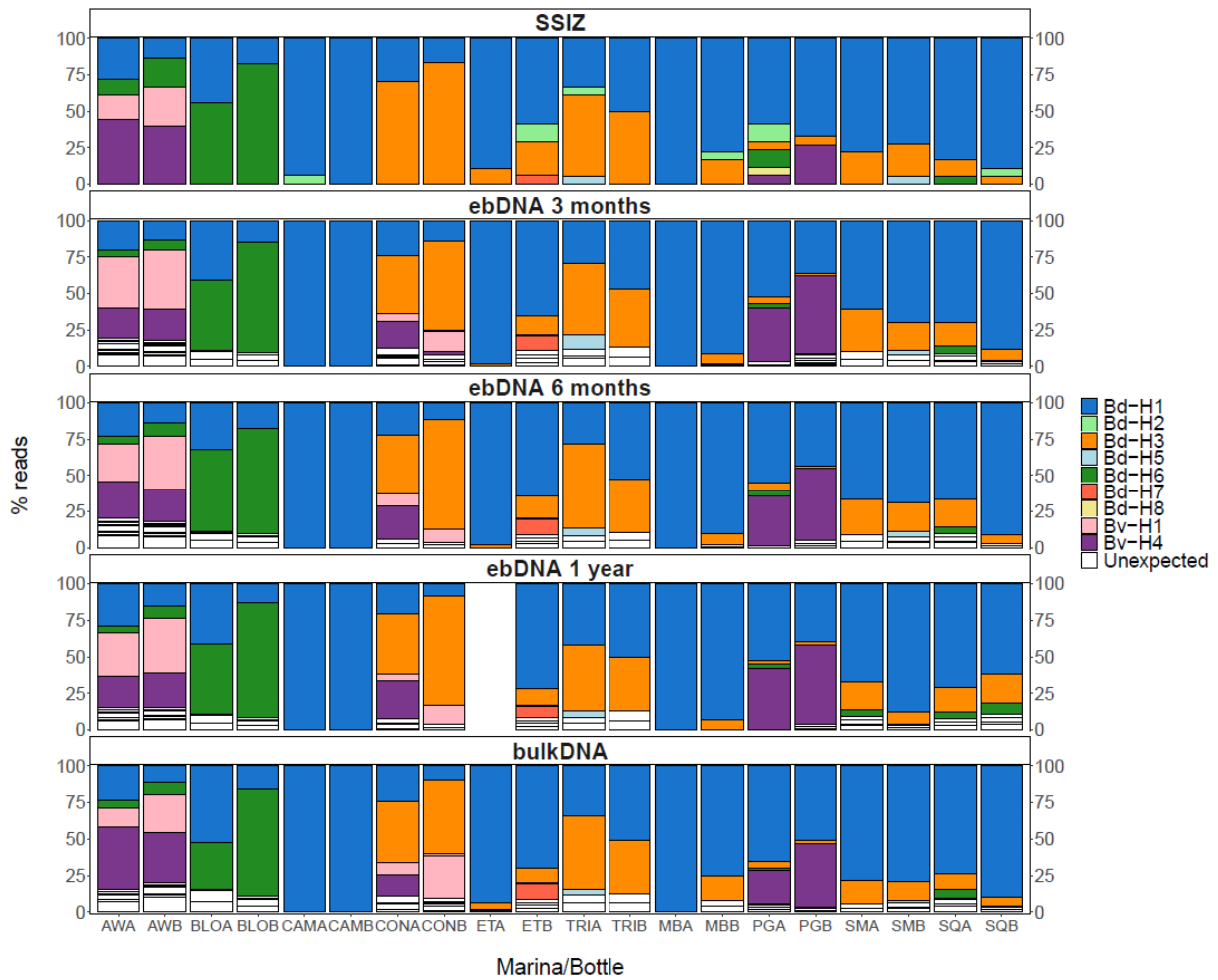


**Figure S7** Haplotype network for *Botrylloides violaceus* COI sequences performed with ASVs and OTUs produced by all six bioinformatics pipelines tested in this study. The size of a node represents the abundance (fourth root of the number of reads) of the corresponding ASV/OTU across all sample types. The contribution of each pipeline to the total amount of reads of each ASV/OTU is illustrated by the pie chart inside nodes. Expected haplotypes (i.e. haplotypes found with Sanger Sequencing of Individual Zooid, SSIZ) are labelled. Bv-H2 is labelled with a star because, despite being known from previous work (Viard, Roby, Turon, Bouchemousse, & Bishop, 2019), this haplotype was not found with SSIZ. The number of crossing lines represents the number of mutations between two nodes. The dashed grey lines figure alternative links. For a better visualization, only alternative links of one or two mutation steps are drawn. The link between *B. violaceus* and *B. diegensis* (presented in Figure S2) has been cut for visualization purposes and the 56 mutations separating the two species are written into brackets.

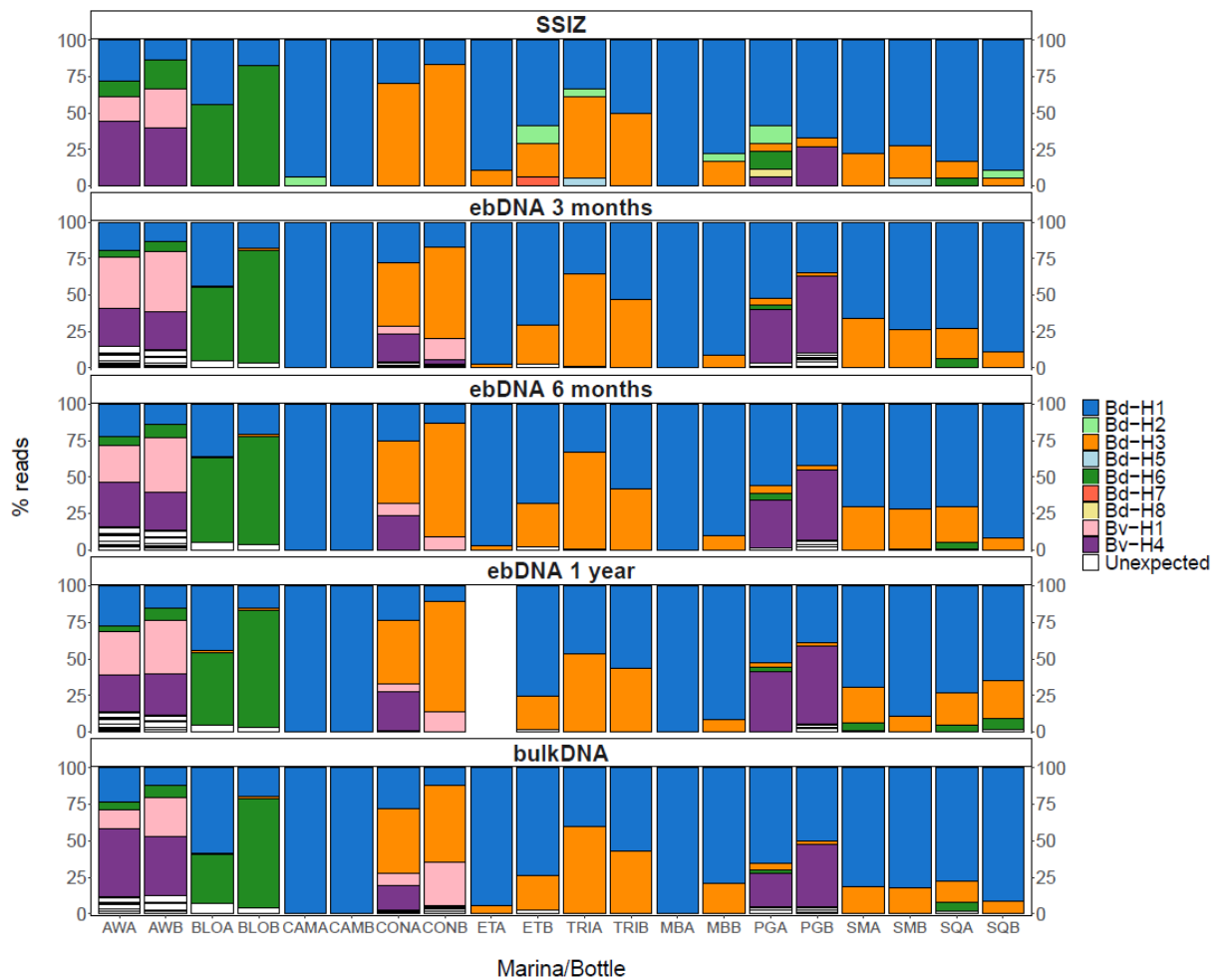




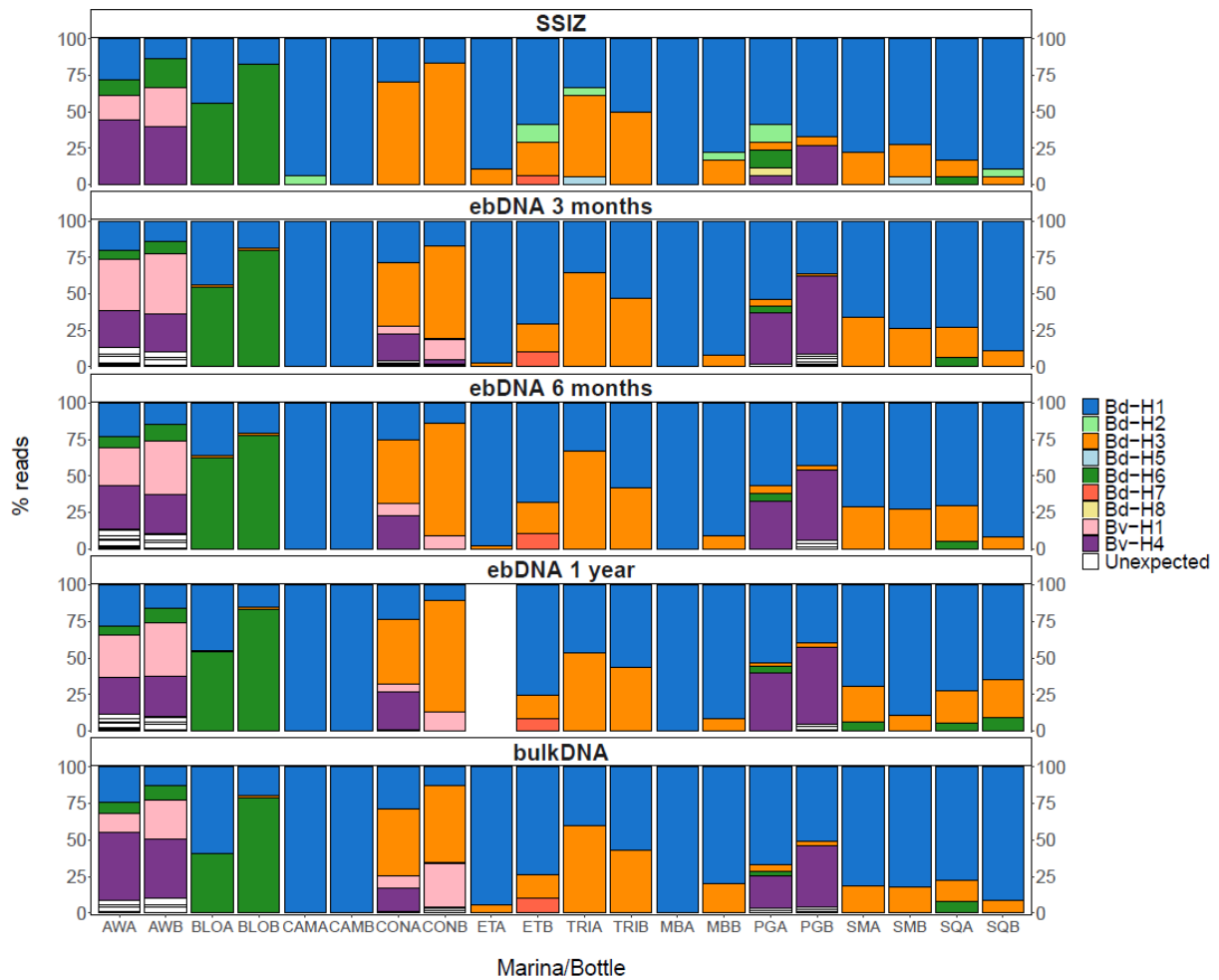
**Figure S8** Proportion of reads (or colonies in SSIZ) associated (100% identity) with each haplotype in every community (jars A and B for each location) as revealed by SSIZ (top panel) or HTSA using obitools for the four types of samples (four lower panels). The sample ETA after one-year storage could not be amplified. See Fig. 1 for marina codes and abbreviations of samples.



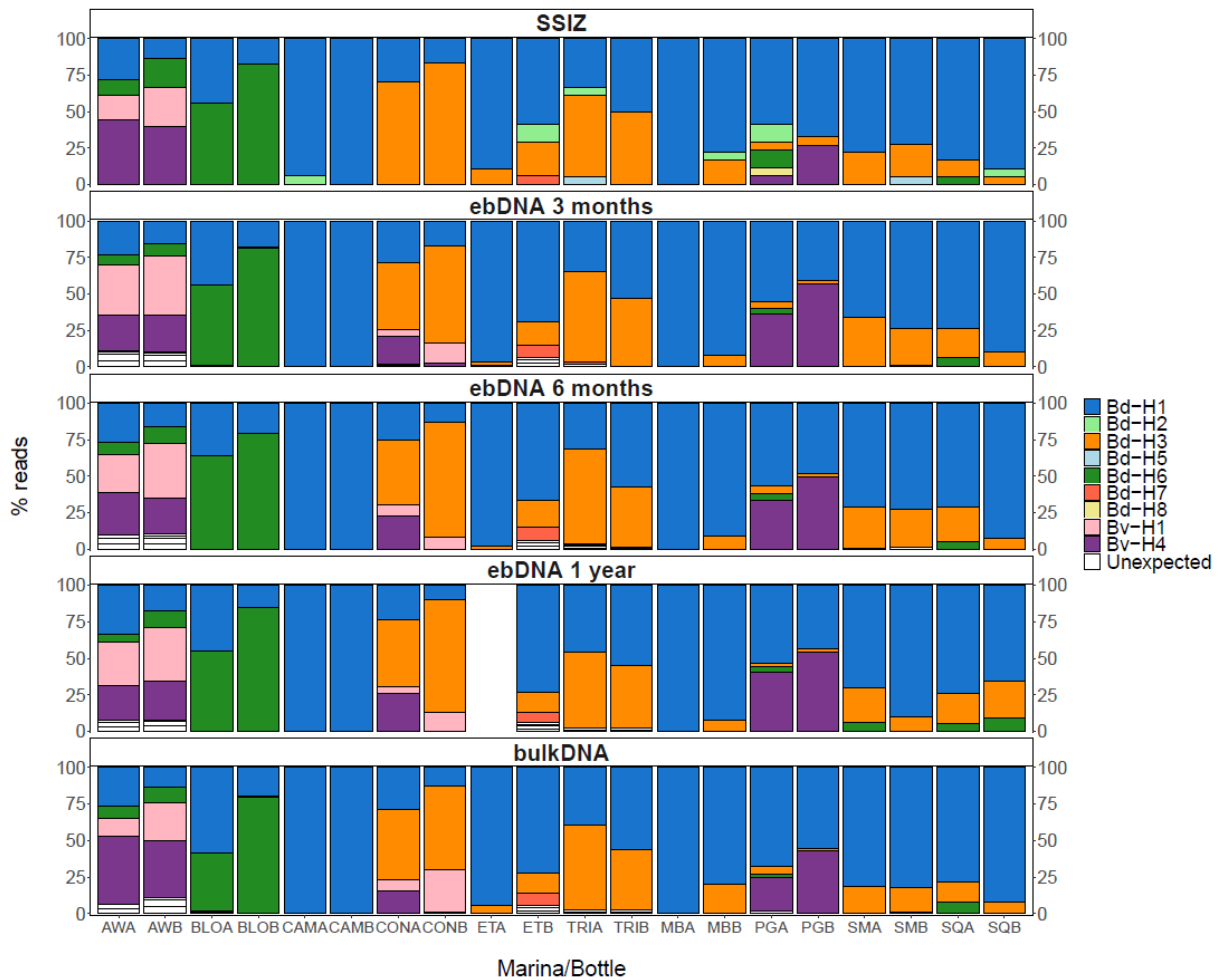
**Figure S9** Proportion of reads (or colonies in SSIZ) associated (100% identity) with each haplotype in every community (jars A and B for each location) as revealed by SSIZ (top panel) or HTSA using vsearch for the four types of samples (four lower panels). The sample ETA after one-year storage could not be amplified. See Fig. 1 for marina codes and abbreviations of samples.



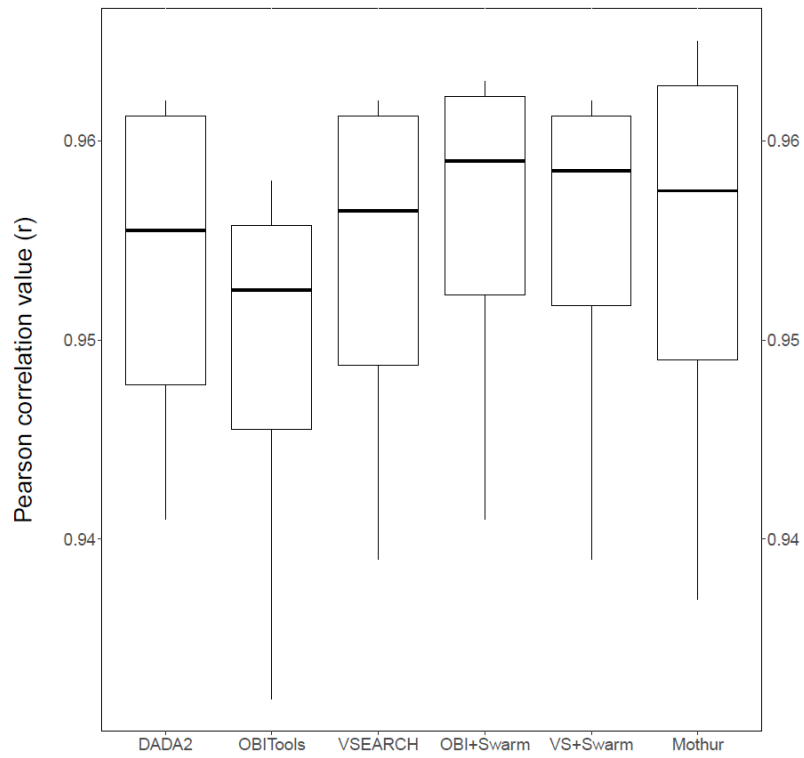
**Figure S10** Proportion of reads (or colonies in SSIZ) associated (100% identity) with each haplotype in every community (jars A and B for each location) as revealed by SSIZ (top panel) or HTSA using obi+swarm for the four types of samples (four lower panels). The sample ETA after one-year storage could not be amplified. See Fig. 1 for marina codes and abbreviations of samples.



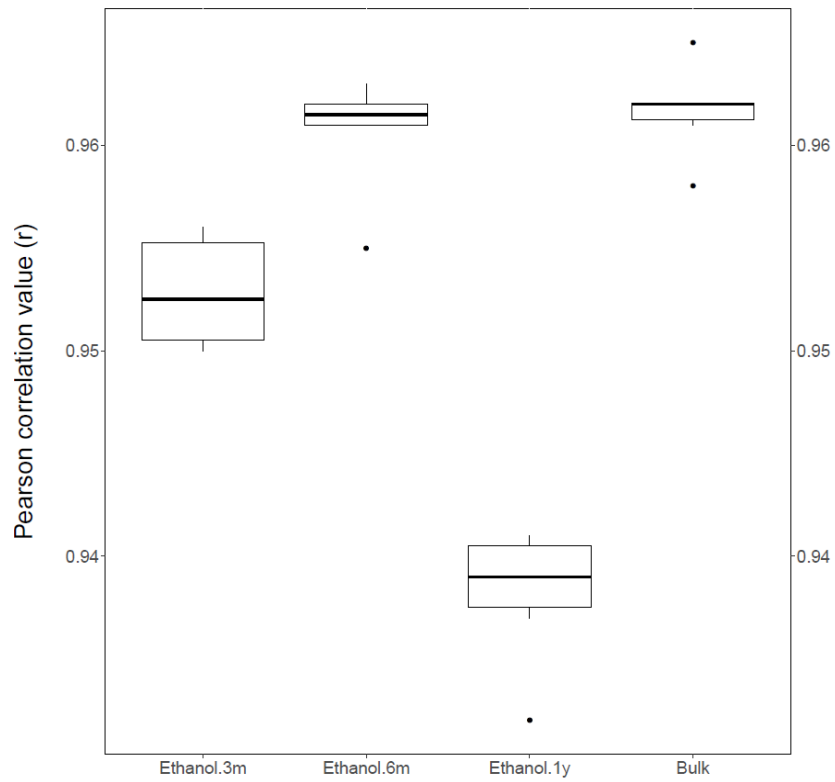
**Figure S11** Proportion of reads (or colonies in SSIZ) associated (100% identity) with each haplotype in every community (jars A and B for each location) as revealed by SSIZ (top panel) or HTSA using vs+swarm for the four types of samples (four lower panels). The sample ETA after one-year storage could not be amplified. See Fig. 1 for marina codes and abbreviations of samples.



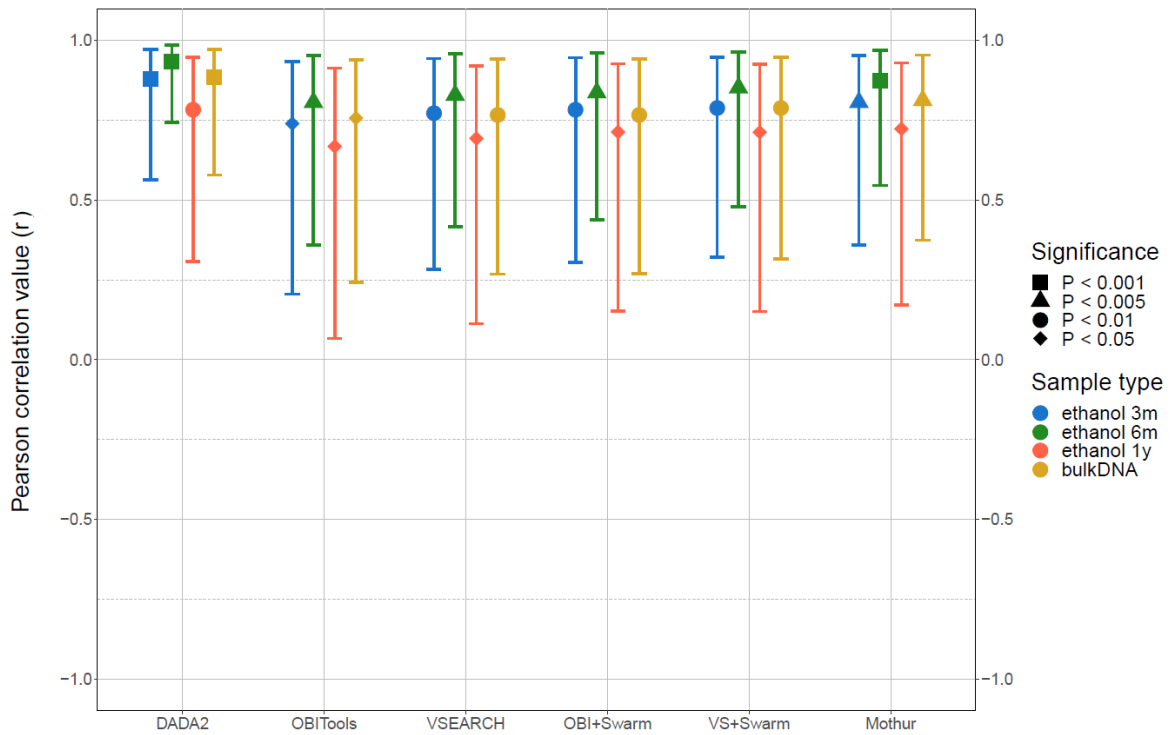
**Figure S12** Proportion of reads (or colonies in SSIZ) associated (100% identity) with each haplotype in every community (jars A and B for each location) as revealed by SSIZ (top panel) or HTSA using mother for the four types of samples (four lower panels). The sample ETA after one-year storage could not be amplified. See Fig. 1 for marina codes and abbreviations of samples.



**Figure S13** Distribution of Pearson correlation coefficient  $r$  values for each pipeline, computed on the relative abundance of reads of a given ASV/OTU in a given community (jar) and the proportion of colonies with this haplotype as determined by SSIZ in that community. All types of samples are considered here.



**Figure S14** Distribution of Pearson correlation coefficient  $r$  values for each type of sample, computed on the relative abundance of reads of a given ASV/OTU in a given community (jar) and the proportion of colonies with this haplotype as determined by SSIZ in that community. All pipelines are considered here.



**Figure S15** Pearson correlation coefficient values (r), for each pipeline and type of sample (ethanol-based DNA after 3 months, 6 months or 1 year of storage, and bulkDNA), between gene diversity per locus estimates (Hs) computed per locality from SSIZ and HTSA datasets. The shape of the points are related to the upper limit of the P-values associated with the correlation values. Vertical bars represents the 95% confidence intervals.



## References

- Benjamini, Y. & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4), 1165-1188 doi: [jstor.org/stable/2674075](https://doi.org/10.2307/2674075).
- Callahan, A. G., Deibel, D., McKenzie, C. H., Hall, J. R. & Rise, M. L. (2010). Survey of harbours in Newfoundland for indigenous and non-indigenous ascidians and an analysis of their cytochrome c oxidase I gene sequences. *Aquatic Invasions*, 5(1), 31-39 doi: [10.3391/ai.2010.5.1.5](https://doi.org/10.3391/ai.2010.5.1.5).
- Darling, J. A., Galil, B. S., Carvalho, G. R., Rius, M., Viard, F. & Piraino, S. (2017). Recommendations for developing and applying genetic tools to assess and manage biological invasions in marine ecosystems. *Marine Policy*, 85, 54-64 doi: [10.1016/j.marpol.2017.08.014](https://doi.org/10.1016/j.marpol.2017.08.014).
- Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology*, 3(5), 294-299 doi: [10.1007/BF02010148](https://doi.org/10.1007/BF02010148).
- Kelly, R. P., O'Donnell, J. L., Lowell, N. C., Shelton, A. O., Samhuri, J. F., Hennessey, S. M., ... Williams, G. D. (2016). Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ*, 4, e2444 doi: [10.7717/peerj.2444](https://doi.org/10.7717/peerj.2444).
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70(12), 3321-3323 doi: [10.1073/pnas.70.12.3321](https://doi.org/10.1073/pnas.70.12.3321).
- Viard, F., Roby, C., Turon, X., Bouchemousse, S. & Bishop, J. D. D. (2019). Cryptic diversity and database errors challenge non-indigenous species surveys: an illustration with *Botrylloides* spp. in the English Channel and Mediterranean Sea. *Frontiers in Marine Science*, 6, 615 doi: [10.3389/fmars.2019.00615](https://doi.org/10.3389/fmars.2019.00615).

## Appendix 2: Supplementary material from chapter II.1

### a. Supplementary tables

**Table S1** Sequences of Nextera-tailed and tagged primers for amplifying eDNA samples over three markers. Stars indicate phosphorothiate (PTO) bounds protecting primers from the 3' exonuclease activity of the Q5 enzyme.

Primer Name	Primer sequence (5' – 3')
NXTtag1_mlcOintF	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTATGCCGGWACWGGWTGAACWGTWTAYCCY*C*C
NXTtag2_mlcOintF	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGGACTAGGWACWGGWTGAACWGTWTAYCCY*C*C
NXTtag3_mlcOintF	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCACGTATGGWACWGGWTGAACWGTWTAYCCY*C*C
NXTtag4_mlcOintF	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTACTCAGGGWACWGGWTGAACWGTWTAYCCY*C*C
NXTtag5_mlcOintF	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATCTTCAGGGWACWGGWTGAACWGTWTAYCCY*C*C
NXTtag1_jgHCO2198	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTATGCCTAIACYTCIGGRTGICCRAARAAY*C*A
NXTtag2_jgHCO2198	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGGACTATAIACYTCIGGRTGICCRAARAAY*C*A
NXTtag3_jgHCO2198	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACGTATTAIACYTCIGGRTGICCRAARAAY*C*A
NXTtag4_jgHCO2198	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACTCAGGTAAIACYTCIGGRTGICCRAARAAY*C*A
NXTtag5_jgHCO2198	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCTTCAGTAIACYTCIGGRTGICCRAARAAY*C*A
NXTtag1_SSU_FO4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTATGCCGCTTGTCTCAAAGATTAAG*C*C
NXTtag2_SSU_FO4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGGACTAGCTTGTCTCAAAGATTAAG*C*C
NXTtag3_SSU_FO4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCACGTATGCTTGTCTCAAAGATTAAG*C*C
NXTtag4_SSU_FO4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTACTCAGGGCTTGTCTCAAAGATTAAG*C*C
NXTtag5_SSU_FO4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATCTTCAGGCTTGTCTCAAAGATTAAG*C*C
NXTtag1_SSU_R22_mod	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTATGCCCTGCTGCCTTCCTTR*G*A
NXTtag2_SSU_R22_mod	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGGACTACCTGCTGCCTTCCTTR*G*A
NXTtag3_SSU_R22_mod	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACGTATCCTGCTGCCTTCCTTR*G*A
NXTtag4_SSU_R22_mod	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACTCAGGCCTGCTGCCTTCCTTR*G*A
NXTtag5_SSU_R22_mod	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCTTCAGCCTGCTGCCTTCCTTR*G*A
NXTtag1_16S_Metazoa_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTATGCCAGTTACYYTAGGGATAACAG*C*G
NXTtag2_16S_Metazoa_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGGACTAAGTTACYYTAGGGATAACAG*C*G
NXTtag3_16S_Metazoa_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCACGTATAGTTACYYTAGGGATAACAG*C*G
NXTtag4_16S_Metazoa_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTACTCAGGAGTTACYYTAGGGATAACAG*C*G
NXTtag5_16S_Metazoa_fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATCTTCAGAGTTACYYTAGGGATAACAG*C*G
NXTtag1_16S_Metazoa_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTATGCCCGGTCTGAACTCAGATCAY*G*T
NXTtag2_16S_Metazoa_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGGACTACCGGTCTGAACTCAGATCAY*G*T
NXTtag3_16S_Metazoa_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACGTATCCGGTCTGAACTCAGATCAY*G*T
NXTtag4_16S_Metazoa_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACTCAGGCCGGTCTGAACTCAGATCAY*G*T
NXTtag5_16S_Metazoa_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCTTCAGCCGGTCTGAACTCAGATCAY*G*T

**Table S2** Nextera XT i5 and i7 indexes used in combination in the second PCR (i5 and i7 indexes are in bold; stars indicate phosphorothiate (PTO) bounds protecting primers from the 3' exonuclease activity of the Q5 enzyme).

Primer name	Primer sequence (5'-3')
i7_N701	CAAGCAGAAGACGGCATAACGAGAT <b>TCGCCTTAGT</b> TCTCGTGGGCTC*G*G
i7_N702	CAAGCAGAAGACGGCATAACGAGAT <b>CTAGTACGGT</b> TCTCGTGGGCTC*G*G
i7_N703	CAAGCAGAAGACGGCATAACGAGAT <b>TTCTGCCTGT</b> TCTCGTGGGCTC*G*G
i7_N705	CAAGCAGAAGACGGCATAACGAGAT <b>AGGAGTCCGT</b> TCTCGTGGGCTC*G*G
i7_N706	CAAGCAGAAGACGGCATAACGAGAT <b>CATGCCTAGT</b> TCTCGTGGGCTC*G*G
i7_N707	CAAGCAGAAGACGGCATAACGAGAT <b>GTAGAGAGGT</b> TCTCGTGGGCTC*G*G
i7_N710	CAAGCAGAAGACGGCATAACGAGAT <b>CAGCCTCGGT</b> TCTCGTGGGCTC*G*G
i7_N711	CAAGCAGAAGACGGCATAACGAGAT <b>TCCTCTTGT</b> TCTCGTGGGCTC*G*G
i7_N712	CAAGCAGAAGACGGCATAACGAGAT <b>TCTCTACGT</b> TCTCGTGGGCTC*G*G
i7_N714	CAAGCAGAAGACGGCATAACGAGAT <b>TCATGAGCGT</b> TCTCGTGGGCTC*G*G
i7_N720	CAAGCAGAAGACGGCATAACGAGAT <b>AGGCTCCGT</b> TCTCGTGGGCTC*G*G
i7_N723	CAAGCAGAAGACGGCATAACGAGAT <b>GAGCGTAGT</b> TCTCGTGGGCTC*G*G
i5_S503	AATGATACGGCGACCACCGAGATCTACACT <b>ATCCTCTT</b> CGTCGGCAGCG*T*C
i5_S505	AATGATACGGCGACCACCGAGATCTACAC <b>GTAAGGAGT</b> CGTCGGCAGCG*T*C
i5_S506	AATGATACGGCGACCACCGAGATCTACAC <b>ACTGCATAT</b> CGTCGGCAGCG*T*C
i5_S510	AATGATACGGCGACCACCGAGATCTACAC <b>CGTCTAATT</b> CGTCGGCAGCG*T*C
i5_S513	AATGATACGGCGACCACCGAGATCTACACT <b>CGACTAGT</b> CGTCGGCAGCG*T*C
i5_S516	AATGATACGGCGACCACCGAGATCTACAC <b>CCTAGAGTT</b> CGTCGGCAGCG*T*C
i5_S517	AATGATACGGCGACCACCGAGATCTACAC <b>GCGTAAGAT</b> CGTCGGCAGCG*T*C
i5_S522	AATGATACGGCGACCACCGAGATCTACACT <b>TATGCGAT</b> CGTCGGCAGCG*T*C

**Table S3** List of values applied to the different parameters during read processing and filtering for each of the three markers used in this study. All parameters not indicated in this table were used with the default value.

Tool	Command	Parameter	18S	COI	16S
CUTADAPT	cutadapt	-e	0.16	0.12	0.14
DADA2	filterAndTrim	truncLen	260 and 180	267 and 230	90 and 90
		truncQ	0	0	0
		rm.phix	False	False	False
	dada	pool	pseudo	pseudo	Pseudo
	mergePairs	minOverlap	60	170	60
	seqtab	Length selection	335:400	303:323	105:125
R	-	Index-jump	1.82%	1.09%	0.23%
		Replicates	2	2	2

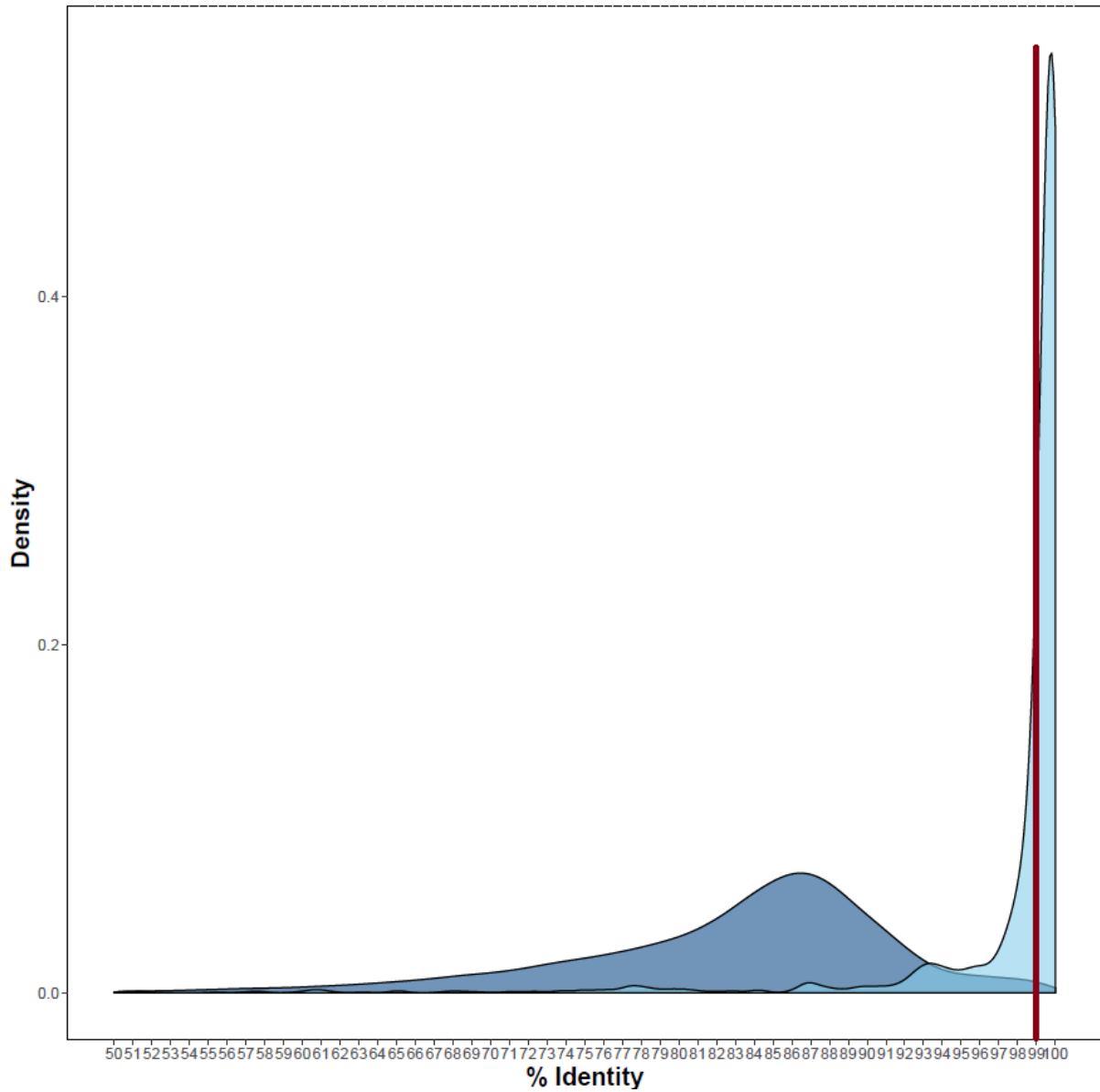
**Table S4** List of species identified with eDNA metabarcoding, in the four targeted metazoan classes (Ascidacea, Bivalvia, Gastropoda, Gymnolaemata) used for comparison with the quadrats dataset. For each marker, the number (and percentage) of reads and the maximum identity percentage of each assignment is given. The last column indicates if the species has already been reported in the study area.

Species	Class	18S		COI		16S		Expected
		Reads	% Ident	Reads	% Ident	Reads	% Ident	
<i>Ascidia conchilega</i>	Ascidacea	1077 (<0.1%)	100	-	-	-	-	Yes
<i>Ascidia mentula</i>	Ascidacea	261 (<0.1%)	100	-	-	-	-	Yes
<i>Asterocarpa humilis</i>	Ascidacea	4186 (<0.1%)	100	-	-	-	-	Yes
<i>Botrylloides diegensis</i>	Ascidacea	115543 (1%)	100	250 (<0.1%)	99.68	-	-	Yes
<i>Botrylloides leachii</i>	Ascidacea	-	-	13 (<0.1%)	99.68	-	-	Yes
<i>Botryllus schlosseri</i>	Ascidacea	6948 (<0.1%)	100	-	-	-	-	Yes
<i>Ciona intestinalis</i>	Ascidacea	341396 (2.9%)	100	1193 (<0.1%)	100	-	-	Yes
<i>Ciona robusta</i>	Ascidacea	4680 (<0.1%)	100	-	-	-	-	Yes
<i>Clavelina lepadiformis</i>	Ascidacea	-	-	12 (<0.1%)	100	-	-	Yes
<i>Clavelina meridionalis</i>	Ascidacea	70309 (0.6%)	99.16	-	-	-	-	No
<i>Corella eumyota</i>	Ascidacea	2319 (<0.1%)	100	-	-	-	-	Yes
<i>Dendrodoa grossularia</i>	Ascidacea	846 (<0.1%)	100	-	-	-	-	Yes
<i>Diplosoma listerianum</i>	Ascidacea	-	-	231 (<0.1%)	97.76	-	-	Yes
<i>Metandrocarpa taylori</i>	Ascidacea	224 (<0.1%)	99.17	-	-	-	-	No
<i>Molgula bleizi</i>	Ascidacea	4661 (<0.1%)	100	-	-	-	-	Yes
<i>Molgula complanata</i>	Ascidacea	100 (<0.1%)	100	-	-	-	-	Yes
<i>Molgula socialis</i>	Ascidacea	6056 (<0.1%)	100	-	-	-	-	Yes
<i>Pelonaia corrugata</i>	Ascidacea	31 (<0.1%)	99.45	-	-	-	-	No
<i>Perophora japonica</i>	Ascidacea	3920 (<0.1%)	100	-	-	-	-	Yes
<i>Phallusia mammillata</i>	Ascidacea	5343 (<0.1%)	100	843 (<0.1%)	100	-	-	Yes
<i>Polycarpa pomaria</i>	Ascidacea	2471 (<0.1%)	99.72	-	-	-	-	Yes
<i>Polycarpa tenera</i>	Ascidacea	-	-	116 (<0.1%)	100	-	-	No
<i>Styela clava</i>	Ascidacea	8070 (<0.1%)	100	-	-	-	-	Yes
<i>Abra alba</i>	Bivalvia	-	-	199 (<0.1%)	100	6317 (<0.1%)	100	Yes
<i>Acanthocardia tuberculata</i>	Bivalvia	-	-	-	-	1134 (<0.1%)	100	Yes
<i>Aequipecten opercularis</i>	Bivalvia	3860 (<0.1%)	100	-	-	319 (<0.1%)	100	Yes
<i>Asbjornsenia pygmaea</i>	Bivalvia	-	-	1073 (<0.1%)	100	-	-	Yes
<i>Callista chione</i>	Bivalvia	-	-	79 (<0.1%)	100	-	-	Yes
<i>Cerastoderma edule</i>	Bivalvia	862 (<0.1%)	100	207 (<0.1%)	100	29154 (0.3%)	100	Yes
<i>Chamelea striatula</i>	Bivalvia	-	-	-	-	31 (<0.1%)	100	Yes
<i>Corbula gibba</i>	Bivalvia	257 (<0.1%)	100	298 (<0.1%)	100	2466 (<0.1%)	100	Yes
<i>Donax trunculus</i>	Bivalvia	31 (<0.1%)	100	-	-	-	-	Yes
<i>Donax vittatus</i>	Bivalvia	186 (<0.1%)	100	-	-	-	-	Yes
<i>Dosinia exoleta</i>	Bivalvia	-	-	2821 (<0.1%)	93.29	-	-	Yes
<i>Hiatella arctica</i>	Bivalvia	9447 (<0.1%)	100	1181 (<0.1%)	99.36	2000 (<0.1%)	99.12	Yes

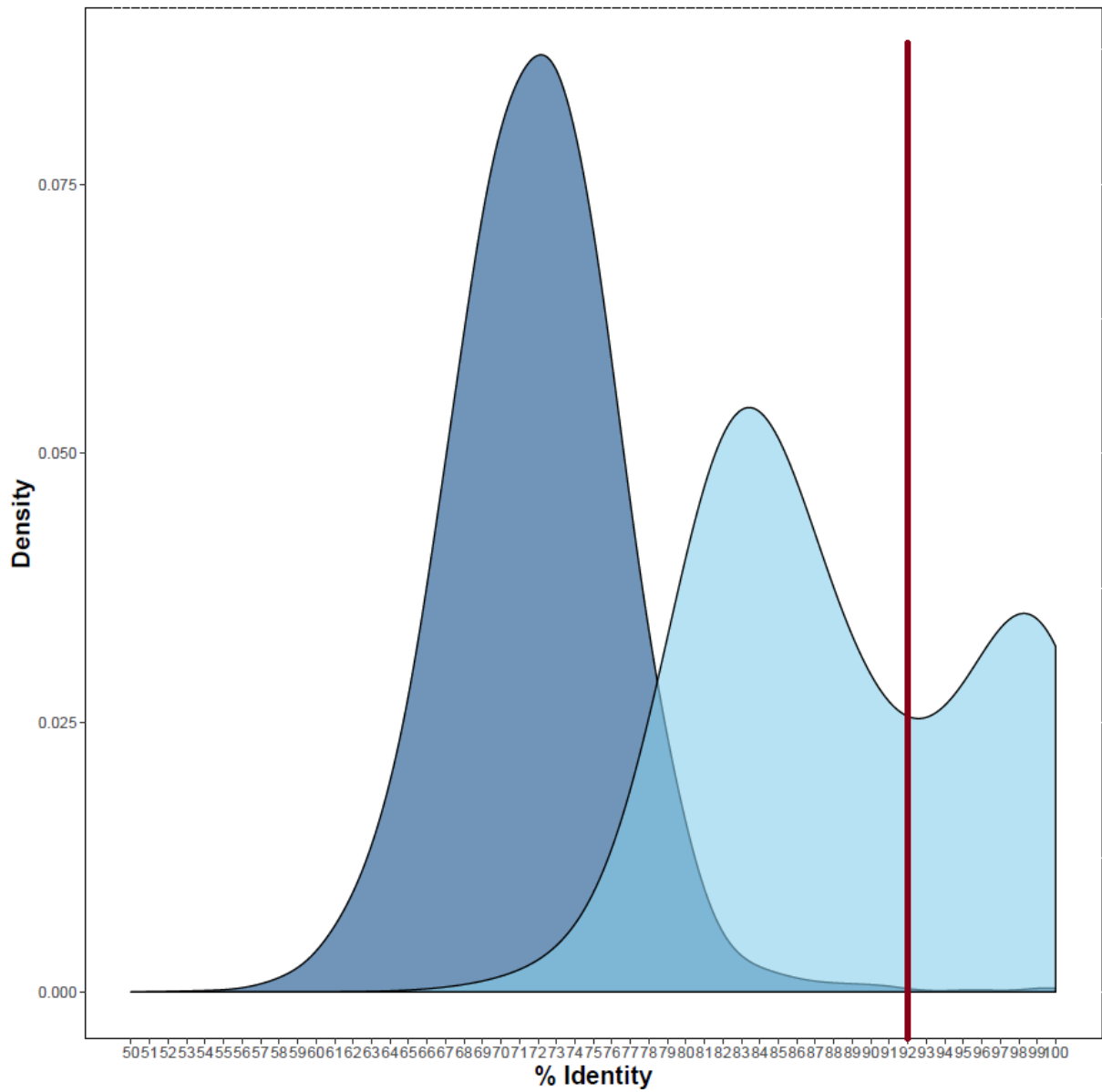
<i>Kurtiella bidentata</i>	Bivalvia	825 (<0.1%)	100	-	-	-	-	Yes
<i>Laevicardium crassum</i>	Bivalvia	-	-	183 (<0.1%)	100	-	-	Yes
<i>Limaria hians</i>	Bivalvia	353 (<0.1%)	100	-	-	-	-	Yes
<i>Macomangulus tenuis</i>	Bivalvia	-	-	-	-	2020 (<0.1%)	100	Yes
<i>Macra stultorum</i>	Bivalvia	513 (<0.1%)	100	17 (<0.1%)	99.36	-	-	Yes
<i>Mimachlamys varia</i>	Bivalvia	-	-	-	-	15971 (0.2%)	100	Yes
<i>Modiolula phaseolina</i>	Bivalvia	199 (<0.1%)	100	-	-	-	-	Yes
<i>Modiolus barbatus</i>	Bivalvia	-	-	-	-	187 (<0.1%)	100	Yes
<i>Mya arenaria</i>	Bivalvia	56 (<0.1%)	100	-	-	-	-	Yes
<i>Mysia undata</i>	Bivalvia	100 (<0.1%)	100	-	-	-	-	Yes
<i>Mytilus edulis</i>	Bivalvia	-	-	949 (<0.1%)	99.68	-	-	Yes
<i>Mytilus galloprovincialis</i>	Bivalvia	-	-	-	-	24591 (0.2%)	100	Yes
<i>Mytilus trossulus</i>	Bivalvia	-	-	749 (<0.1%)	99.68	27231 (0.3%)	100	No
<i>Parvicardium scabrum</i>	Bivalvia	293 (<0.1%)	100	-	-	-	-	Yes
<i>Pharus legumen</i>	Bivalvia	124 (<0.1%)	100	-	-	-	-	Yes
<i>Polititapes aureus</i>	Bivalvia	442 (<0.1%)	100	1086 (<0.1%)	100	433 (<0.1%)	100	Yes
<i>Polititapes rhomboides</i>	Bivalvia	6291 (<0.1%)	100	-	-	1488 (<0.1%)	100	Yes
<i>Ruditapes decussatus</i>	Bivalvia	110 (<0.1%)	100	-	-	116 (<0.1%)	100	Yes
<i>Ruditapes philippinarum</i>	Bivalvia	657 (<0.1%)	100	1225 (<0.1%)	100	138 (<0.1%)	100	Yes
<i>Scrobicularia plana</i>	Bivalvia	77 (<0.1%)	100	-	-	-	-	Yes
<i>Solecurtus divaricatus</i>	Bivalvia	29 (<0.1%)	100	-	-	-	-	No
<i>Solen marginatus</i>	Bivalvia	-	-	-	-	506 (<0.1%)	100	Yes
<i>Spisula solida</i>	Bivalvia	-	-	166 (<0.1%)	100	-	-	Yes
<i>Spisula subtruncata</i>	Bivalvia	-	-	222 (<0.1%)	100	-	-	Yes
<i>Tellimya ferruginosa</i>	Bivalvia	-	-	14 (<0.1%)	100	-	-	Yes
<i>Timoclea ovata</i>	Bivalvia	858 (<0.1%)	100	-	-	-	-	Yes
<i>Venerupis corrugata</i>	Bivalvia	8021 (<0.1%)	100	-	-	18384 (0.2%)	100	Yes
<i>Akera bullata</i>	Gastropoda	-	-	367 (<0.1%)	99.68	-	-	Yes
<i>Alvania tenera</i>	Gastropoda	-	-	-	-	3507 (<0.1%)	100	No
<i>Aplysia parvula</i>	Gastropoda	2172 (<0.1%)	100	-	-	-	-	No
<i>Bittium reticulatum</i>	Gastropoda	-	-	4271 (<0.1%)	99.68	-	-	Yes
<i>Calliostoma zizyphinum</i>	Gastropoda	-	-	39 (<0.1%)	100	-	-	Yes
<i>Calma gobioophaga</i>	Gastropoda	-	-	83 (<0.1%)	98.71	-	-	No
<i>Calyptrea chinensis</i>	Gastropoda	-	-	12 (<0.1%)	100	-	-	Yes
<i>Cerithiopsis petanii</i>	Gastropoda	-	-	485 (<0.1%)	100	-	-	No
<i>Crepidula fornicata</i>	Gastropoda	-	-	1979 (<0.1%)	100	-	-	Yes
<i>Doto coronata</i>	Gastropoda	-	-	519 (<0.1%)	100	-	-	Yes
<i>Elysia viridis</i>	Gastropoda	35 (<0.1%)	100	26 (<0.1%)	99.65	-	-	Yes
<i>Embletonia pulchra</i>	Gastropoda	-	-	225 (<0.1%)	99.27	-	-	Yes
<i>Eucithara coronata</i>	Gastropoda	89 (<0.1%)	99.185	-	-	-	-	No
<i>Favorinus branchialis</i>	Gastropoda	-	-	1667 (<0.1%)	100	-	-	Yes
<i>Fjordia lineata</i>	Gastropoda	-	-	52 (<0.1%)	100	-	-	Yes

<i>Gibbula magus</i>	Gastropoda	-	-	345 (<0.1%)	99.68	-	-	Yes
<i>Haminoea ortei</i>	Gastropoda	-	-	360 (<0.1%)	100	-	-	No
<i>Jorunna tomentosa</i>	Gastropoda	-	-	625 (<0.1%)	99.36	-	-	Yes
<i>Jujubinus striatus</i>	Gastropoda	-	-	1067 (<0.1%)	100	-	-	Yes
<i>Limapontia capitata</i>	Gastropoda	344 (<0.1%)	100	119 (<0.1%)	99.68	-	-	Yes
<i>Limapontia depressa</i>	Gastropoda	-	-	949 (<0.1%)	98.96	-	-	No
<i>Manzonina crassa</i>	Gastropoda	-	-	-	-	139 (<0.1%)	100	Yes
<i>Peringia ulvae</i>	Gastropoda	15973 (0.1%)	100	29230 (0.3%)	100	20 (<0.1%)	100	Yes
<i>Phorcus lineatus</i>	Gastropoda	-	-	2042 (<0.1%)	100	-	-	Yes
<i>Placida cremoniana</i>	Gastropoda	2664 (<0.1%)	99.21	-	-	-	-	No
<i>Polycera quadrilineata</i>	Gastropoda	-	-	36 (<0.1%)	99.04	-	-	Yes
<i>Pruvotfolia pselliotes</i>	Gastropoda	-	-	596 (<0.1%)	100	-	-	Yes
<i>Pusillina inconspicua</i>	Gastropoda	-	-	18 (<0.1%)	98.08	-	-	Yes
<i>Retusa umbilicata</i>	Gastropoda	174 (<0.1%)	99.74	-	-	-	-	Yes
<i>Rissoa auriscalpium</i>	Gastropoda	1193 (<0.1%)	99.73	-	-	-	-	No
<i>Rissoa guerinii</i>	Gastropoda	-	-	558 (<0.1%)	99.19	-	-	Yes
<i>Rissoa membranacea</i>	Gastropoda	16891 (0.1%)	100	-	-	-	-	Yes
<i>Steromphala cineraria</i>	Gastropoda	-	-	8262 (<0.1%)	100	-	-	Yes
<i>Steromphala umbilicalis</i>	Gastropoda	-	-	59 (<0.1%)	100	-	-	Yes
<i>Tergipes tergipes</i>	Gastropoda	63 (<0.1%)	100	53 (<0.1%)	99.36	-	-	Yes
<i>Tricolia pullus</i>	Gastropoda	189 (<0.1%)	100	660 (<0.1%)	99.36	-	-	Yes
<i>Tricolia saxatilis</i>	Gastropoda	-	-	-	-	23699 (0.2%)	100	No
<i>Tritia incrassata</i>	Gastropoda	-	-	841 (<0.1%)	100	-	-	Yes
<i>Tritia nitida</i>	Gastropoda	-	-	2743 (<0.1%)	99.65	-	-	No
<i>Tritia reticulata</i>	Gastropoda	-	-	3724 (<0.1%)	100	-	-	Yes
<i>Turritella communis</i>	Gastropoda	-	-	583 (<0.1%)	100	-	-	Yes
<i>Amathia gracilis</i>	Gymnolaemata	-	-	30 (<0.1%)	100	1932 (<0.1%)	100	Yes
<i>Bugula neritina</i>	Gymnolaemata	67 (<0.1%)	100	2821 (<0.1%)	100	1687 (<0.1%)	100	Yes
<i>Bugulina fulva</i>	Gymnolaemata	-	-	119 (<0.1%)	100	25728 (0.3%)	100	Yes
<i>Cryptosula pallasiana</i>	Gymnolaemata	-	-	-	-	3459 (<0.1%)	100	Yes
<i>Electra pilosa</i>	Gymnolaemata	726 (<0.1%)	100	4145 (<0.1%)	100	1155646 (11.5%)	100	Yes
<i>Flustrellidra hispida</i>	Gymnolaemata	-	-	95 (<0.1%)	100	11176 (0.1%)	100	Yes
<i>Membranipora membranacea</i>	Gymnolaemata	-	-	331 (<0.1%)	100	-	-	Yes
<i>Scruparia chelata</i>	Gymnolaemata	211 (<0.1%)	99.73	-	-	-	-	Yes
<i>Scrupocaberea maderensis</i>	Gymnolaemata	-	-	-	-	11967 (0.1%)	97.37	No
<i>Scrupocellaria scruposa</i>	Gymnolaemata	-	-	17 (<0.1%)	99.68	-	-	Yes
<i>Watersipora subatra</i>	Gymnolaemata	865 (<0.1%)	100	-	-	-	-	Yes

b. Supplementary figures

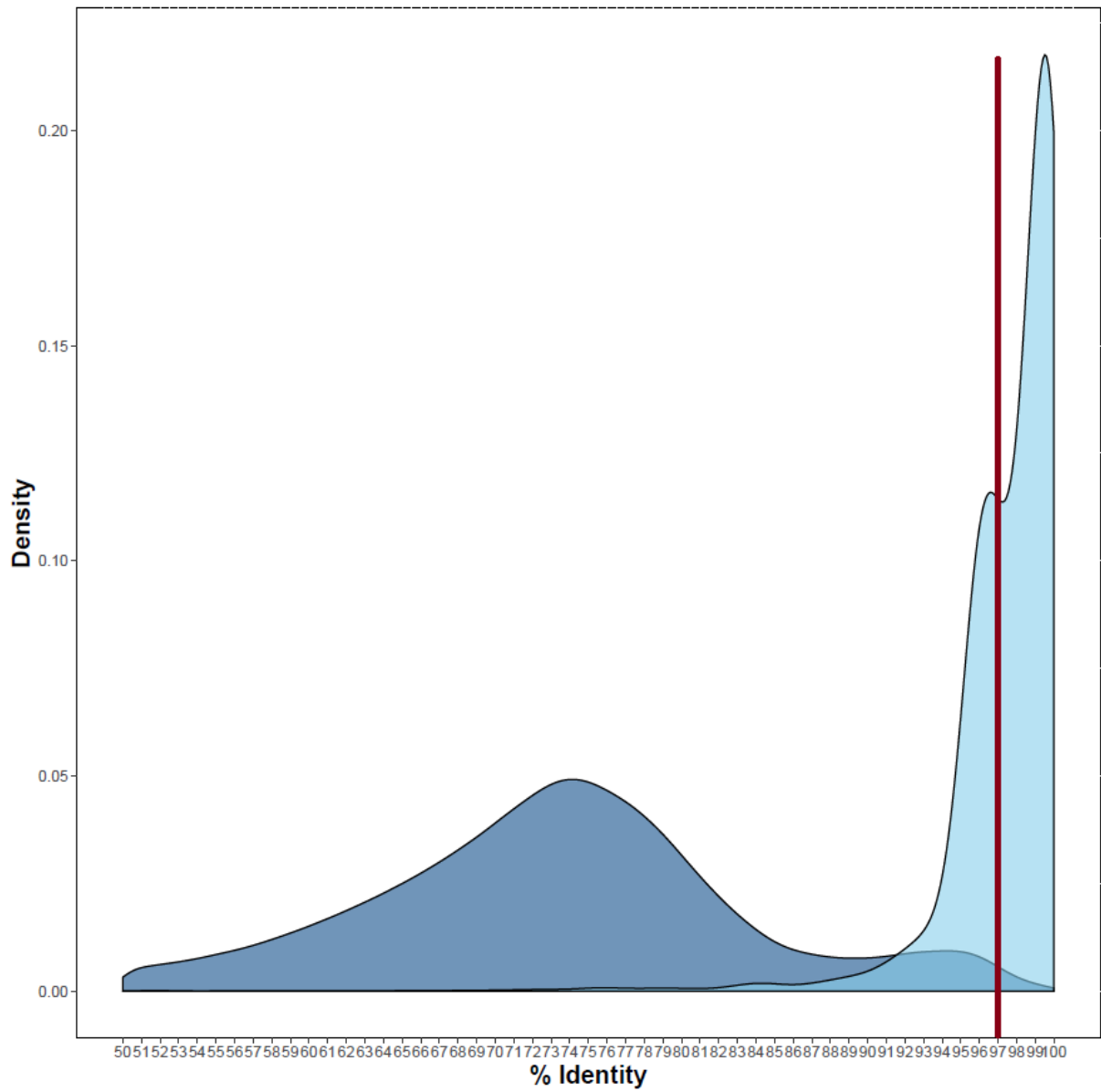


**Figure S1** Density plot of identity percentage distributions within the same species (light blue) and between species (dark blue) based on 18S reference sequences from the Mollusca and Bryozoa phyla as well as the class Ascidiacea. The red line marks the chosen identity threshold for assignment.

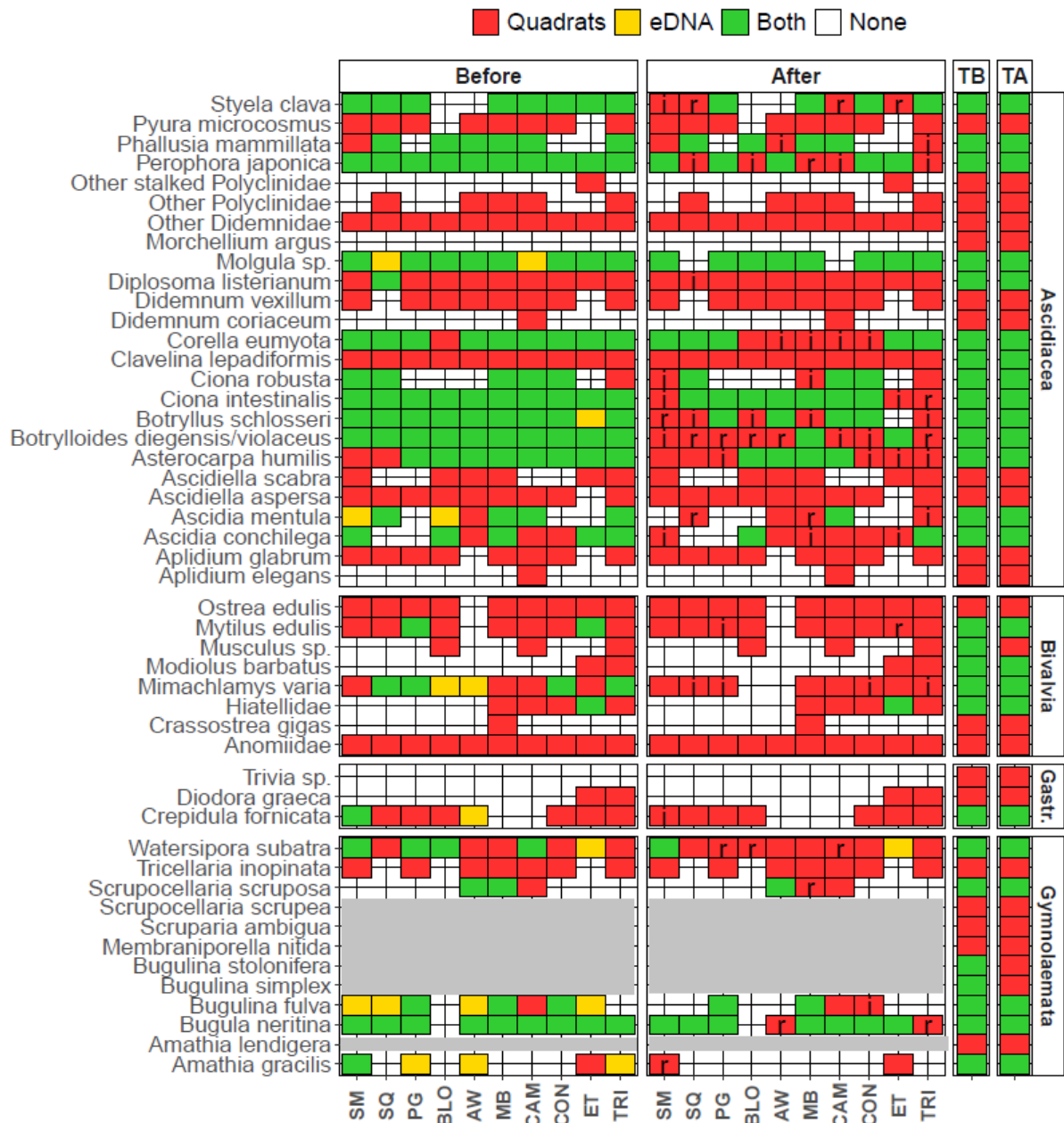


**Figure S2** Density plot of identity percentage distributions within the same species (light blue) and between species (dark blue) based on COI reference sequences from the Bryozoa phyla. The red line marks the chosen identity threshold for assignment.

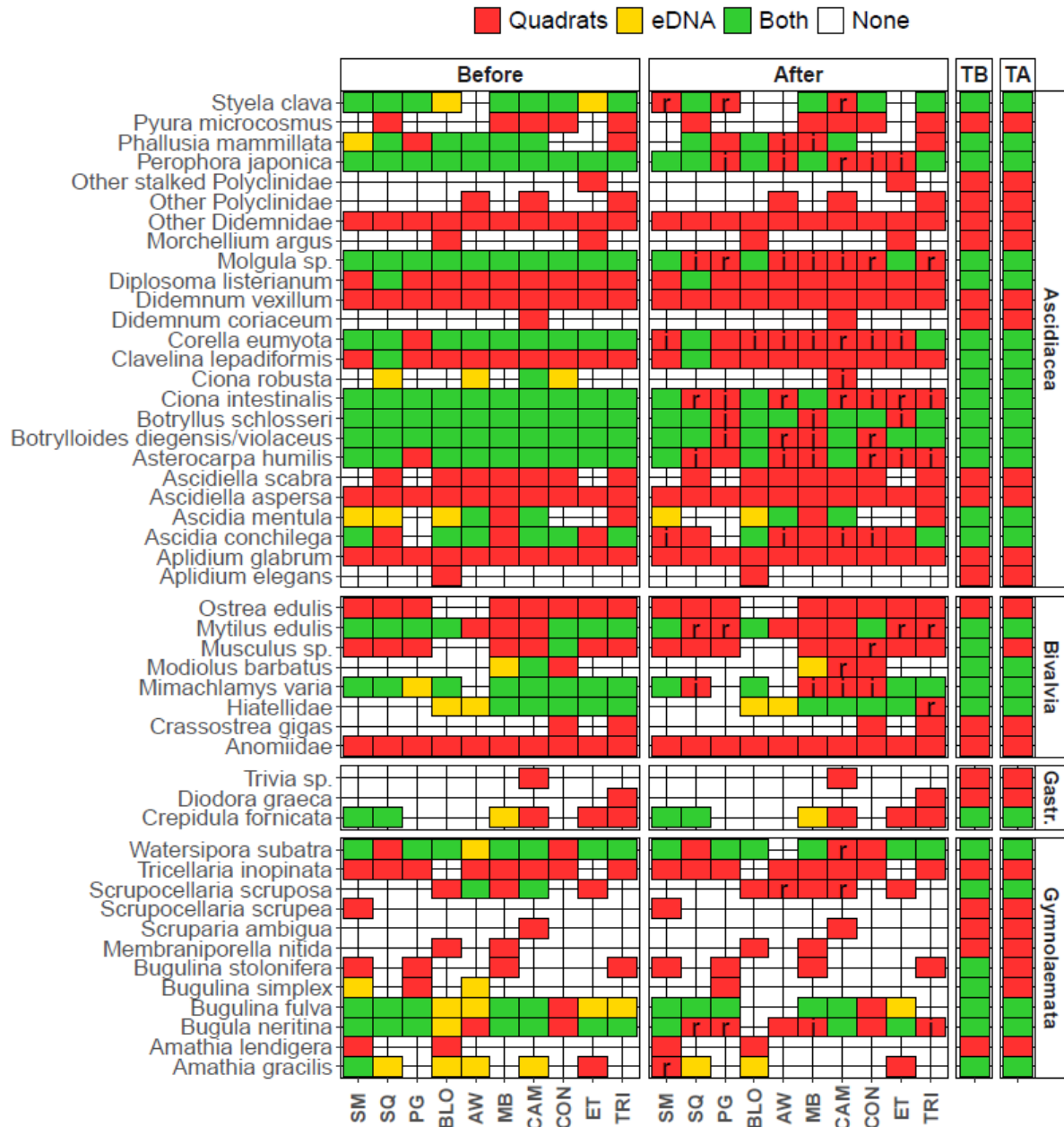




**Figure S3** Density plot of identity percentage distributions within the same species (light blue) and between species (dark blue) based on 16S reference sequences from the Mollusca and Bryozoa phyla as well as the class Ascidiacea. The red line marks the chosen identity threshold for assignment.



**Figure S4** Effect of the filtering steps (index-jump correction and filtering on the number of PCR replicates) on eDNA metabarcoding detection capacity. Metabarcoding data across all markers before and after filtering are displayed for samples from fall 2017. When a taxon disappeared after the filtering steps, a letter indicate whether its removal was due to index-jump correction (i) or PCR replicates filtering (r). The last two columns represent the presence of each taxon in the whole dataset before (TB) and after (TA) the filtering steps. See figure 1 for location codes.



**Figure S5** Effects of the filtering steps (index-jump correction and filtering on the number of PCR replicates) on the ability of eDNA metabarcoding to detect taxa. Metabarcoding data across all markers before and after filtering are displayed for samples of spring 2018 (see Fig. S4 for samples of fall 2017). When a taxon disappeared after the filtering steps, a letter indicate whether its removal was due to index-jump correction (i) or PCR replicates filtering (r). The last two columns represent the presence of each taxon in the whole dataset before (TB) and after (TA) the filtering steps. See figure 1 for location codes.

## Appendix 3: Supplementary material from chapter II.2

---

### a. Supplementary tables

**Table S1** List of values applied to the different parameters during read processing and filtering for each of the three markers used in this study. All parameters not indicated in this table were used with the default value.

Tool	Command	Parameter	18S	COI	16S
CUTADAPT	cutadapt	-e	0.16	0.12	0.14
DADA2	filterAndTrim	truncLen	270 and 180	240 and 180	90 and 90
		truncQ	0	0	0
		rm.phix	False	False	False
	dada	pool	pseudo	pseudo	pseudo
	seqtab	Length selection	335:400	303:323	105:125
SWARM	swarm	-d	1	1	1
		-f -b	500	100	1000
R	-	Index-jump	0.0182	0.0043	0.0021
		Replicates	2	2	2
VEGAN	rrarefy	Sample	97782	98201	32334

**Table S2** List of taxa included in each of the three metabarcoding subsets representing functional groups, for each marker.

Kingdom	Phylum	Subphylum	Class	Subclass	Superorder	Order	Family	AphiaID
<b>Chromista</b>	All	All	All	All	All	All	All	7
<b>Benthic</b>								
Metazoa	Arthropoda	Chelicerata	Pycnogonida	-	All	All	All	1302
		Crustacea	Hexanauplia	Thecostraca	-	All	All	22388
			Malacostraca	Eumalacostraca	Eucarida	Decapoda	All	1130
					Peracarida	All	All	1090
	Chordata	-	Ascidiacea	All	All	All	All	1839
	Cnidaria	-	Anthozoa	All	All	All	All	1292
			Hydrozoa	All	All	All	All	1337
	Annelida	-	Polychaeta	Sedentaria	All	All	All	754175
	Bryozoa	All	All	All	All	All	All	146142
	Echinodermata	All	All	All	All	All	All	1806
	Mollusca	-	Bivalvia	Autobranchia	-	Ostreida	Ostreidae	215
						Mytilida	Mytilidae	211
						Pectinida	Pectinidae	213
							Anomiidae	214
					Imparidentia	Adapedonta	Hiatellidae	251
			Gastropoda	All	All	All	All	101
			Polyplacophora	All	All	All	All	55
	Porifera	All	All	All	All	All	All	558
<b>Mobile</b>								
Metazoa	Arthropoda	Crustacea	Hexanauplia	Copepoda	All	All	All	1080
			Ostracoda	All	All	All	All	1078
			Ichthyostraca	All	All	All	All	845958
			Malacostraca	Eumalacostraca	Eucarida	Euphausiacea	All	1128
	Chordata	Vertebrata	Actinopterygii	All	All	All	All	10194
		Vertebrata	Elasmobranchii	All	All	All	All	10193
		Tunicata	Appendicularia	All	All	All	All	146421
	Cnidaria	Cubozoa	All	All	All	All	All	135219
		Scyphozoa	All	All	All	All	All	135220
		Staurozoa	All	All	All	All	All	265044
	Ctenophora	All	All	All	All	All	All	1248
	Mollusca	-	Cephalopoda	All	All	All	All	11707

**Table S3** List of all databases used for categorizing species according to their introduction status.

Database name	URL or reference
AquaNIS	<a href="http://www.corpi.ku.lt/databases/index.php/aquanis">http://www.corpi.ku.lt/databases/index.php/aquanis</a>
Bryozoan	<a href="http://www.bryozoa.net">http://www.bryozoa.net</a>
DORIS	<a href="http://doris.ffesm.fr/">http://doris.ffesm.fr/</a>
EASIN	<a href="https://easin.jrc.ec.europa.eu/easin">https://easin.jrc.ec.europa.eu/easin</a>
Encyclopedia of Marine Life of Britain and Ireland	<a href="http://www.habitas.org.uk/marinelife/">http://www.habitas.org.uk/marinelife/</a>
European Nature Information System (EUNIS)	<a href="https://eunis.eea.europa.eu/index.jsp">https://eunis.eea.europa.eu/index.jsp</a>
French National Inventory for MSFD - Descriptor D2 (invasive species)	Massé C. etGuérin L. (2018). Évaluation du descripteur 2 «espèces non-indigènes» en France Métropolitaine.Rapport scientifique pour l'évaluation 2018 au titre de la DCSMM. Muséum National d'Histoire Naturelle (UMS 2006 Patrimoine Naturel), stations marines de Dinardet d'Arcachon. 141p. <a href="http://resomar.cnrs.fr/">http://resomar.cnrs.fr/</a>
GBIF	<a href="https://www.gbif.org/">https://www.gbif.org/</a>
INPN	<a href="https://inpn.mnhn.fr">https://inpn.mnhn.fr</a>
Roscoff biological station inventory	<a href="http://abims.sb-roscoff.fr/inventaires">http://abims.sb-roscoff.fr/inventaires</a>
Marine species identification portal	<a href="http://species-identification.org/about.php">http://species-identification.org/about.php</a>
MARLIN	<a href="https://www.marlin.ac.uk/species">https://www.marlin.ac.uk/species</a>
NEMESIS	<a href="http://invasions.si.edu/nemesis/">http://invasions.si.edu/nemesis/</a>
OBIS	<a href="https://obis.org/">https://obis.org/</a>
Polychaete lab	<a href="https://thesimonpolychaetelab.com/">https://thesimonpolychaetelab.com/</a>
Sponges of the North East Atlantic 2.0	<a href="https://sponges-ne-atlantic.linnaeus.naturalis.nl">https://sponges-ne-atlantic.linnaeus.naturalis.nl</a>
Smithsonian Tropical Research Institute	<a href="https://stricollections.org/portal/collections/index.php">https://stricollections.org/portal/collections/index.php</a>
Sponges of Britain & Ireland	<a href="http://www.habitas.org.uk/marinelife/sponge_guide/">http://www.habitas.org.uk/marinelife/sponge_guide/</a>
WoRMS	<a href="http://www.marinespecies.org/index.php">http://www.marinespecies.org/index.php</a>

**Table S4** Comparison of assignments between all ASVs within one OTU and the assignment of this OTU. For each marker, the number of identical assignments between one ASV and its corresponding OTU is given. If the assignment was different, we specify if the two taxa were from the same genus, same family or from different families (“no match”). If either the ASV or the OTU was classified as “unassigned”, the comparison is listed in the corresponding line but if both were classified as “unassigned”, the comparison is classified as “same assignment”.

	18S		COI		16S	
	ASV	OTU	ASV	OTU	ASV	OTU
Same assignment	5741	3059	6654	3154	586	206
Same genus	30	16	30	16	14	7
Same family	7	6	14	8	4	4
No match	2	1	0	0	9	8
Unassigned	22	19	10	8	22	15
Total	5802	3101	6708	3186	635	240

**Table S5** List of all species belonging to the benthic functional group assigned to OTUs across all markers. Their introduction status is indicated in the last column. Species in bold were not previously recorded in our study area and their assignment is questionable due to the lack of reference sequence for a closely related native species for a given marker.

Class	Family	Species	Status
Anthozoa	Aiptasiidae	<b><i>Exaiptasia diaphana</i></b>	NIS
Anthozoa	Cerianthidae	<b><i>Pachycerianthus fimbriatus</i></b>	NIS
Anthozoa	Corallimorphidae	<b><i>Corynactis californica</i></b>	NIS
Anthozoa	Edwardsiidae	<i>Edwardsia longicornis</i>	native
Anthozoa	Edwardsiidae	<i>Edwardsia tuberculata</i>	cryptogenic
Anthozoa	Epizoanthidae	<b><i>Epizoanthus arenaceus</i></b>	NIS
Anthozoa	Hormathiidae	<i>Calliactis parasitica</i>	native
Asciacea	Asciidae	<i>Ascidia mentula</i>	native
Asciacea	Asciidae	<i>Phallusia mammillata</i>	native
Asciacea	Cionidae	<i>Ciona intestinalis</i>	native
Asciacea	Clavelinidae	<i>Clavelina lepadiformis</i>	native
Asciacea	Clavelinidae	<b><i>Clavelina meridionalis</i></b>	NIS
Asciacea	Corellidae	<i>Corella eumyota</i>	NIS
Asciacea	Molgulidae	<i>Molgula bleizi</i>	native
Asciacea	Molgulidae	<i>Molgula complanata</i>	native
Asciacea	Molgulidae	<i>Molgula socialis</i>	native
Asciacea	Perophoridae	<i>Perophora japonica</i>	NIS
Asciacea	Styelidae	<i>Asterocarpa humilis</i>	NIS
Asciacea	Styelidae	<i>Botrylloides diegensis</i>	NIS
Asciacea	Styelidae	<i>Botrylloides leachii</i>	native
Asciacea	Styelidae	<i>Botryllus schlosseri</i>	cryptogenic
Asciacea	Styelidae	<i>Dendrodoa grossularia</i>	native

Ascidacea	Styelidae	<i>Metandrocarpa taylori</i>	NIS
Ascidacea	Styelidae	<i>Pelonaia corrugata</i>	NIS
Ascidacea	Styelidae	<i>Polycarpa pomaria</i>	native
Ascidacea	Styelidae	<b><i>Polycarpa tenera</i></b>	NIS
Ascidacea	Styelidae	<i>Styela clava</i>	NIS
Asteroidea	Asteriidae	<i>Asterias rubens</i>	native
Asteroidea	Asteriidae	<i>Marthasterias glacialis</i>	native
Bivalvia	Hiatellidae	<i>Hiatella arctica</i>	native
Bivalvia	Mytilidae	<i>Modiolula phaseolina</i>	native
Bivalvia	Mytilidae	<i>Modiolus barbatus</i>	native
Bivalvia	Mytilidae	<i>Mytilus edulis</i>	native
Bivalvia	Mytilidae	<b><i>Mytilus trossulus</i></b>	NIS
Bivalvia	Pectinidae	<i>Aequipecten opercularis</i>	native
Bivalvia	Pectinidae	<i>Mimachlamys varia</i>	native
Calcarea	Baeriidae	<i>Leuconia nivea</i>	native
Calcarea	Grantiidae	<i>Leucandra aspera</i>	cryptogenic
Calcarea	Sycettidae	<i>Sycon ciliatum</i>	native
Crinoidea	Antedonidae	<i>Antedon bifida</i>	native
Demospongiae	Chalinidae	<i>Haliclona (Haliclona) oculata</i>	native
Demospongiae	Chalinidae	<i>Haliclona (Haliclona) simulans</i>	native
Demospongiae	Chalinidae	<i>Haliclona (Reniera) aqueductus</i>	NIS
Demospongiae	Chalinidae	<i>Haliclona (Reniera) cinerea</i>	native
Demospongiae	Chalinidae	<b><i>Haliclona (Rhizoniera) curacaoensis</i></b>	NIS
Demospongiae	Chalinidae	<i>Haliclona (Soestella) xena</i>	native
Demospongiae	Clionidae	<i>Cliona celata</i>	native
Demospongiae	Clionidae	<b><i>Sphaciospongia vesparium</i></b>	NIS
Demospongiae	Esperiopsidae	<i>Amphilectus fucorum</i>	native
Demospongiae	Halichondriidae	<i>Halichondria (Halichondria) panicea</i>	native
Demospongiae	Hymedesmiidae	<i>Phorbas dives</i>	native
Demospongiae	Hymedesmiidae	<i>Phorbas plumosus</i>	native
Demospongiae	Hymedesmiidae	<i>Spanioplou armaturum</i>	native
Demospongiae	Microcionidae	<i>Ophlitaspongia papilla</i>	native
Demospongiae	Mycalidae	<i>Mycale (Carmia) macilenta</i>	native
Demospongiae	Myxillidae	<i>Myxilla (Myxilla) rosacea</i>	native
Demospongiae	Myxillidae	<i>Myxilla (Styloptilon) ancorata</i>	native
Demospongiae	Suberitidae	<i>Protosuberites denhartogi</i>	native
Echinoidea	Echinocyamidae	<i>Echinocyamus pusillus</i>	native
Echinoidea	Loveniidae	<i>Echinocardium cordatum</i>	native
Echinoidea	Parechinidae	<i>Psammechinus miliaris</i>	native
Echinoidea	Spatangidae	<i>Spatangus purpureus</i>	native
Gastropoda	Akeridae	<i>Akera bullata</i>	native
Gastropoda	Aplysiidae	<b><i>Aplysia parvula</i></b>	cryptogenic
Gastropoda	Caecidae	<i>Caecum trachea</i>	native
Gastropoda	Calliostomatidae	<i>Calliostoma zizyphinum</i>	native
Gastropoda	Calyptraeidae	<i>Calyptraea chinensis</i>	native
Gastropoda	Calyptraeidae	<i>Crepidula fornicata</i>	NIS



Gastropoda	Cerithiidae	<i>Bittium reticulatum</i>	native
Gastropoda	Cerithiopsidae	<b><i>Cerithiopsis petanii</i></b>	NIS
Gastropoda	Discodorididae	<i>Jorunna tomentosa</i>	native
Gastropoda	Embletoniidae	<i>Embletonia pulchra</i>	native
Gastropoda	Facelinidae	<i>Favorinus branchialis</i>	native
Gastropoda	Facelinidae	<i>Pruvotfolia pselliotes</i>	native
Gastropoda	Flabellinidae	<i>Fjordia lineata</i>	native
Gastropoda	Haminoeidae	<b><i>Haminoea ortei</i></b>	NIS
Gastropoda	Hydrobiidae	<i>Peringia ulvae</i>	native
Gastropoda	Limapontiidae	<i>Limapontia capitata</i>	native
Gastropoda	Limapontiidae	<i>Placida cremoniana</i>	cryptogenic
Gastropoda	Mangeliidae	<b><i>Eucithara coronata</i></b>	NIS
Gastropoda	Nassariidae	<i>Tritia incrassata</i>	native
Gastropoda	Nassariidae	<i>Tritia nitida</i>	native
Gastropoda	Nassariidae	<i>Tritia reticulata</i>	native
Gastropoda	Patellidae	<i>Patella pellucida</i>	native
Gastropoda	Phasianellidae	<i>Tricolia pullus</i>	native
Gastropoda	Phasianellidae	<i>Tricolia saxatilis</i>	NIS
Gastropoda	Plakobranchidae	<i>Elysia viridis</i>	native
Gastropoda	Polyceridae	<i>Polycera quadrilineata</i>	native
Gastropoda	Retusidae	<i>Retusa umbilicata</i>	native
Gastropoda	Rissoidae	<i>Alvania tenera</i>	undetermined
Gastropoda	Rissoidae	<i>Manzonina crassa</i>	native
Gastropoda	Rissoidae	<i>Rissoa auriscalpium</i>	NIS
Gastropoda	Rissoidae	<i>Rissoa guerinii</i>	native
Gastropoda	Rissoidae	<i>Rissoa membranacea</i>	native
Gastropoda	Tergipedidae	<i>Tergipes tergipes</i>	native
Gastropoda	Trochidae	<i>Gibbula magus</i>	native
Gastropoda	Trochidae	<i>Jujubinus striatus</i>	native
Gastropoda	Trochidae	<i>Phorcus lineatus</i>	native
Gastropoda	Trochidae	<i>Steromphala cineraria</i>	native
Gastropoda	Trochidae	<i>Steromphala umbilicalis</i>	native
Gastropoda	Turritellidae	<i>Turritella communis</i>	native
Gymnolaemata	Bugulidae	<i>Bugula neritina</i>	NIS
Gymnolaemata	Bugulidae	<i>Bugulina fulva</i>	cryptogenic
Gymnolaemata	Bugulidae	<i>Bugulina simplex</i>	NIS
Gymnolaemata	Candidae	<i>Scrupocellaria scruposa</i>	native
Gymnolaemata	Cryptosulidae	<i>Cryptosula pallasiana</i>	native
Gymnolaemata	Electridae	<i>Electra pilosa</i>	native
Gymnolaemata	Flustrellidridae	<i>Flustrellidra hispida</i>	native
Gymnolaemata	Membraniporidae	<i>Membranipora membranacea</i>	native
Gymnolaemata	Scrupariidae	<i>Scruparia chelata</i>	native
Gymnolaemata	Vesiculariidae	<i>Amathia gracilis</i>	native
Gymnolaemata	Watersiporidae	<i>Watersipora subatra</i>	NIS
Hexanauplia	Archaeobalanidae	<i>Semibalanus balanoides</i>	native
Hexanauplia	Austrobalanidae	<i>Austrominius modestus</i>	NIS

Hexanauplia	Balanidae	<i>Perforatus perforatus</i>	native
Hexanauplia	Chthamalidae	<i>Chthamalus montagui</i>	native
Hexanauplia	Chthamalidae	<i>Chthamalus stellatus</i>	native
Hexanauplia	Verrucidae	<i>Verruca stroemia</i>	native
Holothuroidea	Synaptidae	<i>Leptosynapta inhaerens</i>	native
Holothuroidea	Synaptidae	<i>Oestergrenia digitata</i>	native
Hydrozoa	Aglaopheniidae	<i>Aglaophenia pluma</i>	native
Hydrozoa	Aglaopheniidae	<i>Aglaophenia tubiformis</i>	native
Hydrozoa	Bougainvilliidae	<i>Bougainvillia muscus</i>	native
Hydrozoa	Bougainvilliidae	<i>Dicoryne conybeari</i>	native
Hydrozoa	Campanulariidae	<i>Clytia hemisphaerica</i>	native
Hydrozoa	Campanulariidae	<i>Clytia paulensis</i>	cryptogenic
Hydrozoa	Campanulariidae	<i>Gonothyrea loveni</i>	NIS
Hydrozoa	Campanulariidae	<i>Laomedea flexuosa</i>	native
Hydrozoa	Campanulariidae	<i>Obelia dichotoma</i>	native
Hydrozoa	Campanulariidae	<i>Obelia geniculata</i>	native
Hydrozoa	Campanulariidae	<i>Obelia longissima</i>	native
Hydrozoa	Campanulinidae	<i>Calycella syringa</i>	native
Hydrozoa	Cladonematidae	<i>Eleutheria dichotoma</i>	native
Hydrozoa	Corynidae	<i>Coryne muscoides</i>	native
Hydrozoa	Corynidae	<i>Coryne pusilla</i>	native
Hydrozoa	Diphyidae	<i>Muggiaea atlantica</i>	cryptogenic
Hydrozoa	Haleciidae	<i>Halecium halecinum</i>	native
Hydrozoa	Hydractiniidae	<i>Clava multicornis</i>	native
Hydrozoa	Kirchenpaueriidae	<i>Kirchenpaueria pinnata</i>	native
Hydrozoa	Mitrocomidae	<i>Mitrocomella brownei</i>	native
Hydrozoa	Pandeidae	<i>Amphinema dinema</i>	native
Hydrozoa	Pandeidae	<i>Leuckartiara octona</i>	cryptogenic
Hydrozoa	Plumulariidae	<i>Plumularia setacea</i>	native
Hydrozoa	Rathkeidae	<i>Lizzia blondina</i>	cryptogenic
Hydrozoa	Sertulariidae	<i>Dynamena pumila</i>	native
Hydrozoa	Sertulariidae	<i>Hydrallmania falcata</i>	native
Hydrozoa	Sertulariidae	<i>Sertularia cupressina</i>	native
Malacostraca	Carcinidae	<i>Carcinus maenas</i>	native
Malacostraca	Dexaminidae	<i>Dexamine spinosa</i>	native
Malacostraca	Ischyroceridae	<i>Jassa slatteryi</i>	cryptogenic
Malacostraca	Nuuanuidae	<i>Gammarella fucicola</i>	native
Malacostraca	Palaemonidae	<b><i>Crinotonia attenuatus</i></b>	NIS
Malacostraca	Palaemonidae	<i>Palaemon elegans</i>	native
Malacostraca	Polybiidae	<i>Necora puber</i>	native
NA	Myzostomatidae	<i>Myzostoma cirriferum</i>	native
Ophiuroidea	Amphiuridae	<i>Amphipholis squamata</i>	native
Ophiuroidea	Amphiuridae	<i>Amphiura filiformis</i>	native
Ophiuroidea	Ophiotomidae	<i>Ophiocomina nigra</i>	native
Ophiuroidea	Ophiotrichidae	<b><i>Ophiothrix (Ophiothrix) oerstedii</i></b>	NIS
Ophiuroidea	Ophiotrichidae	<i>Ophiothrix fragilis</i>	native

Polychaeta	Ampharetidae	<i>Ampharete santillani</i>	native
Polychaeta	Chaetopteridae	<i>Chaetopterus variopedatus</i>	native
Polychaeta	Pectinariidae	<i>Amphictene auricoma</i>	native
Polychaeta	Pectinariidae	<i>Lagis koreni</i>	native
Polychaeta	Poecilochaetidae	<i>Poecilochaetus serpens</i>	native
Polychaeta	Sabellariidae	<i>Sabellaria spinulosa</i>	native
Polychaeta	Sabellidae	<i>Parasabella saxicola</i>	native
Polychaeta	Sabellidae	<i>Sabella pavonina</i>	native
Polychaeta	Serpulidae	<i>Hydroides norvegica</i>	native
Polychaeta	Serpulidae	<i>Laeospira corallinae</i>	native
Polychaeta	Serpulidae	<b><i>Neodexiospira alveolata</i></b>	NIS
Polychaeta	Serpulidae	<i>Spirobranchus triqueter</i>	native
Polychaeta	Serpulidae	<i>Spirorbis (Spirorbis) rupestris</i>	native
Polychaeta	Serpulidae	<i>Spirorbis (Spirorbis) spirorbis</i>	native
Polychaeta	Serpulidae	<b><i>Vermiliopsis striaticeps</i></b>	NIS
Polychaeta	Spionidae	<i>Aonides oxycephala</i>	native
Polychaeta	Spionidae	<i>Dipolydora capensis</i>	NIS
Polychaeta	Spionidae	<i>Laonice cirrata</i>	native
Polychaeta	Spionidae	<i>Malacoceros fuliginosus</i>	native
Polychaeta	Spionidae	<i>Polydora cornuta</i>	cryptogenic
Polychaeta	Spionidae	<i>Polydora hoplura</i>	native
Polychaeta	Spionidae	<i>Pseudopolydora paucibranchiata</i>	NIS
Polychaeta	Spionidae	<i>Pygospio elegans</i>	native
Polychaeta	Spionidae	<b><i>Scolecopsis laoncola</i></b>	NIS
Polychaeta	Spionidae	<i>Spiophanes bombyx</i>	cryptogenic
Polychaeta	Spionidae	<i>Streblospio benedicti</i>	NIS
Polychaeta	Terebellidae	<i>Amphitritides gracilis</i>	native
Polychaeta	Terebellidae	<i>Artacama proboscidea</i>	native
Polychaeta	Terebellidae	<i>Lanice conchilega</i>	native
Polychaeta	Terebellidae	<i>Neoamphitrite figulus</i>	native
Polychaeta	Terebellidae	<i>Terebella lapidaria</i>	native
Polychaeta	Terebellidae	<i>Thelepus cincinnatus</i>	native
Polychaeta	Trichobanchidae	<i>Terebellides stroemii</i>	native
Polychaeta	Trichobanchidae	<i>Trichobanchus glacialis</i>	native
Polyplacophora	Callochitonidae	<i>Callochiton septemvalvis</i>	native
Polyplacophora	Lepidochitonidae	<i>Lepidochitona cinerea</i>	native
Stenolaemata	Tubuliporidae	<i>Tubulipora liliacea</i>	native

**Table S6** Results of PERMANOVA and pairwise tests performed on the subsets comprising all OTUs without taxonomic assignment for each of the three markers. Community compositions were compared between biogeographic regions, between marinas within regions, and between seasons. PERMANOVA were performed with the adonis2 function of the VEGAN-2.5.2 R package, with 9999 permutations. Factors were added sequentially in the order presented below. Pairwise tests were computed using the pairwise.adonis2 function from the PAIRWISE.ADONIS R package. Due to the positive interaction between seasons and regions, permutations were constrained within seasons for each region (9999 permutations).

PERMANOVA															
	18S					COI					16S				
	df	Sum of squares	R <sup>2</sup>	F	P	df	Sum of squares	R <sup>2</sup>	F	P	df	Sum of squares	R <sup>2</sup>	F	P
Region	2	4.238	0.135	13.625	<0.001	2	3.620	0.147	16.479	<0.001	2	2.339	0.086	2.935	<0.001
Season	1	4.432	0.142	28.494	<0.001	1	4.773	0.194	43.457	<0.001	1	2.209	0.081	5.543	<0.001
Marina (Region)	7	9.205	0.294	8.456	<0.001	7	6.081	0.248	7.909	<0.001	7	8.328	0.305	2.986	<0.001
Region x Season	2	3.124	0.100	10.045	<0.001	2	2.932	0.119	13.347	<0.001	2	1.635	0.060	2.052	<0.001
Marina (Region) x Season	7	7.170	0.229	6.586	<0.001	7	4.945	0.201	6.432	<0.001	7	4.821	0.177	1.728	<0.001
Residuals	20	3.111	0.099			20	2.197	0.089			20	7.970	0.292		
Pairwise tests															
	18S					COI					16S				
	df	Sum of squares	R <sup>2</sup>	F	P	df	Sum of squares	R <sup>2</sup>	F	P	df	Sum of squares	R <sup>2</sup>	F	P
WEC vs. SB	1	1.366	0.124	4.243	<0.001	1	1.288	0.146	5.130	<0.001	1	0.797	0.100	3.322	<0.001
WEC vs. IS	1	1.046	0.110	3.222	<0.001	1	0.838	0.116	3.399	<0.001	1	0.479	0.071	1.997	0.003
SB vs. IS	1	0.972	0.147	3.099	<0.001	1	0.655	0.128	2.647	<0.001	1	0.436	0.096	1.916	0.011

**Table S7** Results of PERMANOVA performed on the three subsets describing “functional” groups for each of the three markers. Community compositions were compared between biogeographic regions, between marinas within regions, and between seasons. PERMANOVA were performed with the adonis2 function of the VEGAN-2.5.2 R package, with 9999 permutations. Factors were added sequentially in the order presented below.

18S															
	Chromista					Mobile					Benthic				
	df	Sum of squares	R <sup>2</sup>	F	P	df	Sum of squares	R <sup>2</sup>	F	P	df	Sum of squares	R <sup>2</sup>	F	P
Region	2	3.988	0.127	22.645	<0.001	2	3.972	0.122	4.456	<0.001	2	2.501	0.077	2.846	<0.001
Season	1	5.015	0.160	56.954	<0.001	1	2.446	0.075	5.489	<0.001	1	3.456	0.106	7.867	<0.001
Marina (Region)	7	9.001	0.288	14.603	<0.001	7	9.964	0.307	3.194	<0.001	7	9.336	0.286	3.036	<0.001
Region x Season	2	3.301	0.106	18.748	<0.001	2	1.610	0.050	1.807	0.014	2	1.912	0.059	2.176	<0.001
Marina (Region) x Season	7	8.218	0.263	13.334	<0.001	7	5.570	0.172	1.786	<0.001	7	6.635	0.203	2.158	<0.001
Residuals	20	1.761	0.057			20	8.913	0.274			20	8.786	0.269		
COI															
	Chromista					Mobile					Benthic				
	df	Sum of squares	R <sup>2</sup>	F	P	df	Sum of squares	R <sup>2</sup>	F	P	df	Sum of squares	R <sup>2</sup>	F	P
Region	2	4.139	0.145	45.947	<0.001	2	2.965	0.094	3.317	<0.001	2	2.851	0.088	2.684	<0.001
Season	1	6.167	0.216	136.94	<0.001	1	3.376	0.108	7.553	<0.001	1	2.093	0.064	3.941	<0.001
Marina (Region)	7	6.771	0.237	21.479	<0.001	7	9.406	0.300	3.007	<0.001	7	8.102	0.249	2.180	<0.001
Region x Season	2	4.425	0.155	49.129	<0.001	2	1.468	0.047	1.642	0.036	2	2.117	0.065	1.994	<0.001
Marina (Region) x Season	7	6.183	0.216	19.613	<0.001	7	5.249	0.167	1.678	0.003	7	6.763	0.208	1.819	<0.001
Residuals	20	0.901	0.032			20	8.938	0.285			20	10.621	0.326		
16S															
	Chromista					Mobile					Benthic				
	df	Sum of squares	R <sup>2</sup>	F	P	df	Sum of squares	R <sup>2</sup>	F	P	df	Sum of squares	R <sup>2</sup>	F	P
Region						2	2.006	0.087	3.528	<0.001	2	2.402	0.103	3.284	<0.001
Season						1	0.585	0.025	2.057	0.042	1	2.318	0.099	6.340	<0.001
Marina (Region)						7	9.663	0.420	4.856	<0.001	7	6.303	0.270	2.463	<0.001
Region x Season						2	1.072	0.047	1.885	0.027	2	1.333	0.057	1.823	0.022
Marina (Region) x Season						7	4.005	0.174	2.013	<0.001	7	3.699	0.158	1.445	0.027
Residuals						20	5.685	0.247			20	7.313	0.313		

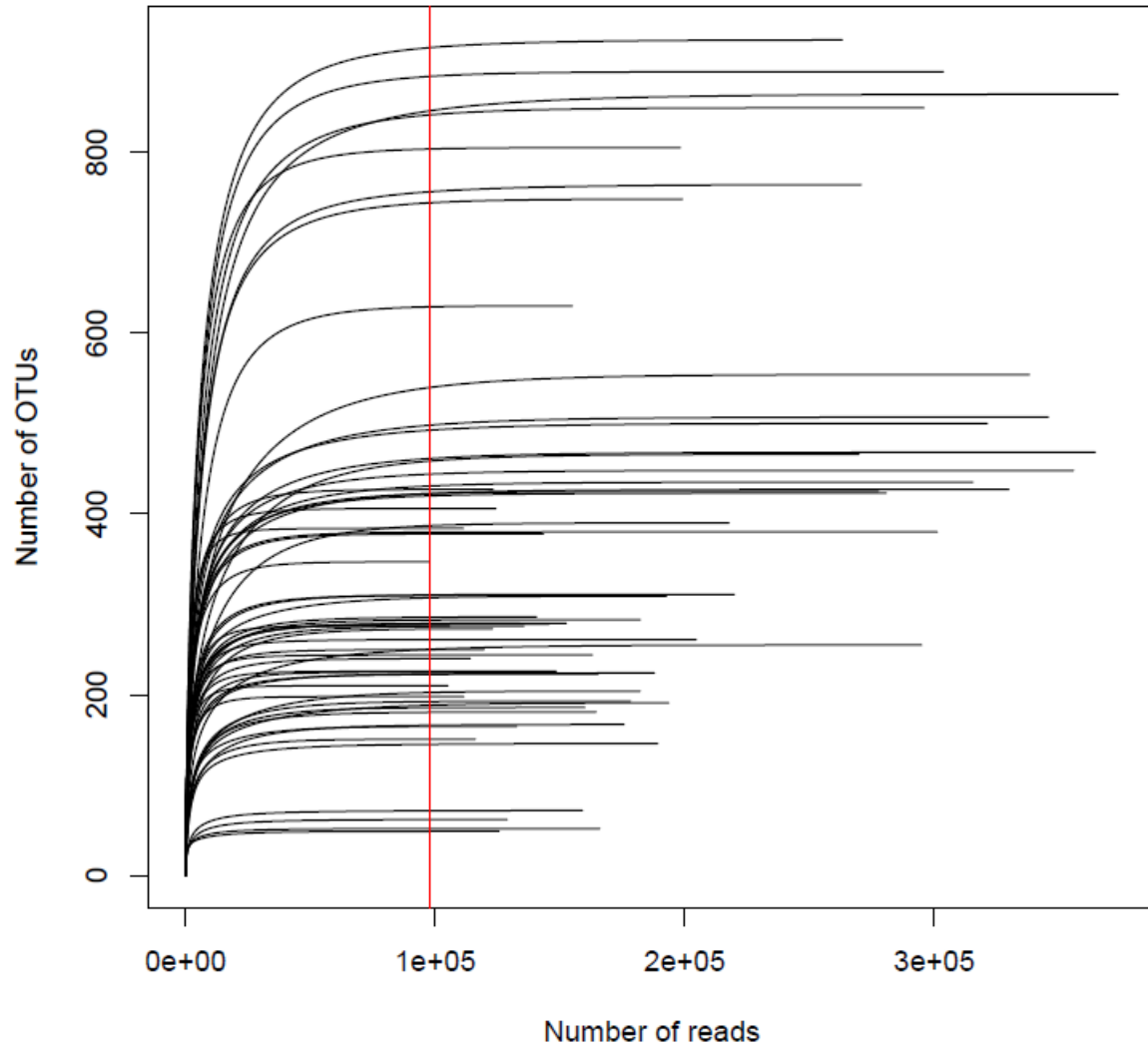
**Table S8** Results of pairwise tests performed on the three subsets describing “functional” groups for each of the three markers. Pairwise tests were computed using the pairwise.adonis2 function from the PAIRWISE.ADONIS R package. Due to the positive interaction between seasons and regions (see Table S10), permutations were constrained within seasons for each region (9999 permutations). Non-significant values are in bold.

18S																	
	Chromista						Mobile						Benthic				
	df	Sum of squares	R <sup>2</sup>	F	P		df	Sum of squares	R <sup>2</sup>	F	P		df	Sum of squares	R <sup>2</sup>	F	P
WEC vs. SB	1	1.545	0.139	4.837	<0.001		1	0.904	0.076	2.481	0.004		1	0.734	0.071	2.125	0.002
WEC vs. IS	1	1.049	0.112	3.272	<0.001		1	1.213	0.116	3.417	<0.001		1	0.707	0.075	1.946	0.002
SB vs. IS	1	0.998	0.151	3.193	<0.001		1	0.871	0.138	2.890	0.001		1	0.642	0.092	1.833	0.018
COI																	
	Chromista						Mobile						Benthic				
	df	Sum of squares	R <sup>2</sup>	F	P		df	Sum of squares	R <sup>2</sup>	F	P		df	Sum of squares	R <sup>2</sup>	F	P
WEC vs. SB	1	1.239	0.145	5.085	<0.001		1	1.133	0.096	3.198	<0.001		1	0.823	0.075	2.436	<0.001
WEC vs. IS	1	0.812	0.110	3.206	<0.001		1	0.687	0.067	1.880	0.038		1	0.639	0.067	1.880	0.001
SB vs. IS	1	0.600	0.125	2.578	<0.001		1	0.573	0.092	1.821	0.020		1	0.701	0.106	2.145	<0.001
16S																	
	Chromista						Mobile						Benthic				
	df	Sum of squares	R <sup>2</sup>	F	P		df	Sum of squares	R <sup>2</sup>	F	P		df	Sum of squares	R <sup>2</sup>	F	P
WEC vs. SB							<b>1</b>	<b>0.398</b>	<b>0.062</b>	<b>1.987</b>	<b>0.080</b>		1	1.009	0.121	4.133	<0.001
WEC vs. IS							<b>1</b>	<b>0.358</b>	<b>0.067</b>	<b>1.858</b>	<b>0.074</b>		1	0.607	0.085	2.406	0.002
SB vs. IS							1	0.474	0.131	2.713	0.033		1	<b>0.257</b>	<b>0.060</b>	<b>1.149</b>	<b>0.171</b>

**Table S9** Local contributions to beta diversity (LCBD) calculated for each season separately and for two datasets (Quadrats, taxa identified from quadrat sampling; eDNA all markers, benthic taxa identified from OTUs across all markers combined). P values were adjusted with the Holm correction. Significant values are in bold.

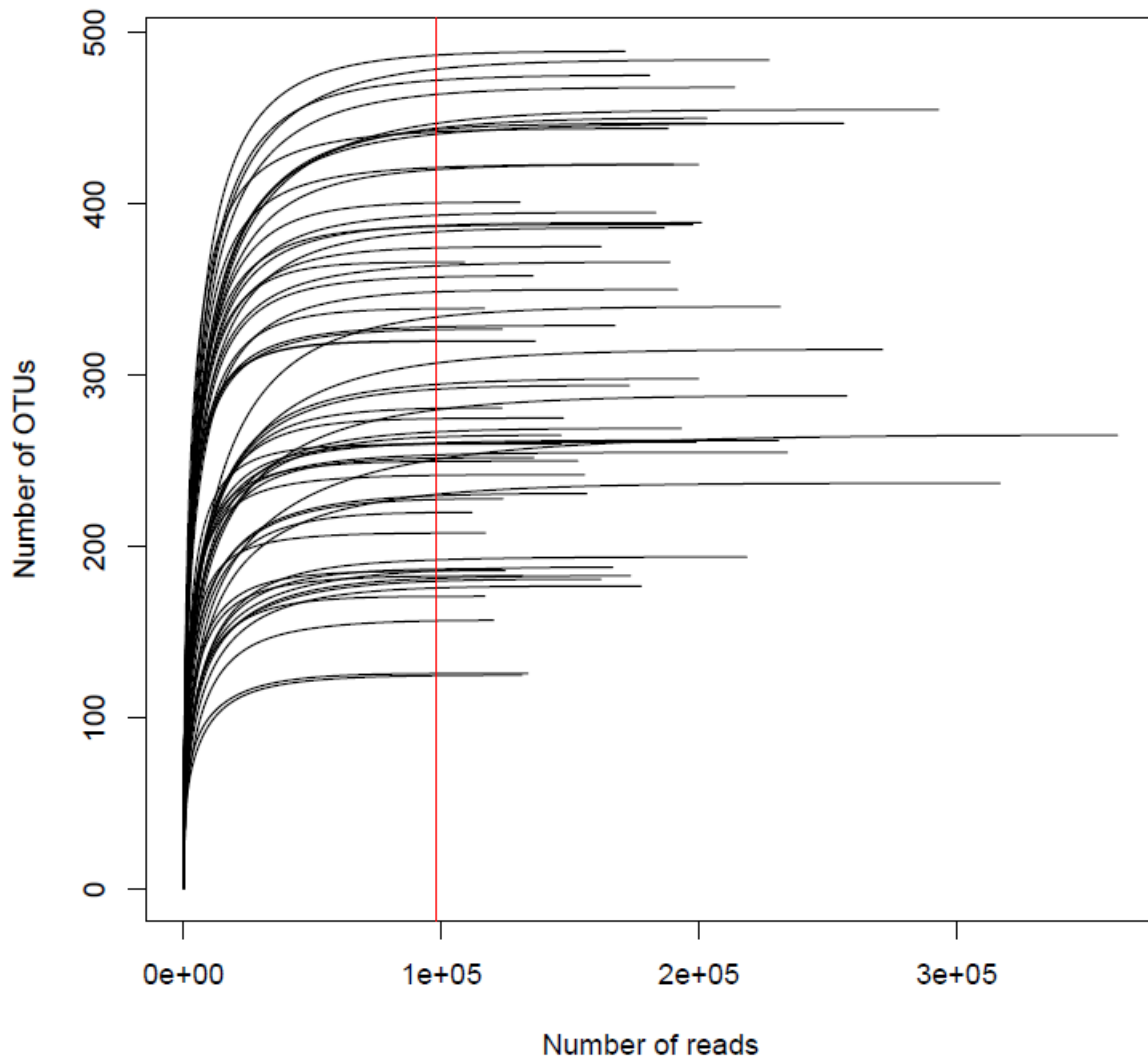
Marina	Pontoon	Season	Quadrats		eDNA all markers	
			LCBD	Adjusted P	LCBD	Adjusted P
AW	1	Fall	0.076	0.324	0.059	0.822
AW	2	Fall	0.076	0.318	0.053	1.000
BL	1	Fall	0.068	0.826	0.056	1.000
BL	2	Fall	0.056	1.000	0.057	1.000
CAM	1	Fall	0.047	1.000	0.055	1.000
CAM	2	Fall	0.043	1.000	0.056	1.000
CON	1	Fall	0.030	1.000	0.048	1.000
CON	2	Fall	0.034	1.000	0.050	1.000
Et	1	Fall	0.064	1.000	0.054	1.000
Et	2	Fall	0.073	0.318	0.049	1.000
MB	1	Fall	0.040	1.000	0.047	1.000
MB	2	Fall	0.034	1.000	0.043	1.000
PG	1	Fall	0.043	1.000	0.044	1.000
PG	2	Fall	0.054	1.000	0.043	1.000
SM	1	Fall	0.037	1.000	0.048	1.000
SM	2	Fall	0.026	1.000	0.044	1.000
SQ	1	Fall	0.053	1.000	0.054	1.000
SQ	2	Fall	0.042	1.000	0.048	1.000
TR	1	Fall	0.071	0.374	0.047	1.000
TR	2	Fall	0.033	1.000	0.046	1.000
AW	1	Spring	0.070	0.792	<b>0.062</b>	<b>0.036</b>
AW	2	Spring	0.066	1.000	0.053	1.000
BL	1	Spring	0.064	1.000	0.042	1.000
BL	2	Spring	0.080	0.062	0.045	1.000
CAM	1	Spring	0.070	0.792	0.047	1.000
CAM	2	Spring	0.050	1.000	0.053	1.000
CON	1	Spring	0.047	1.000	0.051	1.000
CON	2	Spring	0.029	1.000	0.058	0.385
Et	1	Spring	0.065	1.000	0.048	1.000
Et	2	Spring	0.065	0.988	0.046	1.000
MB	1	Spring	0.030	1.000	0.048	1.000
MB	2	Spring	0.031	1.000	0.046	1.000
PG	1	Spring	0.046	1.000	0.061	0.068
PG	2	Spring	0.039	1.000	0.055	1.000
SM	1	Spring	0.041	1.000	0.047	1.000
SM	2	Spring	0.039	1.000	0.047	1.000
SQ	1	Spring	0.039	1.000	0.047	1.000
SQ	2	Spring	0.044	1.000	0.048	1.000
TR	1	Spring	0.028	1.000	0.047	1.000
TR	2	Spring	0.056	1.000	0.050	1.000

b. Supplementary figures

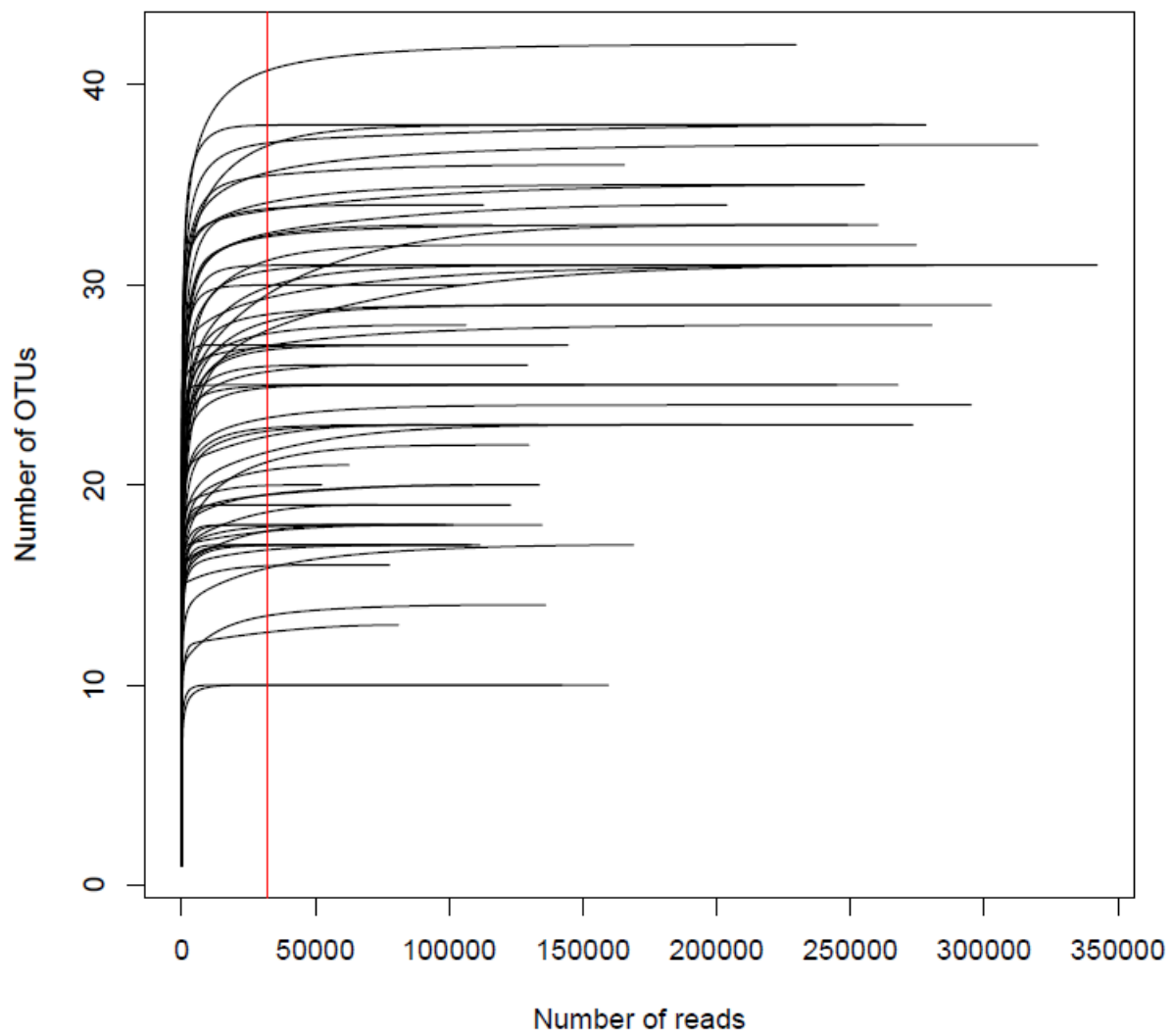


**Figure S1** Rarefaction curves on OTUs obtained with the 18S marker for all samples. The red line indicates the number of reads chosen as a cut-off.

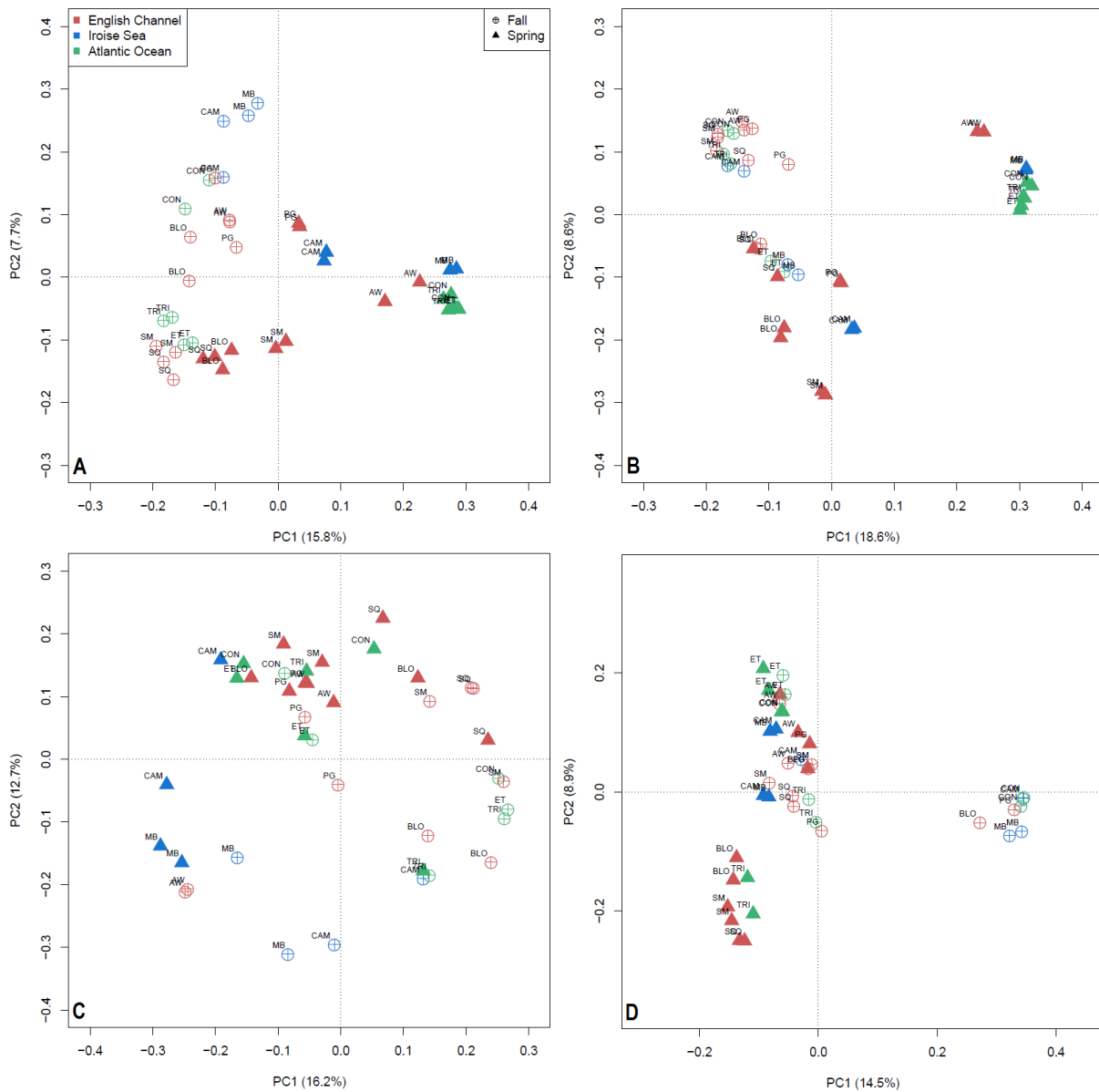




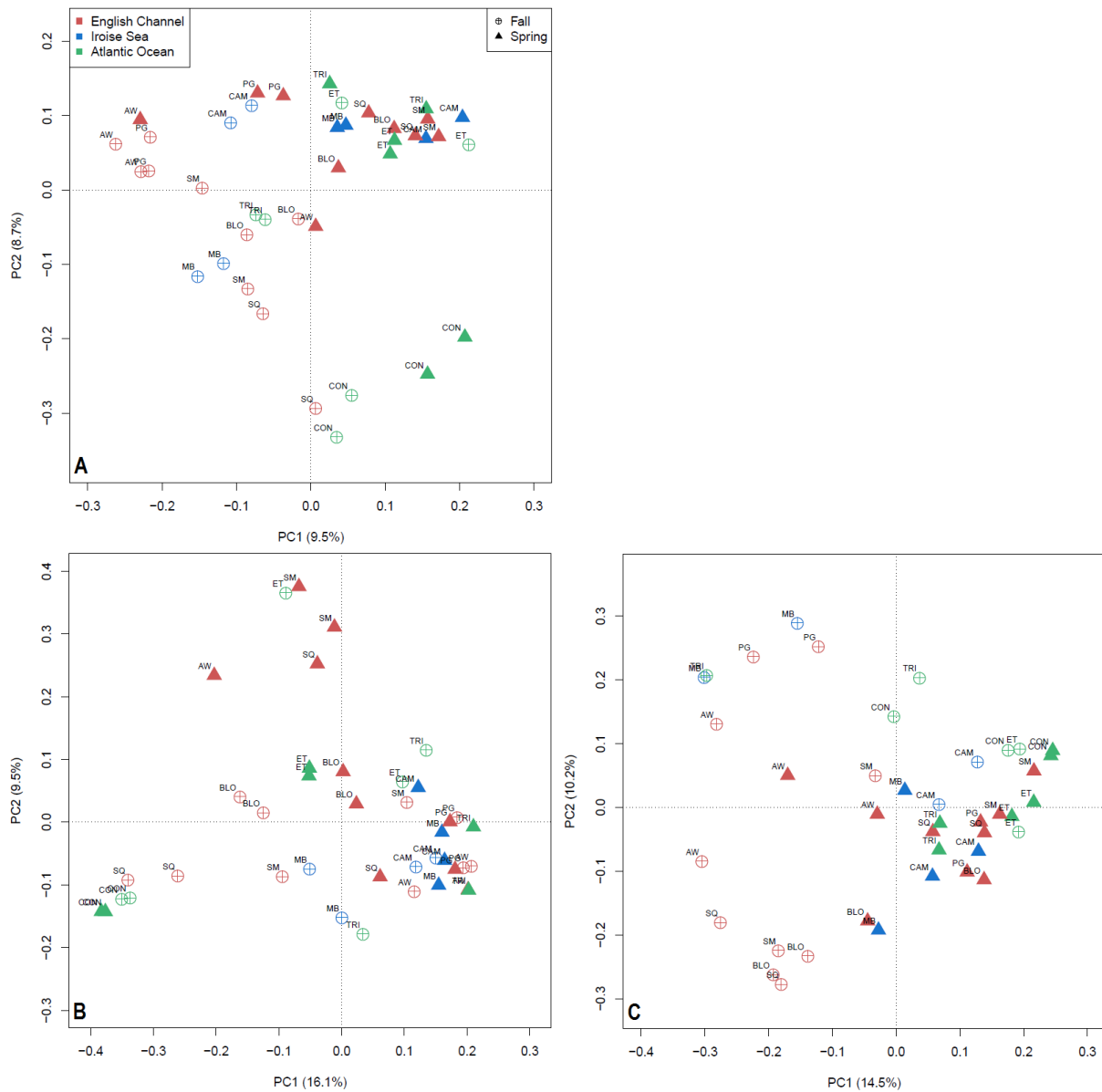
**Figure S2** Rarefaction curves on OTUs obtained with the COI marker for all samples. The red line indicates the number of reads chosen as a cut-off.



**Figure S3** Rarefaction curves on OTUs obtained with the 16S marker for all samples. The red line indicates the number of reads chosen as a cut-off.



**Figure S4** Ordination plots of principal component analysis results from Hellinger-transformed OTU abundances for pontoons from each locality according to their region (colors) and the season of sampling (shape) on four datasets: All OTUs obtained with 18S (A), OTUs from the Chromista subset (B), OTUs from the Pelagic subset (C), and OTUs from the Benthic subset (D). Sample scores are displayed in scaling 1. See figure 1 for location codes.



**Figure S5** Ordination plots of principal component analysis results from Hellinger-transformed OTU abundances for pontoons from each locality according to their region (colors) and the season of sampling (shape) on three datasets: All OTUs obtained with 16S (A), OTUs from the Pelagic subset (B), and OTUs from the Benthic subset (C). Sample scores are displayed in scaling 1. See figure 1 for location codes.

## Appendix 4: Supplementary material from chapter III.1

### a. Supplementary tables

**Table S1** List of values applied to the different parameters during read processing and filtering for each of the three markers used in this study. All parameters not indicated in this table were used with the default value. W describe parameters specific to water eDNA datasets whereas P stands for plates bulkDNA. When the type of DNA is not indicated, the same parameters were used for both.

Tool	Command	Parameter	18S	COI	16S
<b>CUTADAPT</b>	cutadapt	-e	0.16	0.12	0.14
<b>DADA2</b>	filterAndTrim	truncLen	W: 270 and 180 P: 230 and 180	240 and 180	90 and 90
		truncQ	0	0	0
		rm.phix	False	False	False
	dada	pool	pseudo	pseudo	pseudo
	mergePairs	minOverlap	W: 20 P: 20	W: 60 P: 50	W: 30 P: 40
	seqtab	Length selection	W: 335:430 P: 340:390	303:323	110:125
<b>R</b>	-	Index-jump	W: 0.0182 P: 0.0031	W: 0.0114 P: 0.0059	W: 0.0175 P: 0.0068
		Replicates	2	2	2

**Table S2** List of taxa identified by inspecting settlement plates under a microscope. Introduction status is indicated for species.

Phylum	Class	Family	Taxon	Status
Chordata	Asciacea	Cionidae	<i>Ciona intestinalis</i>	native
Chordata	Asciacea	Asciidae	<i>Asciella aspersa</i>	native
Chordata	Asciacea	Asciidae	<i>Asciella scabra</i>	native
Chordata	Asciacea	Corellidae	<i>Corella eumyota</i>	NIS
Chordata	Asciacea	Asciidae	<i>Phallusia mammillata</i>	native
Chordata	Asciacea	Molgulidae	<i>Molgula</i> sp.	native
Chordata	Asciacea	Styelidae	<i>Asterocarpa humilis</i>	NIS
Chordata	Asciacea	Pyuridae	<i>Pyura tessellata</i>	native
Chordata	Asciacea	Clavelinidae	<i>Clavelina lepadiformis</i>	native
Chordata	Asciacea	Perophoridae	<i>Perophora</i> sp.	NA
Chordata	Asciacea	Styelidae	<i>Dendrodoa grossularia</i>	native
Chordata	Asciacea	Polyclinidae	<i>Morchellium argus</i>	native
Chordata	Asciacea	Polyclinidae	Other Polyclinidae	NA
Chordata	Asciacea	Didemnidae	<i>Diplosoma</i> spp.	cryptogenic
Chordata	Asciacea	Didemnidae	Other Didemnidae	NA
Chordata	Asciacea	Styelidae	<i>Botryllus schlosseri</i>	cryptogenic
Chordata	Asciacea	Styelidae	<i>Botrylloides</i> spp.	NA
Bryozoa	Gymnolaemata	Candidae	<i>Tricellaria inopinata</i>	NIS
Bryozoa	Gymnolaemata	Bugulidae	<i>Bugula neritina</i>	NIS
Bryozoa	Gymnolaemata	Bugulidae	<i>Bugulina fulva</i>	cryptogenic
Bryozoa	Gymnolaemata	Bugulidae	<i>Bugulina flabellata</i>	native
Bryozoa	Gymnolaemata	Bugulidae	<i>Crisularia plumosa</i>	native
Bryozoa	Gymnolaemata	Bugulidae	<i>Bugulina stolonifera</i>	NIS
Bryozoa	Gymnolaemata	Bugulidae	<i>Bugulina</i> spp.	NA
Bryozoa	Stenolaemata	Crisiidae	<i>Crisia eburnea</i>	native
Bryozoa	Stenolaemata	Crisiidae	<i>Filicrisia geniculata</i>	native
Bryozoa	Gymnolaemata	Bugulidae	<i>Bicellariella ciliata</i>	native
Bryozoa	Gymnolaemata	Candidae	<i>Caberea boryi</i>	native
Bryozoa	Gymnolaemata	Watersiporidae	<i>Watersipora subatra</i>	NIS
Bryozoa	Gymnolaemata	Cryptosulidae	<i>Cryptosula pallasiana</i>	native
Bryozoa	Gymnolaemata	Electridae	<i>Electra pilosa</i>	native
Bryozoa	Gymnolaemata	Candidae	<i>Scrupocellaria</i> spp.	native
Bryozoa	Gymnolaemata	Candidae	<i>Cradoscrupocellaria reptans</i>	native
Bryozoa	Gymnolaemata	Calloporidae	<i>Callopora</i> spp.	native
Bryozoa	Gymnolaemata	Escharinidae	<i>Phaeostachys spinifera</i>	native
Bryozoa	Gymnolaemata	Hippothoidae	<i>Celleporella hyalina</i>	native
Bryozoa	Gymnolaemata	Celleporidae	<i>Celleporina</i> sp.	NA
Bryozoa	Gymnolaemata	Haplopomidae	<i>Haplopoma impressum</i>	NIS
Bryozoa	Gymnolaemata	Cribrilinidae	<i>Membraniporella nitida</i>	native
Bryozoa	Gymnolaemata	Microporellidae	<i>Microporella ciliata</i>	native
Bryozoa	Gymnolaemata	Bitectiporidae	<i>Schizomavella</i> spp.	native
Bryozoa	Gymnolaemata	Celleporidae	<i>Omalosecosa ramulosa</i>	native
Bryozoa	Gymnolaemata	Escharellidae	<i>Escharella</i> sp.	NA
Bryozoa	Gymnolaemata	Exochellidae	<i>Escharoides coccinea</i>	native
Bryozoa	Gymnolaemata	Chorizoporidae	<i>Chorizopora brongniartii</i>	cryptogenic
				native

Bryozoa	Gymnolaemata	Alcyonidiidae	<i>Alcyonidium cellarioides</i>	native
Bryozoa	Gymnolaemata	Scrupariidae	<i>Scruparia</i> spp.	NA
Bryozoa	Gymnolaemata	Aeteidae	<i>Aetea sica</i>	cryptogenic
Bryozoa	Stenolaemata	Tubuliporidae	<i>Tubulipora</i> spp.	NA
Bryozoa	Stenolaemata	Lichenoporidae	<i>Disporella hispida</i>	native
Bryozoa	Stenolaemata	Lichenoporidae	<i>Patinella radiata</i>	cryptogenic
Bryozoa	Stenolaemata	Plagioeciidae	<i>Plagioecia patina</i>	cryptogenic
Arthropoda	Hexanauplia	Austrobalanidae	<i>Austrominius modestus</i>	NIS
Arthropoda	Hexanauplia	Verrucidae	<i>Verruca stroemia</i>	native
Arthropoda	Hexanauplia	Balanidae	<i>Perforatus perforatus</i>	native
Arthropoda	Hexanauplia	Balanidae	Other Balanidae	NA
Mollusca	Bivalvia	Anomiidae	Anomiidae	NA
Mollusca	Bivalvia	Pectinidae	Pectinidae	NA
Annelida	Polychaeta	Spirorbidae	Spirorbinae	NA
Annelida	Polychaeta	Serpulidae	Serpulidae	NA
Annelida	Polychaeta	NA	Other Polychaeta	NA
Cnidaria	Anthozoa	NA	Anthozoa	NA
Cnidaria	Hydrozoa	Tubulariidae	<i>Ectopleura</i> sp.	NA
Cnidaria	Hydrozoa	Halopterididae	<i>Antennella</i> sp.	NA
Cnidaria	Hydrozoa	Plumulariidae	<i>Plumularia setacea</i>	cryptogenic
Cnidaria	Hydrozoa	NA	Other Hydrozoa	NA
Porifera	Calcarea	NA	Calcarea	native
Porifera	NA	NA	Other Porifera	NA
Arthropoda	Malacostraca	NA	Amphipoda	NA

**Table S3** Chi-squared values for pairwise tests of NIS/cryptogenic species proportions between localities. The values below the diagonal were calculated based on Morphology results whereas the values above the diagonal are based on data from the plates bulkDNA dataset. Corresponding P values are given in Table S3.

	BLO	BBL	AST	MEL	FIG	BdF
BLO	-	0.000	0.845	3.818	0.042	0.797
BBL	1.097	-	0.585	3.273	0.001	0.525
AST	1.559	0.000	-	0.642	0.642	0.000
MEL	1.927	0.000	0.000	-	2.788	0.996
FIG	0.538	0.032	0.115	0.297	-	0.280
BdF	0.320	0.055	0.152	0.341	0.000	-

**Table S4** P values associated with chi-square pairwise tests of NIS/cryptogenic species proportions between localities. The values below the diagonal were calculated based on Morphology results whereas the values above the diagonal are based on data from the plates bulkDNA dataset.

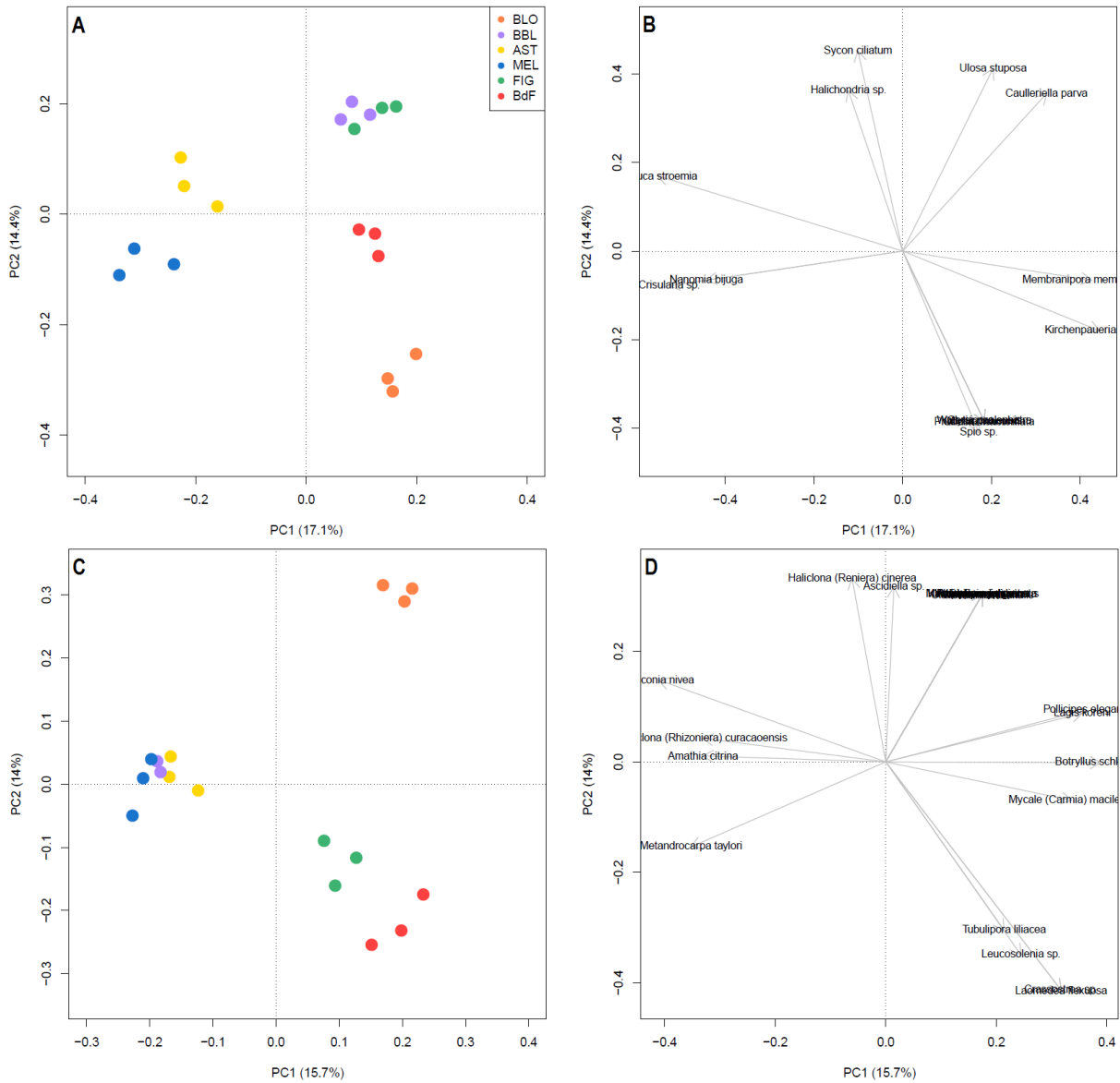
	<b>BLO</b>	<b>BBL</b>	<b>AST</b>	<b>MEL</b>	<b>FIG</b>	<b>BdF</b>
<b>BLO</b>	-	1.000	0.358	0.051	0.839	0.372
<b>BBL</b>	0.295	-	0.445	0.070	0.973	0.469
<b>AST</b>	0.212	1.000	-	0.423	0.423	1.000
<b>MEL</b>	0.165	1.000	1.000	-	0.095	0.318
<b>FIG</b>	0.463	0.858	0.734	0.586	-	0.597
<b>BdF</b>	0.572	0.815	0.697	0.559	1.000	-

**Table S5** Chi-square values (above diagonal) and associated P values (below diagonal) for pairwise tests of NIS/cryptogenic species proportions between localities for the water eDNA dataset.

	<b>BLO</b>	<b>BBL</b>	<b>AST</b>	<b>MEL</b>	<b>FIG</b>	<b>BdF</b>
<b>BLO</b>	-	0.000	0.211	1.852	0.080	0.000
<b>BBL</b>	1.000	-	0.186	1.753	0.077	0.000
<b>AST</b>	0.646	0.666	-	0.365	0.889	0.263
<b>MEL</b>	0.174	0.186	0.546	-	3.339	1.953
<b>FIG</b>	0.777	0.781	0.346	0.068	-	0.031
<b>BdF</b>	1.000	1.000	0.608	0.162	0.859	-



## b. Supplementary figures



**Figure S1** Ordination plot of principal component analyses results from Hellinger-transformed taxa occurrences observed in water eDNA samples collected at the time of positioning of the settlement structures. The two period of sampling have been treated separately (April: A-B; August: C-D). Sample scores are displayed in scaling 1. Colours indicate the different sampled localities.

## Appendix 5: Supplementary material from chapter III.2

---

### a. Supplementary tables

**Table S1** Details concerning the 42 target non-indigenous species. The total number of reference sequences, retrieved from public databases or produced locally (number in parenthesis) is indicated for each marker. The identification of the species following our metabarcoding analysis is indicated in the columns "detected", for each of the two markers used. Whenever possible, individual DNA was tested for amplification bias for COI. The results are reported in the last column with "yes" meaning that no (or weak) amplification was visible for the COI marker, and "no" meaning that amplification was visible (see figure S3). No bias was observed for the 18S marker.

Class	Species	Dispersal mode <sup>2</sup>	Present in the study bay	Native range <sup>3</sup>	18S	COI	Detected with 18S	Detected with COI	Amplification with COI
Asciacea	<i>Asterocarpa humilis</i>	short disperser	yes	S Pacific	2 (1)	4 (2)	yes	no	no
	<i>Botrylloides diegensis</i>	short disperser	yes	NE Pacific	1 (1)	2 (2)	no <sup>4</sup>	yes	yes
	<i>Botrylloides violaceus</i>	short disperser	yes	NW Pacific	1 (1)	5 (1)	no <sup>4</sup>	yes	yes
	<i>Botrylloides</i> sp X <sup>5</sup>	short disperser	no	Unknown	1 (1)	1 (1)	no	no	yes
	<i>Ciona robusta</i>	short disperser	yes <sup>6</sup>	NW Pacific	1 (1)	55 (55)	no	no	no
	<i>Corella eumyota</i>	short disperser	yes	cryptogenic (native from the southern hemisphere)	1 (1)	11 (11)	yes	no	no
	<i>Didemnum vexillum</i>	short disperser	yes	cryptogenic (introduced in Europe)	1 (-)	21 (8)	no	no	yes
	<i>Molgula manhattensis</i>	short disperser	no	NW Atlantic	1 (-)	19 (17)	no	no	not tested
	<i>Perophora japonica</i>	short disperser	yes	NW Pacific	1 (1)	1 (1)	no	no	no <sup>7</sup>
	<i>Styela clava</i>	short disperser	yes	NW Pacific	1 (1)	27 (1)	no	no	weak
Bivalvia	<i>Anadara kagoshimensis</i>	long disperser	no	NW Pacific	2 (-)	64 (-)	no	no	not tested
	<i>Arcuatula senhousia</i>	long disperser	no	NW Pacific	1 (-)	59 (-)	no	no	not tested

<sup>a</sup> Short and long disperser describe species with a benthic-pelagic life cycle for which the larvae spend either less or more than 2 days in the water column, respectively, based on literature data.

<sup>3</sup> Cryptogenic refers to species for which the native range is unknown

<sup>4</sup> No unique variants assigned to this species but some were assigned to the genus

<sup>5</sup> *Botrylloides* sp X is a cryptic species recently discovered within the genus *Botrylloides* (Viard *et al.* in prep.; Wood *et al.* (2015)). Reference: Wood C, Bishop J, Yunnice A (2015) Comprehensive reassessment of NNS in Welsh marinas, online report available at: <http://plymsea.ac.uk/id/eprint/7138/1/Comprehensive%20Reassessment%20of%20NNS%20in%20Welsh%20marinas.pdf>

<sup>6</sup> This species has been reported in 2012 but disappeared after that (Bouchemousse *et al.* 2017; Authors, personal observation). Reference: Bouchemousse S, Lévêque L and Viard F (2017) Do settlement dynamics influence competitive interactions between an alien tunicate and its native congener? *Ecology and Evolution* 7: 200-213.

<sup>7</sup> The lack of PCR amplification for *P. japonica* is not shown in Figure S3 but has been observed in previous experiments (data not shown).

	<i>Corbicula fluminea</i>	short disperser	no	NW Pacific	4 (-)	15 (-)	no	no	not tested
	<i>Ensis leei</i>	long disperser	no	NW Atlantic	3 (-)	9 (-)	no <sup>4</sup>	no	not tested
	<i>Crassostrea gigas</i> <sup>8</sup>	long disperser	yes / Aquaculture	NW Pacific	4 (1)	70 (1)	yes	no	yes
	<i>Mercenaria mercenaria</i>	long disperser	no	NW Atlantic	5 (-)	8 (-)	yes	no	yes
	<i>Mizuhopecten yessoensis</i>	long disperser	no	NW Pacific	1 (-)	7 (-)	no	no	not tested
	<i>Mya arenaria</i>	long disperser	yes	NW Atlantic	2 (1)	12 (-)	yes	yes	yes
	<i>Mytilopsis leucophaeata</i>	long disperser	no	NW Atlantic & Gulf of Mexico	4 (3)	2 (1)	no	no	weak
	<i>Petricolaria pholadiformis</i>	long disperser	no	NW Atlantic	1 (-)	1 (-)	no	no	not tested
	<i>Rangia cuneata</i>	long disperser	no	Gulf of Mexico	1 (1)	5 (3)	no	no	yes
	<i>Ruditapes philippinarum</i>	long disperser	no <sup>9</sup>	NW Pacific	2 (1)	91(-)	yes	yes	yes
	<i>Xenostrobus securis</i>	long disperser	no	South Pacific	2 (-)	23 (-)	no	no	not tested
Gastropoda	<i>Corambe obscura</i>	long disperser	no	NW Atlantic and Gulf of Mexico	1 (-)	2 (-)	no	no	not tested
	<i>Crepidula fornicata</i>	long disperser	yes	NW Atlantic	2 (-)	68 (50)	yes	yes	yes
	<i>Crepidatella dilatata</i>	direct developer	no	SE Pacific	2 (2)	35 (3)	yes	no	no
	<i>Gracilipurpura rostrata</i>	direct developer	yes	Mediterranean Sea	1 (1)	1 (1)	no	no	weak
	<i>Haminoea japonica</i>	short disperser	no	Indo NW Pacific	0 <sup>10</sup>	36 (-)	no	no	not tested
	<i>Hexaplex trunculus</i>	direct developer	no	Mediterranean Sea & Macaronesian islands	0	37 (-)	no	no	not tested
	<i>Ocenebrellus inornatus</i>	direct developer	no	NW Pacific	0	8 (-)	no	no	not tested
	<i>Rapana venosa</i>	long disperser	no	NW Pacific	1 (-)	45 (-)	no	no	not tested

<sup>8</sup> Note that two species names are currently accepted according to the World Register of Marine Species, *Crassostrea gigas* and *Magallana gigas*

<sup>9</sup> This species has not been officially reported in the bay of Morlaix but some individuals have been collected by authors and other researchers.

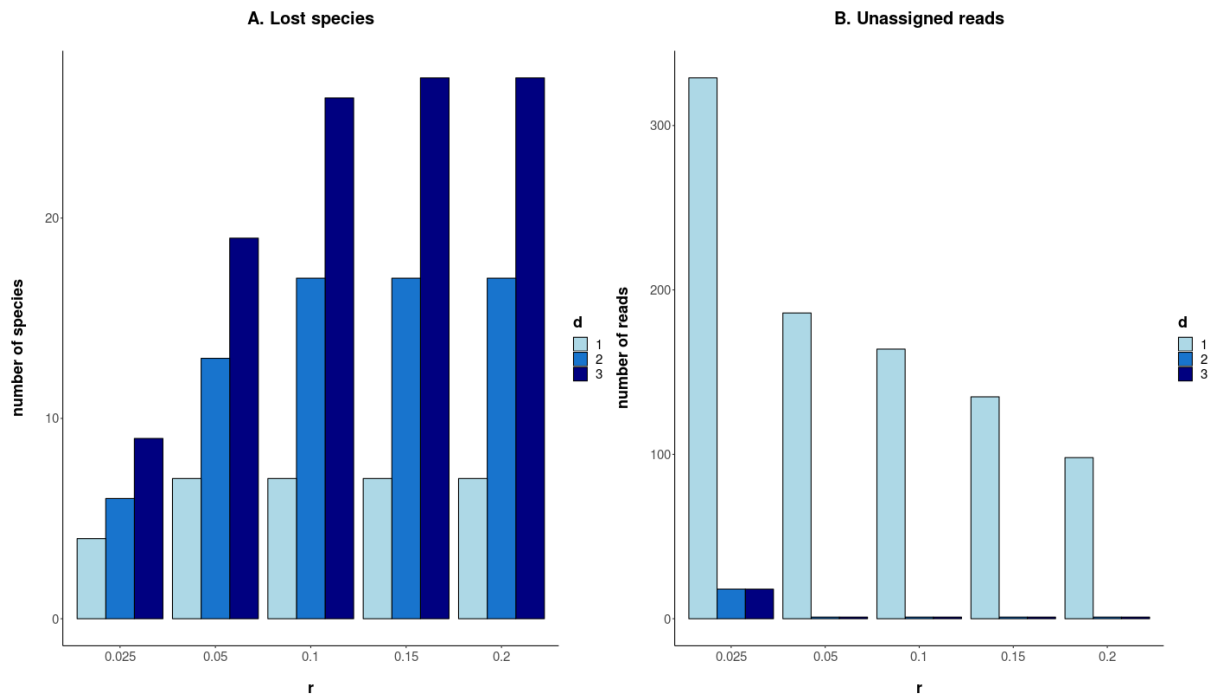
<sup>10</sup> No 18S sequence available for *H. japonica* but one attributed to *Haminoea* sp. was used in this study.

	<i>Tritia neritea</i>	direct developer	yes	Mediterranean and Black Seas	0	18 (-)	no <sup>4</sup>	no	not tested
	<i>Tritia pellucida</i>	- <sup>11</sup>	no	Mediterranean Sea	0	1 (-)	no <sup>4</sup>	no	not tested
	<i>Urosalpinx cinerea</i>	direct developer	no	NW Atlantic	0	11 (-)	no	no	not tested
Gymnolaemata	<i>Bugula neritina</i>	short disperser	yes	South Pacific	1 (1)	11 (1)	yes	yes	yes
	<i>Bugulina fulva</i>	short disperser	yes	NW Atlantic	1 (-)	2 (1)	no	no	yes
	<i>Bugulina simplex</i>	short disperser	no	cryptogenic (Presumably from the Mediterranean Sea)	1 (1)	2 (1)	no	no	yes
	<i>Bugulina stolonifera</i>	short disperser	yes	NW Pacific	1 (-)	2 (1)	no	no	yes
	<i>Celleporaria brunnea</i>	short disperser	no	NE Pacific	1 (1)	1 (1)	no	no	yes
	<i>Schizoporella japonica</i>	short disperser	no	NW Pacific	1 (1)	1 (1)	no	no	yes
	<i>Tricellaria inopinata</i>	short disperser	yes	cryptogenic (Presumably NE Pacific)	0	1 (-)	no	no	not tested
	<i>Watersipora subatra</i> <sup>12</sup>	short disperser	yes	cryptogenic	1 (1)	3 (3)	yes	no	yes

<sup>11</sup> Information not available for this species

<sup>12</sup> A recent revision of the genus by Vieira et al. (2014) revealed that the bryozoan *W. subtorquata* previously reported as an introduced species in Europe was actually *W. subatra*.  
Référence: Viera LM, Spencer Jones M, Taylor PD (2014) The identity of the invasive fouling bryozoan *Watersipora subtorquata* (d'Orbigny) and some other congeneric species. *Zootaxa* 3857(2): 151-182

## b. Supplementary figures

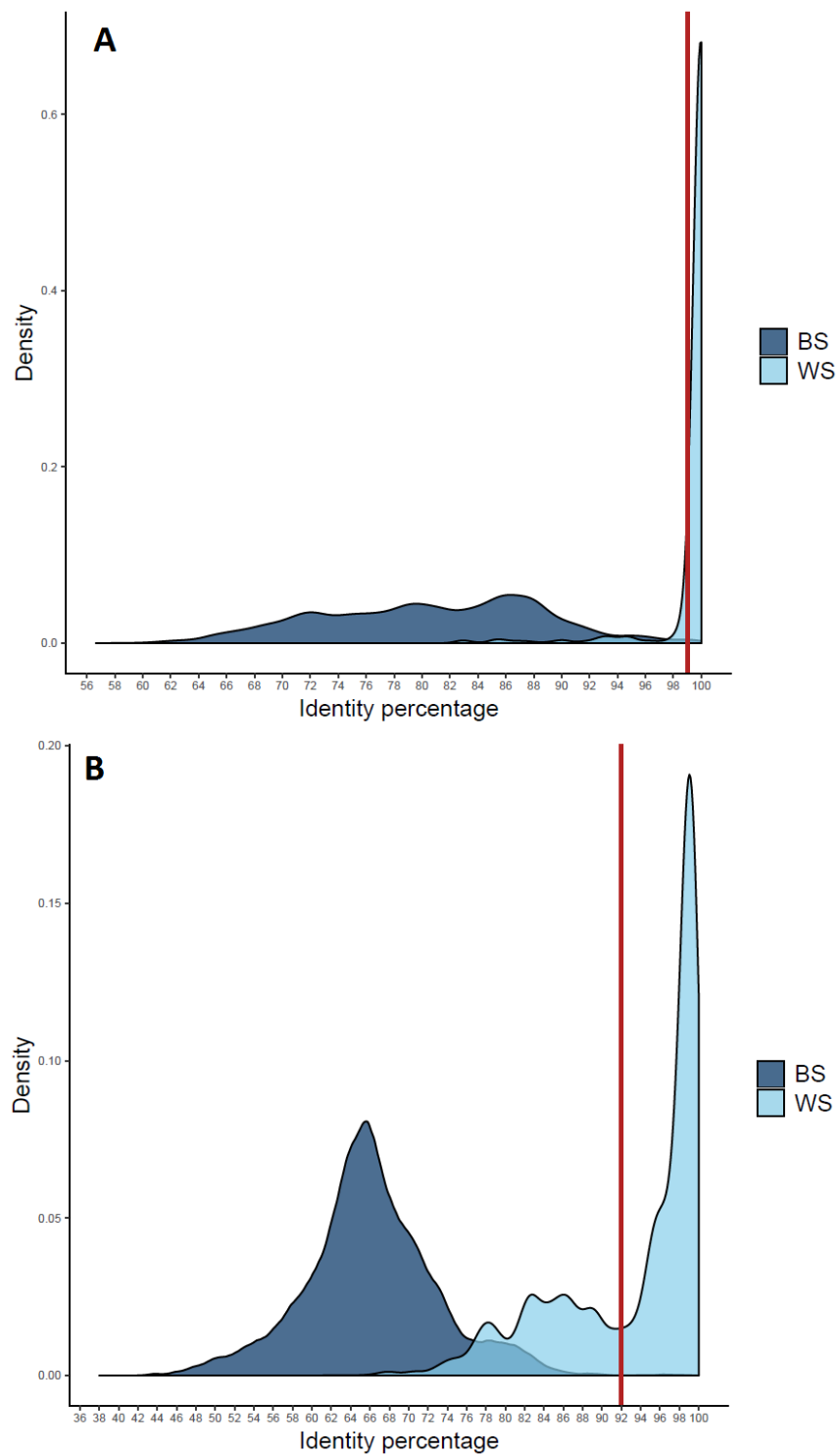


**Figure S1** Use of an *in-silico* mock community to determine the values for  $-d$  and  $-r$  parameters used in the *obiclean* tool

In order to evaluate the most appropriate values to use for two parameters of the *obiclean* tool available in the OBITools suite v.1.2.11, an *in-silico* mock community was created. A set of 294 sequences for the 18S marker, representing 254 species across 41 families within our four classes of interest were gathered from the SILVA public database. Some of these sequences were multiplied to mimic the variations in abundance which could be observed in a real dataset (between 1 and 130 sequences for each species). Sequencing was simulated using ART (Huang et al. 2012) and the quality profile of our real dataset was applied to the artificial one. The pipeline described above was applied to the 15,480 produced reads, and the *obiclean* tool was used several times with different values of  $-d$  (number of differences allowed for a read to be considered as an error produced from a variant) and  $-r$  (ratio between the abundance of two variants below which the less abundant one will be discarded). All variants passing the filtering process were then compared to the 294 starting sequences using BLAST<sup>®</sup>, and assigned to a species name when matching with more

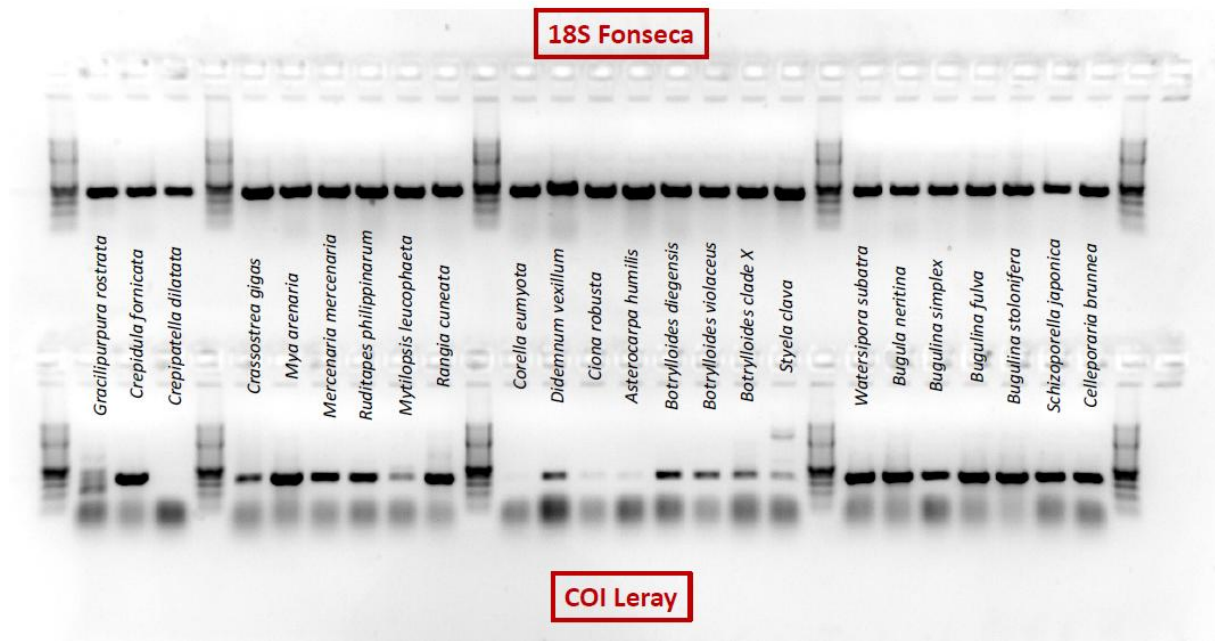
than 99% query cover and 100% identity. The number of species not retrieved (representing reads wrongly considered as errors) is shown in A, and the number of unassigned variants (representing undetected errors) is shown in B for the three values of  $-d$  and the five values of  $-r$  tested.

**Reference:** Huang W, Li L, Myers JR, Marth GT (2012) ART: a next-generation sequencing read simulator. *Bioinformatics* 28(4): 593-594



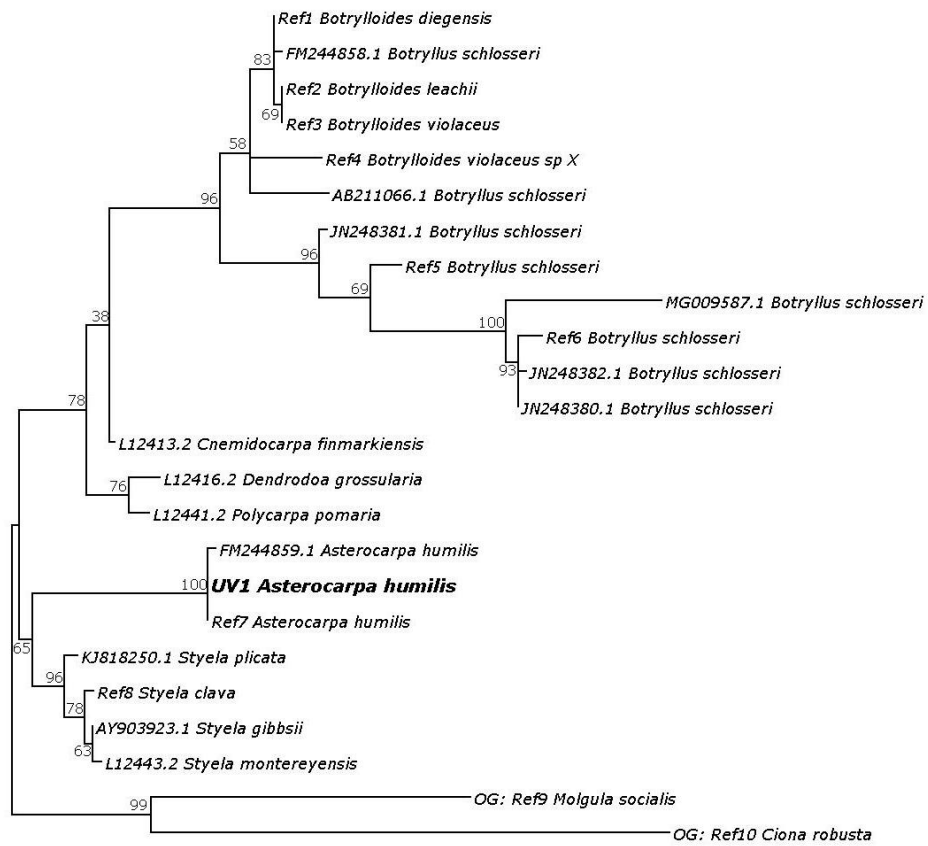
**Figure S2** Frequency distribution of pairwise sequence identity (%) between (BS) and within (WS) species using data from our custom-designed reference database for 18S (A) and COI (B) markers. The thresholds chosen for taxonomic assignment (i.e. 99% for 18S and 92% for COI) are indicated by red lines.



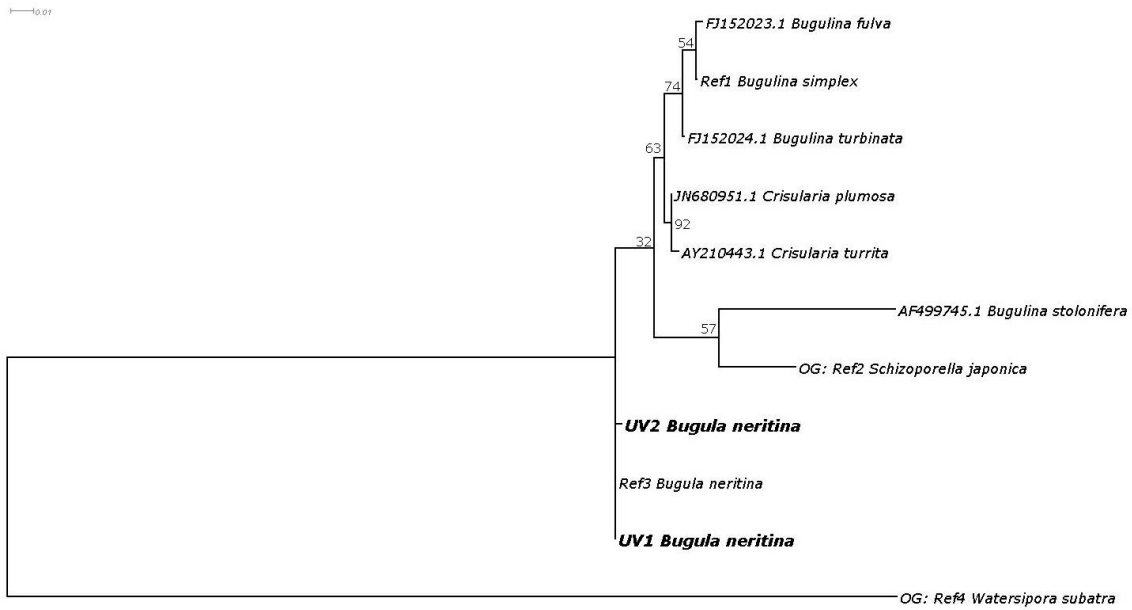


**Figure S3** Taxonomic amplification efficiency for 18S and COI. The picture displays some of the amplification results obtained using DNA extracted from 24 species to test for failure vs. success of amplification.

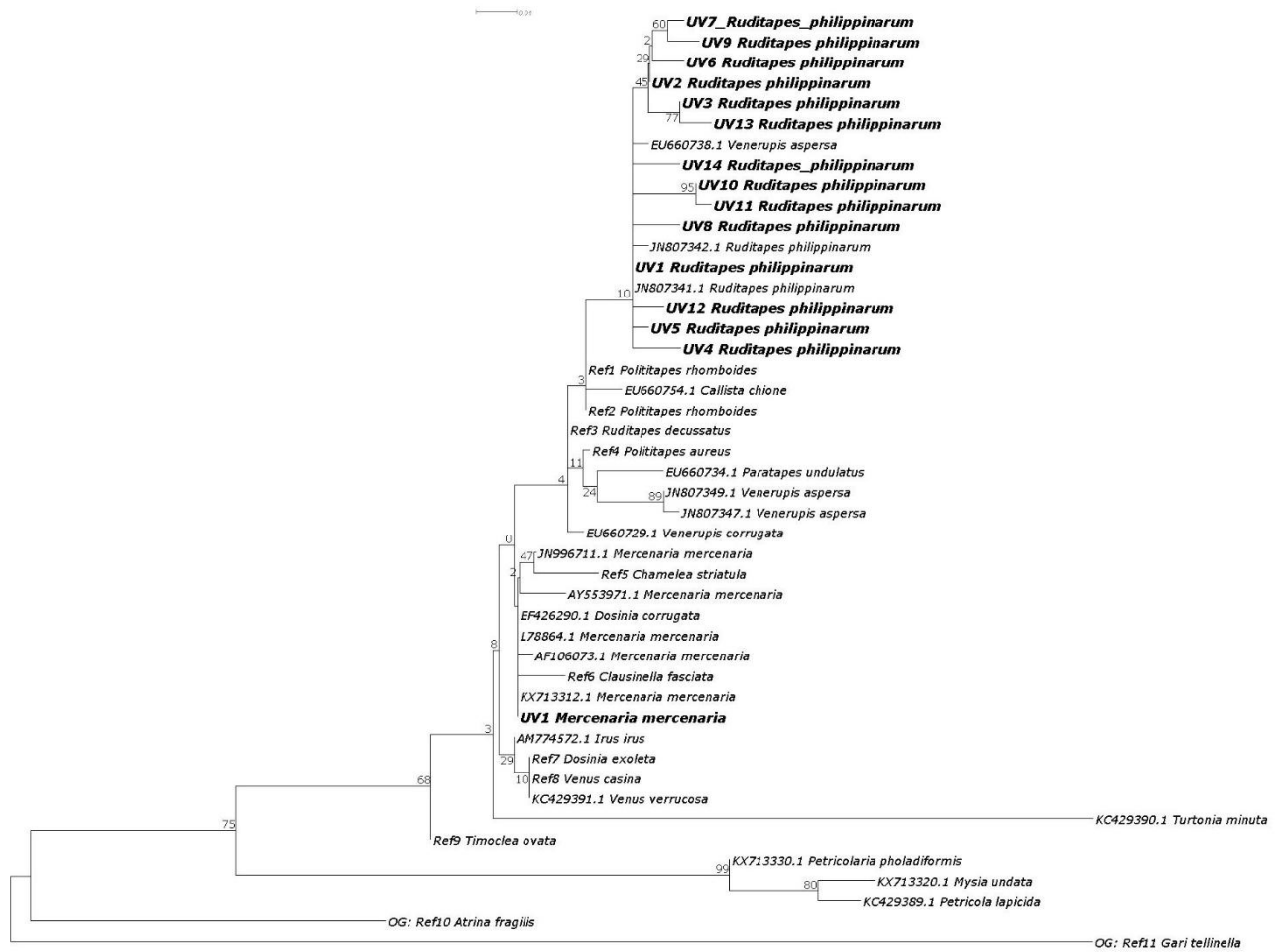
10.01



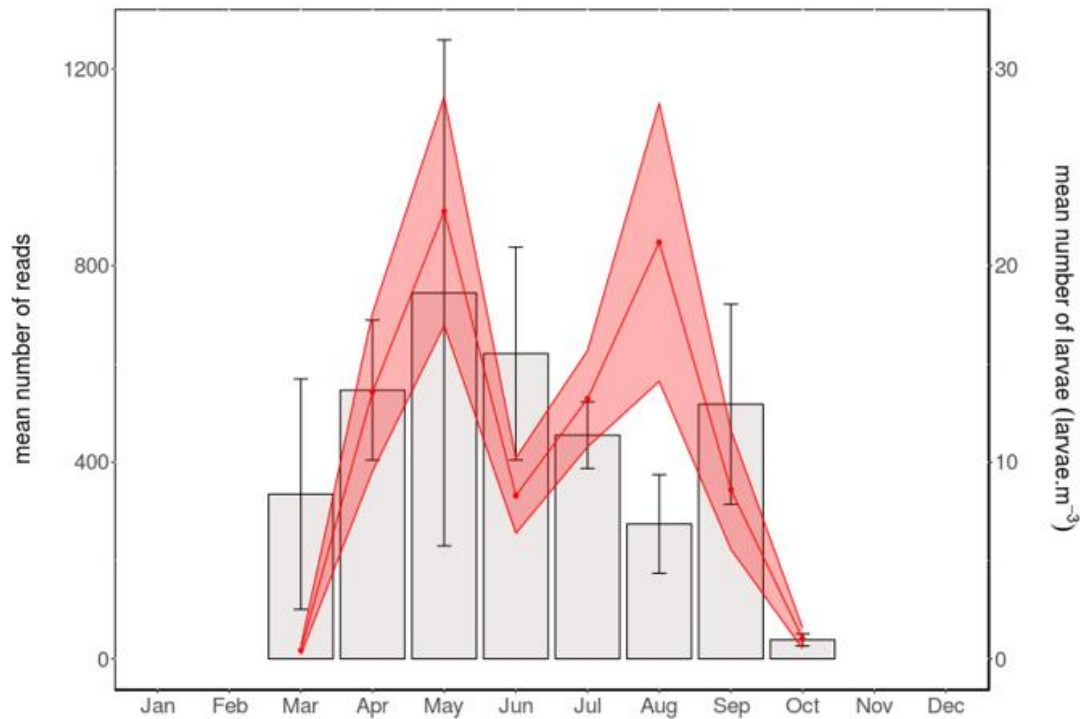
**Figure S4** Phylogenetic tree of the Styelidae family inferred from 18S sequences by using the maximum likelihood method based on the Kimura-2-parameter model (log likelihood = -1563.8858). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories). Bootstrap values (percentage over 1000 permutations) are shown for each visible node. This tree was computed using public references available for this family as well as all locally produced references ("Ref"), unique variants assigned to a Styelidae species ("UV", bold), and two outgroups ("OG").



**Figure S5** Phylogenetic tree of the Bugulidae family inferred from 18S sequences by using the maximum likelihood method based on the Kimura-2-parameter model (log likelihood = -1160.7132). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories). Bootstrap values (percentage over 1000 permutations) are shown for each visible node. This tree was computed using public references available for this family as well as locally produced references ("Ref"), unique variants assigned to a Bugulidae species ("UV", bold), and two outgroups ("OG").



**Figure S6** Phylogenetic tree of the Veneridae family inferred from 18S sequences by using the maximum likelihood method based on the Kimura-2-parameter model (log likelihood = -1618.0341). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories). Bootstrap values (over 1000 permutations) are shown for each visible node. This tree was computed using public references available for this family as well as locally produced references ("Ref"), unique variants assigned to a Veneridae species ("UV", bold), and two outgroups ("OG").



**Figure S7** Monthly variations in abundance of slipper limpet (*Crepidula fornicata*) larvae based on morphological identification (larvae m<sup>-3</sup>), averaged across seven years (red, right axis), and monthly variations in abundance of reads assigned to *C. fornicata* based metabarcoding data using the COI marker, averaged across replicates and samples of both years (grey bars, left axis). Each point/bar represent the mean number of observations at a given month. Black error bars and the red area represent the standard error of mean.

## Abstract

Human movements throughout the globe are constantly increasing, displacing concurrently numerous species. When arriving in a new environment, non-indigenous species (NIS) can establish and spread, with a series of economic and ecological consequences. Prevention and early detection being the most effective management strategies, it is of the utmost importance to be able to detect NIS in introduction hotspots, such as marinas, and when they are still in low abundance. Traditional methods for NIS survey require trained experts, are labor-costly and sometimes ineffective for specimen identification (e.g. early stages, cryptic species). DNA barcoding is often used as a complement but requires to process specimen one by one. Reducing cost and time of samples processing, High-Throughput Sequencing (HTS) techniques had been increasingly used in the past five years for NIS monitoring and biological invasions studies. However, work is still needed to assess their complementary values and limitations. In this thesis, we focused on marine benthic sessile organisms, found in biofouling communities in which numerous NIS are present. In a first chapter, we investigated both taxonomic and genetic (i.e. infra-specific) diversity, two important facets of biodiversity, using HTS techniques applied to assemblages of colonies of the genus *Botrylloides*, a colonial ascidian, sampled in the wild. The six bioinformatic pipelines that were compared did not perform equally, although all of them provide reasonable assessment of species and genetic diversity. In a second chapter, we studied environmental DNA obtained in ten marinas around Brittany, and two seasons. We first compared results obtained with traditional survey methods (specimens sampled in quadrats and identified in situ) to those produced by HTS. We showed that, although HTS-based techniques provided information on a much larger taxonomic coverage and detected numerous ENI, they failed to identify important target NIS. False positives were also observed, which might be a severe limitation of the approach for NIS early detection. Further work on primers, markers and reference databases to be used can circumvent these issues. In a second part, we described marina's communities in the same marinas and showed that, despite potential transport via recreational boating, they exhibit different assemblages and a strong spatio-temporal structure. Finally, in a third chapter, we focused on NIS reproduction and dispersal abilities in the wild. In a first part, we assessed the potential spread of marine benthic sessile NIS from introduction hotspots, based on HTS and morphological data obtained from specimens settled on experimental settlement plates deployed in and outside one marina. While the assemblages within the marina were very different from natural surrounding habitats, most NIS observed in the marina were also present in natural habitats. Then, we evaluated the reproductive output (period/intensity) of several benthic NIS by applying a metabarcoding approach on zooplankton samples from a time-series over two years. Metabarcoding was efficient in detecting NIS whatever their life-cycle, and in inferring reproductive period for species with long-lived larvae. Altogether this work showed that HTS-based and metabarcoding are effective methods for NIS surveys and studies, if precautions can be taken to circumvent the reported uncertainties of the method.

## Résumé

La circulation des hommes et de leurs marchandises à travers le monde est en constante augmentation, déplaçant simultanément de nombreuses espèces. Les espèces non indigènes (ENI) ainsi introduites peuvent s'établir et se propager, avec une série de conséquences socio-économiques et écologiques. La prévention et la détection précoce étant les stratégies de gestion les plus efficaces, il est crucial de pouvoir détecter les ENI dans les points chauds d'introduction, tels que les ports, et lorsqu'elles sont encore en faible abondance. Les méthodes traditionnelles d'identification, basées sur la morphologie, des ENI nécessitent de l'expertise, sont coûteuses et parfois inefficaces (cas par exemple des stades larvaires et des espèces cryptiques). Le barcoding ADN est alors souvent utilisé mais il nécessite aussi de traiter les spécimens un par un. Réduisant le coût et le temps de traitement, les techniques de séquençage à haut débit (HTS) sont de plus en plus utilisées. Toutefois, des recherches sont encore nécessaires pour évaluer leurs limites et valeur ajoutée. Dans cette thèse, nous nous sommes intéressés aux organismes benthiques marins sessiles trouvés dans le biofouling, dans lequel se trouvent de nombreuses ENI. Dans un premier chapitre, nous avons étudié la diversité taxonomique et génétique (infraspécifique), deux importantes facettes de la biodiversité, en appliquant des techniques HTS à des assemblages de colonies d'ascidies du genre *Botrylloides*, échantillonnés sur le terrain. Les six pipelines bioinformatiques testés n'ont pas les mêmes performances, bien qu'ils fournissent tous une évaluation raisonnable de la diversité taxinomique et génétique. Dans un deuxième chapitre, nous avons étudié l'ADN environnemental obtenu dans dix ports de plaisance de Bretagne, et ce en deux saisons, et comparé les résultats à ceux issus d'une méthode traditionnelle (identification sur le terrain de spécimens prélevés dans un cadrat). Bien que les données obtenues par HTS fournissent des informations sur une couverture taxinomique large et détectent de nombreuses ENI, elles ne permettent pas d'identifier toutes les ENI observées sur le terrain. Des faux positifs ont également été observés, ce qui est une limite potentiellement importante pour la détection précoce de nouvelles ENI. Ces limites pourraient être levées avec un travail plus approfondi sur les amorces, marqueurs et bases de données de référence. Dans une deuxième partie, nous avons décrit les communautés établies dans ces mêmes ports de plaisance. Malgré le transport potentiel d'espèces par la navigation de plaisance, elles sont variables dans le temps et l'espace. Enfin, dans un troisième chapitre, nous nous sommes concentrés sur les capacités de reproduction et de dispersion des ENI dans les habitats naturels. Nous avons d'abord évalué la propagation des ENI benthiques sessiles à partir d'un point chaud d'introduction (un port), en nous basant sur des données HTS et morphologiques, à partir de spécimens ayant colonisé des plaques expérimentales déployées au sein et hors du port. Si les assemblages au sein du port étaient différents des habitats naturels environnants, la plupart des ENI observées dans le port était également présent dans les habitats naturels. Ensuite, nous avons évalué la reproduction (période/intensité) de plusieurs ENI benthiques par metabarcoding de zooplancton provenant d'une série chronologique sur deux ans. Le metabarcoding est là encore efficace pour détecter les ENI, ayant différents cycle de vie, et il permet de suivre la reproduction des espèces à larves longévives. Dans l'ensemble, ces travaux ont montré que les approches basées sur le HTS et le metabarcodage sont efficaces pour les suivis et études des NEI, bien que des précautions doivent être prises pour contourner les incertitudes de la méthode.



