UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

Tell Me How You Look Like And I'll Tell You Where You Come From: Identification Of Bivalves Shells' Origin

Inês Rebolo Martins

Mestrado em Ecologia Marinha

Dissertação orientada por:
Doutora Paula Chainho
Doutora Marta Rufino

2022

Agradecimentos

Queria expressar o meu agradecimento às pessoas que mais me ajudaram na realização deste trabalho, quer de forma direta ou indireta, e sem as quais eu não estaria neste ponto agora.

Às minhas orientadoras, a Professora Paula Chainho e a Professora Marta Rufino, que me deram esta oportunidade de trabalhar sobre este tema desafiante e pelo apoio e disponibilidade durante todo este processo. Agradeço ainda pela grande disposição para ajudar e tirar as dúvidas que eu tinha (não eram poucas). Sem o vosso apoio nunca teria conseguido realizar um projeto tão intrincado como este.

Ao MARE e ao IPMA por acolherem-me com imensa simpatia e ajudarem-me durante este processo. Em particular ao projeto NIPOGES que disponibilizou a grande maioria das amostras sem as quais este trabalho não seria possível.

Ao Thomas Goulding por me ter ajudado imenso neste trabalho sobretudo na parte inicial quando mais precisava de rumo e apoio.

Ao grupinho de trabalho da tese que me motivou a trabalhar sem nunca desistir (e não foi pouca a vontade). Ao Pedro que me aturou todos os santos dias e vivenciou grande parte dos meus problemas, que me aconselhou e, que me ouviu a desabafar toda a minha frustração e sabe o quão difícil tudo isto foi. À Marta Salvado (e ao seu carrinho lindo, nunca me esquecerei das boleias) que, não só me motivou, mas, também, me acompanhou ao IPMA durante muito tempo, e que, também, sabe os cabelos brancos que este trabalho me criou. À Daniela Morgado e à Débora que trabalharam comigo em ocasiões diferentes e fizeram pressão para avançar no meu próprio trabalho.

A todos os meus amigos, quer da licenciatura/mestrado, quer da terrinha, que acabaram por saber das minhas aventuras e me apoiaram do melhor que puderam.

À minha família, que me incentivou e me ouviu reclamar de tudo e mais alguma coisa, que esperou mais do que devia pelo fim deste trabalho e, que espero que estejam felizes e contentes pelo que consegui realizar.

At last, to my mutuals who were always present and accompanied me through everything, through the highs and lows, who never got tired of me and my breakdowns, when even I was fed up, who shared with me indignation, laughs and, above all, gave me courage and motivation every day.

A todos, obrigada.

Resumo

A zona costeira é um foco importante de comércio, residência e lazer em todo o mundo. Um dos serviços mais relevantes fornecidos por esta zona é a produção de alimento e inclui os bivalves, que são espécies chave em habitats estuarinos e costeiros.

O berbigão (*Cerastoderma* spp.) é um bivalve de interesse comercial e ecológico que ocorre e é capturado em vários sistemas aquáticos costeiros da Europa. Em Portugal, encontra-se presente em vários sistemas aquáticos costeiros nomeadamente na Ria de Aveiro e na Ria Formosa. O berbigão é o bivalve mais capturado em todo o país contabilizando valores superiores a 66 mil toneladas no período de 2010 a 2020. Devido à importância deste bivalve na economia e pesca em Portugal e na Europa, é essencial, em muitas situações, averiguar a origem dos indivíduos por via da rastreabilidade dos espécimes, nomeadamente para o caso da verificação de pesca ilegal ou em caso de rótulos fraudulentos. Em Portugal, ocorrem duas espécies de berbigão que vivem em simpatria na maioria dos locais, *Cerastoderma edule* (Linnaeus 1768) e *Cerastoderma glaucum* (Bruguière 1799). Estes são difíceis de diferenciar devido às suas similitudes morfológicas. Apesar desse facto, diversos estudos sobre a identificação da origem geográfica ou da identidade destas espécies usam caraterísticas morfológicas.

Um dos métodos de análise das caraterísticas morfológicas mais reconhecido é a morfometria geométrica, *i.e.*, a representação visual de mudanças de forma e a produção de variáveis que podem ser analisadas estatisticamente. Uma das abordagens da morfometria geométrica é baseada em *landmarks i.e.*, pontos de referência de interesse biológico que são considerados como homólogos e identificáveis em todos os espécimes. As *landmarks*, depois de serem analisadas com a análise generalizada de Procrustes (GPA), podem produzir variáveis de forma que podem ser analisadas estatisticamente.

O objetivo deste trabalho foi estudar a variação geográfica da forma da concha de duas espécies de berbigão com interesse comercial (*Cerastoderma edule* e *Cerastoderma glaucum*) em cinco sistemas aquáticos costeiros portugueses (Ria de Aveiro, lagoa de Óbidos, estuário do Tejo, lagoa de Albufeira e estuário do Sado), explorando a possibilidade de inferir a origem dos indivíduos. Para tal, usou-se a morfologia da valva direita da concha de 504 indivíduos e vários métodos envolvendo morfometria geométrica, testando se esses métodos podem ajudar no desenvolvimento de técnicas de rastreabilidade da forma, associadas à origem de bivalves económica- e ecologicamente importantes como é o caso de *Cerastoderma* spp.

Os métodos de morfometria geométrica usados neste estudo têm baixo custo, uma vez que não necessitam do uso de reagentes ou de técnicos especializados em laboratório. Estes métodos foram baseados em 16 *landmarks*, localizadas nas valvas da concha direita em locais homólogos de interesse biológico (*e.g.*, cicatriz do músculo adutor anterior e posterior, linha paleal, etc.). Com base nessas *landmarks*, foram feitas diversas análises estatísticas. Cada uma das análises foi realizada utilizando o programa de computador R e R Studio e desvendou aspetos diferentes da forma de uma estrutura complexa, como é a concha dos bivalves, nos cinco sistemas aquáticos costeiros portugueses diferentes.

A análise de variância (ANOVA) permitiu descrever padrões de variação de forma para o conjunto de coordenadas Procrustes previamente analisado por via da GPA. A ANOVA aferiu a hipótese estatística de existir diferenças de forma da concha relativas ao tamanho do espécime, à espécie, ao sistema aquático e a interação entre a espécie e o sistema aquático.

A análise de componentes principais (PCA) permitiu a visualização de padrões de variação de forma das conchas relativamente às duas espécies (*Cerastoderma edule* e *Cerastoderma glaucum*) e aos

cinco sistemas aquáticos (Ria de Aveiro, lagoa de Óbidos, estuário do Tejo, lagoa de Albufeira e estuário do Sado). Após a análise dos padrões observados, foi feita a comparação detalhada da posição das *landmarks* de diferentes grupos com interesse, com a ajuda de vetores.

A análise de trajetória permitiu quantificar e comparar as variações de forma da concha observadas com as outras análises realizadas, expondo os sistemas aquáticos que apresentavam mais diferenças de forma entre os cinco que foram estudados. Isto foi possível através da análise de magnitudes, correlações e direções de trajetórias que representam atributos geométricos de variação de forma das diferentes espécies e dos diferentes sistemas aquáticos costeiros.

A análise de disparidade morfológica representou a diversidade de morfologias presente em cada um dos sistemas aquáticos costeiros e para cada uma das espécies em estudo, estando ligada à diversidade funcional e ecológica dos espécimes. Esta análise realizou-se com o uso de dois descritores, um baseado num *proxy* do tamanho da concha, o *centroid size*, e outro baseado no modelo ANOVA. Cada descritor capturou e realçou aspetos diferentes da morfometria da concha dos espécimes estudados, permitindo uma análise mais completa dos dados.

A análise de variáveis canónicas permitiu a classificação da forma das conchas em cada um dos cinco sistemas aquáticos estudados, sendo útil na identificação de locais onde as morfologias são semelhantes e, portanto, ajudando na rastreabilidade do berbigão. Com a matriz de confusão obtida a partir desta análise, foi possível calcular a proporção de erro, *i.e.*, as classificações erradas divididas pelo total observado, realçando as diferenças na morfologia das conchas de cada um dos cinco sistemas aquáticos costeiros portugueses estudados.

O *Random forest*, um método de *machine learning* popular pelo seu excelente desempenho e grande sucesso em vários contextos (*e.g.*, ecológico, taxonómico, etc.), que permitiu a previsão da diferenciação entre espécies, entre sistemas aquáticos, e entre ambos simultaneamente, com uma exatidão de alta percentagem. O *Random forest* efetua tarefas de previsão rapidamente e com grande facilidade, apresenta classificações sólidas e precisas baseadas num algoritmo de *machine learning* que usa múltiplas árvores de decisão individuais construídas a partir de um subconjunto de dados aleatórios, reservando o restante dos dados para validação cruzada interna. Com este método e as suas matrizes de confusão, foi calculado o erro, *i.e.*, previsões erradas divididas pelo total observado, em percentagem para cada um dos grupos em análise, de forma a ter informação mais detalhada acerca da variação de morfologias presente nas conchas recolhidas neste estudo.

O uso das diferentes análises da morfométrica geométrica baseada em *landmarks*, método que já provou a sua credibilidade por via de diversos estudos, permitiu, no presente trabalho, analisar a morfometria e a variação de forma das conchas de duas espécies de bivalves de interesse comercial, *Cerastoderma edule* e *Cerastoderma glaucum*. Foi feita a comparação das diferentes morfologias das conchas dos espécimes recolhidos nos cinco sistemas aquáticos costeiros estudados de forma a capturar diferenças subtis de forma nas conchas, permitindo, assim, a caraterização e discriminação das populações de berbigão estudadas. Por mérito dos métodos de morfometria geométrica baseada em *landmarks*, foi possível não só a diferenciação geográfica dos cinco sistemas aquáticos costeiros estudados, mas também, o estudo da morfometria da concha do berbigão em grande detalhe.

Os resultados obtidos no presente estudo permitem apoiar o uso da morfometria geométrica baseada em *landmarks* para análises com objetivos semelhantes aos que foram apresentados neste trabalho. Para além disso, os resultados deste estudo fornecem boas perspetivas para trabalhos futuros visando a diferenciar populações de berbigão presentes em zonas geográficas variadas. As metodologias usadas

neste trabalho podem ser alargadas a outras populações de bivalves e sistemas aquáticos. Por serem técnicas de discriminação precisas, eficazes, fáceis e de baixo custo, estas podem tornar-se uma contribuição benéfica para a rastreabilidade de populações de diversas espécies incluindo as estudadas aqui, e, também, podem representar um recurso valioso na pesca em geral.

Palavras-chave: *Cerastoderma* spp., morfometria geométrica baseada em *landmarks*, *Random forest*, rastreabilidade, Portugal

Abstract

Cerastoderma edule (Linnaeus 1768) and *Cerastoderma glaucum* (Bruguière 1799) are two sibling species of cockles that can occur in sympatry. They are ecologically and economically important, have a wide geographical distribution and are harvested all over Europe. In Portugal, cockles are commonly found along all coastal areas and were the bivalve species with the highest landings in the country between 2010 and 2020. This importance emphasises the need for unravelling the origin of an individual using traceability techniques, when the harvesting location is doubtful, identifying and verifying illegal fishing activity or benefiting consumers through the detection of mislabelled seafood products.

Distinguishing cockles' populations based on morphological characteristics is difficult due to their morphological similarities, however, species and geographic identification studies often use morphological characteristics as tools to acquire most of this bivalve's information. Geometric morphometrics is one of the most successful methods based on morphometry, providing a visual representation of shape change and producing shape variables that can be analysed through statistics. The aim of this work was to study the geographical variation of the shell shape of *Cerastoderma* spp. common in Portuguese aquatic systems such as Ria de Aveiro, the Tagus and Sado estuaries, the Albufeira and Óbidos lagoons, etc. The possibility of inferring the origin of an individual using shell morphology and multiple landmark-based geometric morphometrics methodologies was explored alongside the possibility of using such methods as traceability techniques.

This work supports the use of landmark-based geometric morphometrics methods, being easy, low cost and effective in the study of shell shape variation. Using these methods allowed the discrimination of cockle populations from different aquatic systems on the Portuguese coast. Such methodologies could be extended to other bivalve populations and aquatic systems, becoming beneficial for the traceability of other populations and a valuable asset in fisheries in general.

Keywords: *Cerastoderma* spp., landmark-based geometric morphometrics, Random forest, traceability, Portugal

Index

List of figures and tables

List of abbreviations

ANOVA – Analysis of variance

Csize – Centroid size

CVA – Canonical variates analysis

DA – Disparity analysis

DGRM – Direção-Geral de Recursos Naturais, Segurança e Serviços Marítimos

FAO – Food and Agriculture Organization of the United Nations

GM – Geometric morphometrics

GPA – Generalised Procrustes analysis

LM – Landmark

MD – Morphological disparity

NIPOGES project – Estado atual das populações de amêijoa-japonesa da Ria de Aveiro, lagoa de Óbidos e estuários do Tejo e Sado – bases científicas para uma gestão sustentável do recurso

PC – Principal component

PCA – Principal component analysis

PCC – Probability of correct classification

RF – Random forest

TA – Trajectory analysis

1. Introduction

The coastal zone is an important focus of trade, residence, and recreation all around the world [1]. One of the most prominent services provided by this area is food production, which includes bivalves [1], which are key species in estuarine and coastal habitats [2]. They are tolerant to harsh environments [3], living in different types of habitats like soft, muddy sediments or lotic habitats [4, 5], they also exhibit some variation in mode of life [6], and many of them are long-lived [7].

Cerastoderma edule (Linnaeus 1768) and *Cerastoderma glaucum* (Bruguière 1799) are two sibling species of cockles [8-11], that may co-exist and form mixed populations [12, 13], and, since they are wild-harvested, their value and ecosystem importance are often overlooked, when compared with other commercial bivalve species [1]. Despite this, cockles play an essential role in the ecosystem and, considering they are a sediment-dwelling active filter, they are an important link between primary producers and consumers such as crabs, shrimps, fishes, and birds [14-21].

Both the common cockle (*C. edule*) and the lagoon cockle (*C. glaucum*) are shallow burrowing bivalves, buried in the surface of the sediment in sandy mud, sand, or fine gravel bottoms, or in the intertidal zone, until a few meters, and shallow subtidal areas [18, 22-28]. They can occur in dense populations of several thousand individuals per square metre, comprising up to 60% of the biomass of an area [29-31] and have been characterised by having considerable temporal and spatial variability [18, 29, 32-34]. These two distinct species belonging to the genus *Cerastoderma* exhibit considerable morphological similarities and a high phenotypic variation [11, 12]. Their shell is a strong shell with marked ribs that evolved into and thicker spheroidal shape [12, 28, 35], mostly composed of aragonite [36]. This substantially reduced *Cerastoderma*'s capacity to burrow but provided the ability to withstand severe changes in temperature and desiccation during aerial exposure at low tides [12, 28, 37, 38]. They can also tolerate wide ranges of salinity from 5 to 45 for *C. glaucum* and 3 to 35 for *C. edule* [39-42]. They have high resistance to extreme and unstable environmental conditions [43-45] and can live up to 10 years in some habitats, more commonly, up to six years or down to two, during the occurrence of high mortality events [1, 46].

The *C. glaucum* and its congeneric *C. edule* have been the subject of much attention in the 70s, in physiological and ecological studies as many authors have thoroughly described the differences between the two species, since there was some controversy about their actual geographic distribution and local occurrence [8, 9, 12, 13, 39-41, 43, 47, 48]. Since then, genetically based studies published in the 90s have confirmed the existence of these two morphologically distinguishable species along the eastern Atlantic and Mediterranean coastal waters [11, 12].

The cockles have a wide geographical distribution along the north-eastern Atlantic coastline from the western region of the Barents Sea and the Baltic Sea to the Iberian Peninsula, into the Mediterranean, the Black and Caspian Seas, and south along the coast of West Africa to Senegal [1, 21, 23, 28, 36, 46, 49]. Several species of cockles are commercially harvested and cultured around the world and are consumed for their taste and nutritional value [20, 28, 46, 50, 51]. Since this bivalve is edible and widespread in Western and Northern Europe, it is one of the main non-cultured bivalve species most captured in Western European waters and one of the most abundant mollusc species in European bays and estuaries [1, 18, 20, 36]. Harvesting cockles is deeply ingrained in the history and culture of European countries, and according to the Food and Agriculture Organization of the United Nations (FAO), the average annual harvest of edible cockles was 17,073 tonnes (from 2008 till 2014) [1, 36, 52]. Annual production of cockles in Europe varied between 14,000 and 26,000 tonnes from 2014 to

2017, with production dominated by the United Kingdom, Spain, Portugal, and Denmark [1]. In Portugal, this species is very common [11, 53] along all coastal areas, including Ria de Aveiro, the Óbidos lagoon, the Tagus and Sado estuaries, and the Albufeira lagoon, among others [11, 54]. Cockles fisheries landings reached 2,285 tonnes in 2019 and 3,921 tonnes in 2021 [55, 56]. Additionally, in the ten years period between 2010 to 2020, *Cerastoderma* was the bivalve with the highest landings in the country with 66,186 tonnes, surpassing the surf clam (41,049 tonnes, *Spisula solida*) and the grooved carpet shell (or Palourde clam, *Ruditapes decussatus*) (41,713 tonnes) (Source: *Direção-Geral de Recursos Naturais, Segurança e Serviços Marítimos* - DGRM).

In most areas, both species occur in sympatry, and distinguishing them based on morphological characteristics is very difficult [11, 45, 48]. Despite that, species and geographic identification studies often use morphological characteristics as tools to acquire most of these species' information [12, 45].

1.1 Morphological context

Morphological diversity is the product of genetic, developmental, and environmental influences as well as their interactions [57-63]. It has long been reported in the literature that bivalve shell morphology (including *Cerastoderma* spp.'s) has great phenotypic plasticity in order to adapt to the different environmental and ecological conditions [12, 39, 64-71]. Furthermore, it is widely accepted that ontogenetic changes in shell morphology (size and shape) and certain shell morphologic features (e. g., ribs) are a result of an adaptative strategy to withstand specific environmental conditions, life habits (above or below the substrate, how it feeds, and how/if it locomotes) and habitat preferences [6, 12, 13, 62, 70-74]. Several environmental factors are known to influence shell shape, some endogenous (genetic and physiological) and others exogenous (biotic and abiotic interactions) (*e.g.*, latitude; depth; tidal exposure or shore level; water hydrodynamic factors such as waves, turbulence, and currents; type of substrate; salinity; trophic conditions; type of natural predator; and parasitism) (*e.g.*, [6, 12, 13, 62, 67, 69, 70, 75-79]). Aside from that, burrowing ability, efficiency and behaviour also affect the relative growth and shape of bivalve species (including its shell) [6, 69, 75, 79]. Since shape is a key morphological characteristic that reflects both environmental and ecological aspects, diverse ecological studies cover morphological variation in shell and can be correlated with the adaptation of the species to its habitat [6, 13, 74, 80, 81].

Establishing the relationships between bivalve morphology and habitat can be useful for a multitude of ecologically related topics including conservation and monitoring [61, 71, 82]. For example, shell shape is routinely used for morphological recognition of the taxonomy of bivalves [80]. Since bivalves (such as *Cerastoderma* spp.) have hard shells, they are excellent candidates for morphological analysis based on shape, because no deformation can occur during manipulation [61, 79, 83-86]. Such studies can consist of geographically based analysis of the morphological variance that occurs in individuals from different groups solely based on the analysis of their shape profiles [69, 72].

Additionally, morphological diversity can used for traceability [87-89], which is an important concept in a fisheries context [89]. Traceability is based on population discrimination and stock fingerprinting and has been developing exponentially [89-91]. The core principle of population discrimination is that individuals from different areas display individual traits [89]. Applying this, traceability techniques in a fisheries context can be used to determine the origin of specimens for which the harvest location is in doubt [92]. Its application can lie in the identification and verification of illegal fishing activity, for instance [89] and references therein. It can also benefit consumers through the

detection of fraudulent claims about the origin of seafood products, since the increase of farming activity might lead to mislabelling farmed species as wild-caught species [93].

1.2 Morphometry

Morphometry is the quantification of shape [94]. Traditional or classic morphometry use bivariate or multivariate methods applied to basic linear measurements (such as width, height, and length), generally taken directly from the animal, and widely used to describe, compare, and distinguish shape [71, 86, 95-99]. However, basic linear measurements might not discriminate between size and shape of the structure (since they tend to be autocorrelated with the size of animals, therefore being poor shape descriptors) and are rarely sufficient to capture significant shape information about an organism [86, 98, 100-104]. The shape of biological structures is complex (for example bivalve shells, which have high morphological plasticity) and, because of that, crucial information may be lost if only linear morphometric measurements are applied [86, 102-104]. Nevertheless, the study of morphometry of organisms has experienced a revitalised growing interest in several scientific fields, in the last decades, associated with the development of geometric morphometrics (GM) [71, 94, 95, 97, 105-110]. These techniques focus on the geometry of form in several dimensions, providing more accurate descriptions of morphological shape and allowing direct visualisation of the shape difference of the objects studied [86, 96, 97, 103-105], capturing the relationships between measurements in ways that traditional morphometrics do not [111, 112].

1.3 Geometric morphometrics

While morphometry is the study of the shape variation and covariation among variables, GM provide a visual representation of shape change, and produce shape variables that can be analysed using statistics (*e.g.*, [61, 96, 104, 113]). Therefore, these methods allow the quantification of shapes in great detail and the respective statistical processing of their variations by removing the factor of size and symmetry (*e.g.*, [94, 104, 106, 108, 109, 114-118]). GM methods allow the derivation of components of shape and size, preserving the main geometric proprieties of the studied individuals (*e.g.*, [94-96, 104, 106, 108, 114-116]). It has been successfully applied in many fields, for example, biology, ecology, evolution, ontogeny, taxonomy, fishery, among others (*e.g.*, [84, 113-115, 118-123]). GM techniques were initially based on two main approaches: (1) analysis of outline/contour, where the border of a specimen is analysed (continue matrix data), and (2) Landmark (LM) analysis, where the change in the position of identified homologous points in each sample are quantified (two or three dimensions) [61, 82, 84, 86, 99, 104]. Both approaches provide distinct information and can be applied in different circumstances [61, 82]. Moreover, both technics are an important tool for the analysis of geographical variation in morphology and/or species discrimination (*e.g.*, [69, 83-86, 114-116, 124, 125]). Morphological variation in bivalves has been addressed in several studies dealing with changes in shell morphology (even if subtle), using GM techniques which are turning into a standard approach to study the patterns of bivalve shell shape variation [61, 79, 82, 83, 86].

1.4 Landmarks and landmark-based geometric morphometrics

To correctly compare shapes, it is important to define some of the structural properties of objects (bottom, left side, anterior part, etc) [109]. A distinct mathematical point (*i.e.*, defined by coordinates in two or three dimensions) for which position is comparable among objects is a LM [109, 119, 113]. For biological objects, LMs correspond to an anatomical position comparable among objects and is defined as an anatomical LM [104, 108, 109, 113]. Anatomical LMs are assumed to be homologous having a similar embryological, anatomical, or historical origin [98, 109, 119]. An example of a non-homologous LM would be if some specimens had two scars and others just one scar, thus being dubious and hardly comparable, which is common in bivalves [86, 109]. In this work, LMs are considered homologous points which can be identified in all specimens and bear information on the geometry of biological forms [98, 108, 126].

Most morphometric methods, including linear methods, use the locations of LMs, since inter-landmark measurements are collected when measuring distances with a calliper, for example, length or height of a shell [109]. Three main categories of LMs have been described [108]: Type I, Type II, and Type III [109, 113]. These categories are not always mutually exclusive as the same LM can be classified into two categories [113]. Type I LMs are points that correspond to discrete juxtaposition of tissues or to sufficiently small features to be defined by a single point, such as a point where two or more bones meet, or a small foramen [109, 113]. This kind of LM is preferred since it has a true biological homology origin [109]. Type II LMs are points where correspondence is only indicated by geometry [113]. Self-evident features occurring along elements that can be completely homologized (*e.g.*, bones) serve as representations of the type of geometry that directs the placement of these LMs [109, 113]. Its homological basis may have weaker biological grounds than Type I [109]. Type III LMs are points that are again indicated only by geometry, but where the correspondence is much more loosely supported and no longer self-evident, such as the endpoints of diameters or the intersection of inter-landmark segments [109, 113]. There is also another type of LMs that sometimes is not differentiated from Type III called semi-landmarks or pseudo-landmarks [109]. These are points defined by a construction that permits the identification of the LMs in question and require a subjective choice of the observer [109, 113]. Considering what was said earlier about the criteria for homology, Type I and II LMs represent morphologically equivalent structures (primary homologies) in different organisms most accurately [113]. On the other hand, Type III LMs (and semi or pseudo-landmarks) can be less objective since the LMs do not accurately reflect primary homologies due to their dependence on the subjective choices of the observer [113]. If it is possible to identify LMs, in spite of their type, and measure their coordinates, it becomes easy to calculate distances, angles, etc. These coordinates contain the richest geometric information available in any object [109]. While using LMs as a GM method, the shape of anatomical objects from the coordinates of homologous locations can be defined as the information left in a figure after location, scale and orientation data are removed [99, 107, 108, 120-123, 127]. GM techniques that use LMs have demonstrated high sensitivities, specificities, and accuracies, and do not require highly skilled technicians nor the use of costly reagents and equipment [99]. It has been used to determine shell shape variation in multiple bivalves, being able to detect subtle changes in certain structures or parts of the specimen under study [86, 117].

An integral part of primary homology assessment are alignments aiming to maximise the similarity of the LM configurations, which is why LM configurations are usually aligned before being used for analysis [86, 113]. The most prevalent type of alignment for GM data is the least-squares Procrustes superimposition [104], which is a procedure that minimises the sum-of-squared differences between configurations in a multivariate Euclidean space, by keeping one configuration as a reference while the

rest is rotated sequentially until the sum-of-the-squared residuals between the corresponding coordinates is minimised [63, 86, 119, 128]. The sum-of-squares value is smaller as the coherence between the datasets is greater [63]. To summarise, the Procrustes distance, which is the conventional metric for overall shape dissimilarity in GM, is the Euclidean distance between shapes superimposed to minimise the sum-of-squared distances between homologous LMs, a minimization carried out by translation, rotation, and rescaling, preserving shape [129]. This alignment retains a series of properties that make it suitable for further statistical testing [97, 102, 104, 127, 130-132].

1.5 Classification

Classification is the assignment of individuals based on measured variables, to predefined groups [133]. Classification techniques can be used to identify and compare groups, therefore being useful in species and location identification (thus traceability) (*e.g.*, [106; 134; 135]). A canonical variates analysis (CVA) is a linear discriminant analysis (LDA), a multivariate classification technique frequently applied to morphometric data (including Procrustes shape coordinates) in a multitude of contexts (*e.g.*, ecology, paleontology, forensics, etc…) (*e.g.*, [106, 119, 129, 133, 136]). One of the most important advantages of GM is its power to analyse large datasets and to visualise results as actual shapes or shape deformations (*e.g.*, [95, 108, 110, 137]). However, the large number of variables is a disadvantage to the application of LDA or CVA [133]. Moreover, it has been argued that this application to shape coordinates is inappropriate since discriminant functions cannot be interpreted as biological parameters [108] and an ordination based on a CVA does not preserve the Procrustes geometry [133]. CVA generally offers no interpretation as biological factors, is unlikely to correspond to any particular developmental process, gene, or environmental effect, and is simply a tool for human decision-making [133].

As an alternative to CVA, a Random forest (RF) classifier can be applied. RF is an ensemble classifier tree-based learning algorithm, *i.e.*, the algorithm averages the predictions over multiple individual decision trees [138, 139]. The individual trees are built on bootstrap samples, where each classifier in the ensemble is trained on a random subset of a training samples set, rather than on the original sample [138-141]. This approach is called bootstrap aggregation or bagging, and it reduces overfitting [138, 139]. Usually, about 70% of the samples (in-bag samples) are used to train the trees (training dataset) and the remaining 30% (out-of-the bag samples) are used in an internal cross-validation technique to estimate how well the resulting RF model performs (test or validation dataset) [138, 139, 141-144]. In substance, the RF classifier yields reliable classifications using predictions derived from an ensemble of decision trees [142]. The algorithm creates trees that have high variance and low bias [142]. It easily adapts to nonlinearities found in the data and performs greatly on prediction tasks [139]. In recent years, the use of machine-learning algorithms has increased in the scientific field [139]. RF as devised by [142], is one of the best-performing learning algorithms and most successful methods currently available to handle small and large datasets [138-140]. RF has received increasing attention due to the accuracy of classification results obtained, the speed of processing (*e.g.*, [141, 145-148]), and simplicity of use [140]. This methodology has been successfully involved in various works (*e.g.*, [141, 144, 147-150]).

1.6 Objectives

The aim of this work is to study the geographical variation of the shell shape of two bivalve species with commercial interest, the common cockle (*Cerastoderma edule*) and the lagoon cockle (*Cerastoderma glaucum*), common in several estuarine and lagoon systems of the Portuguese coast, exploring the possibility of inferring the origin of an individual using shell morphology and multiple landmark-based geometric morphometrics methodologies and secondly, to test if this can be used as a traceability technique to unravel the origin of economically and ecologically important marine products, such as the cockles.

2. Materials and methods

2.1 Study area

This study took place at five different aquatic systems located along the western Portuguese coast (Fig. 2.1): Ria de Aveiro (coastal lagoon), the Óbidos coastal lagoon, the Tagus and Sado estuaries, and the Albufeira coastal lagoon. For easier computation and labelling of the results, each system's name was simplified to "Aveiro", "Óbidos", "Tagus", "Sado", and "Albufeira", respectively.

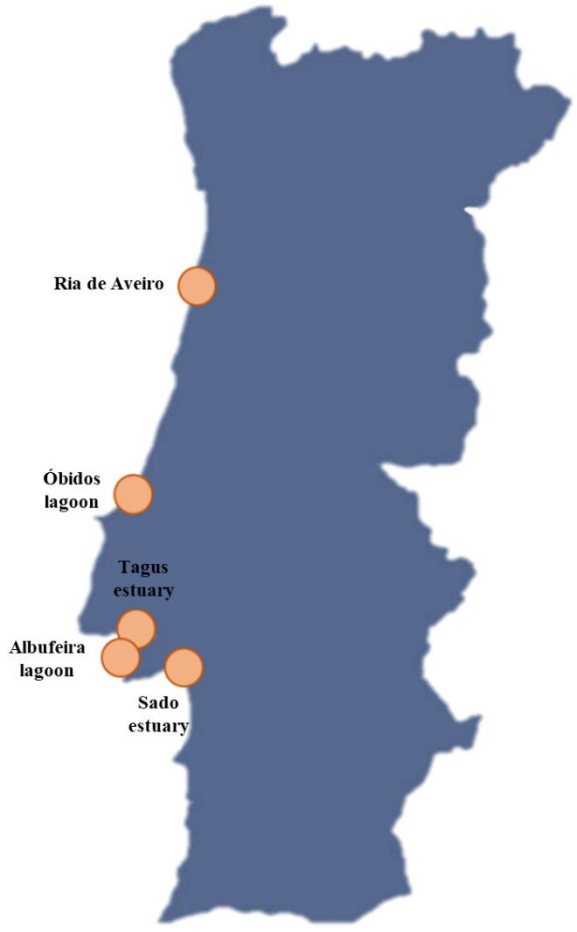


Fig. 2. 1 - Location of the cockles' (*Cerastoderma edule* and *Cerastoderma glaucum*) sampling sites (Ria de Aveiro; the Óbidos coastal lagoon; the Tagus estuary; the Albufeira coastal lagoon; the Sado estuary).

The Ria de Aveiro, located on the Northwest Atlantic coast of Portugal, is 45 km long and 10 km wide, covering an area extending from 66 to 83 km^2, at low and high tide respectively [151]. It is characterised by two main zones: (1) the central area under tidal influence, with strong currents; (2) two major channels, consisting mainly of very shallow intertidal areas, lower currents (~ 0.1 $m.s^{-1}$) [151]. Ria de Aveiro is typically characterised by lowest water temperatures in the winter (minimum temperatures around 13 ºC), progressive warming through spring until summer (maximum temperatures around 19 ºC), and a gradual cooling from autumn to winter [152]. Cockles are often found in this system [54, 153] and are the most important bivalve fishery resource in Ria de Aveiro [28]. For the local fishing communities, it serves as a source of food and work, and it generates significant profits in the area [28]. Between 2015 to 2020, the cockles represented a total of 98.947 tonnes of captures, representing around 88% of the total landings of this species in Portugal (Source: *Direção-Geral de Recursos Naturais, Segurança e Serviços Marítimos* - DGRM). This data is undervalued because an unknown percentage of trading is unreported and commercialised via a clandestine market [28]. Cockles are collected year-round due to their abundance, ease of capture and strong market demand, and, on the most accessible cockle beds, there is also an unknown amount of unreported fishing effort since it requires little investment in manpower and equipment [28]. Therefore, despite being a species of local high density, a substantial decrease in abundance has been observed and it is hypothesised to be due to overfishing, recruitment failures, or a combination of all these factors [153, 154].

The Óbidos lagoon is a semi-closed coastal lagoon located 150 km north of Lisbon with an area of 7 km^2, being one of the largest coastal lagoons in Portugal, permanently connected to the sea [155-157]. Although this ecosystem is mostly oriented NW-SE, its direction can change based on tidal oscillation and the system's sand entrance, being also very unstable because of its proximity to the sea [157]. The Óbidos lagoon is characterised by semi-diurnal tides and includes areas with different morphological and sedimentary characteristics presenting an average depth of 1.5 to 2 m and having a central area and two main channels: the Braço do Bom Sucesso and the Braço da Barrosa [155, 158]. This lagoon supports a wide range of diverse and intricate habitats [157], with the cockles (*Cerastoderma* spp.) as one of the most common species [54, 159]. Landings of this species accounted for 11.531 tonnes between 2015 and 2020 (Source: *Direção-Geral de Recursos Naturais, Segurança e Serviços Marítimos* - DGRM).

The Tagus estuary, located on the Portuguese southwestern coast, is the largest estuary in Portugal and one of the largest in Europe, covering an area of approximately 325 km^2 [160, 161]. It is a mesotidal estuary, roughly 30 km long and exhibits a complex morphology that can be divided into three distinct zones: upstream area, middle area, and external area [50, 160, 161]. The upstream area of the estuary connects with the river through a channel of 200 m width [161]. The middle area is the largest part being 15 km wide and with an average depth of about 7 m [161]. The estuary is connected to the Atlantic Ocean, through a deep, long, and narrow channel, with 12 km length, 2 km width, and 45 m depth [161, 162]. This estuary also presents local wave formation conditions due to its orientation toward the prevailing winds [50]. Cockles occur frequently in this aquatic system, despite being a less important capture (0 tonnes landed between 2015 and 2020) when compared to other coastal regions (*e.g.*, Ria de Aveiro, the Óbidos lagoon) [54, 163].

The Sado estuary is the second largest Portuguese estuary and one of the largest in Europe, occupying an area of approximately 240 km^2 [164]. It is a mesotidal estuary and presents an elongated shape, partially divided into two channels: the North channel, with an average depth of 10 m, dominated by a weaker flood current, and the South channel, with an average depth superior to 25 m, dominated by a more intense ebb current [165-167]. Although having a lower harvesting relevance compared to

other commercial species (*e.g. Ruditapes philippinarum*), cockles presented total landings of 1.807 tonnes between 2015 and 2020 [54, 163] (Source: *Direção-Geral de Recursos Naturais, Segurança e Serviços Marítimos* - DGRM).

The Albufeira semi-enclosed coastal lagoon is located on a moderately exposed mesotidal coast (the Western Portuguese Coast) [168, 169]. It has a NE-SW orientation in relation to the coast and a flooded surface of 1.3 km^2, with maximum depth of 15 m, width of 625 m and length of 3.5 m [168, 170]. The lagoon is made up of two basins separated by a narrow and shallow channel: the main body named Lagoa Grande, with an average depth of 10 m, and the smaller body, Lagoa Pequena, with 2 m in depth on average [170]. The lagoon is sheltered from the ocean by a sand barrier with variable width (300 m on average) [170]. Cockles are frequently found and harvested in this system [54], presenting a total landing of 48.816 tonnes between 2015 and 2020 (Source: *Direção-Geral de Recursos Naturais, Segurança e Serviços Marítimos* - DGRM).

2.2 Sampling and laboratory work

Sampling was conducted with the support of professional harvesters and as part of the NIPOGES project (*Estado atual das populações de amêijoa-japonesa da Ria de Aveiro, lagoa de Óbidos e estuários do Tejo e Sado – bases científicas para uma gestão sustentável do recurso*). The specimens were collected at different locations in five systems (Ria de Aveiro, the Óbidos coastal lagoon, the Tagus and Sado estuaries, and the Albufeira coastal lagoon) between the end of summer and beginning of autumn of 2021. For the Tagus estuary in particular, samples were collected in 12 specific stations (Fig. 2.2) (within the scope of the NIPOGES project). Stations number 8, 12, and 14 were located in the upstream area, while stations 26, 29, 30, 31, and at the beach were located downstream. Stations 23, 23A, 23B, and 23C were in between the previous locations (Fig. 2.2).

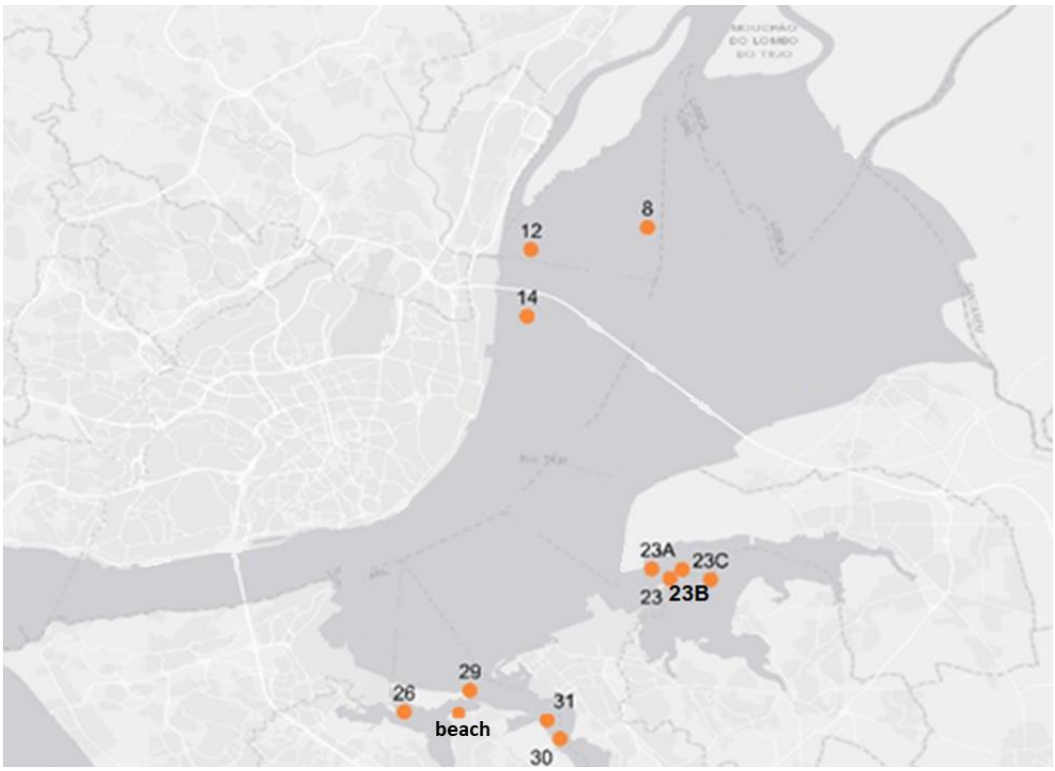


Fig. 2. 2 - Location of the cockles' (*Cerastoderma edule* and *Cerastoderma glaucum*) sampling stations in the Tagus estuary.

Collected specimens were taken to the laboratory, cleaned, and prepared for analysis. Their taxonomic identification was done based on morphological features (*e.g.*, [11, 72, 171]).

2.3. Morphometric methods

2.3.1. Linear morphometrics methods

Height, length, and width were measured for each specimen, using a digital calliper to the lowest millimetre. The mean height, length, width, and shell roundness (calculated dividing width by length) were estimated along with the respective standard error (95%) using the R program (version 4.2.0) [172], specifically RStudio under the version 2022.02.2+485 [173]. The R packages used for the linear morphometric analysis were "ggplot2" [174], "readxl" [175], "car" [176], and "gridExtra" [177].

2.3.2. Geometric morphometrics

For the landmark (LM) locations, 16 points that could be relevant to be used as LMs for the statistical analysis were marked with a pencil on each organism's right valve (504 specimens in total) (Fig. 2.3) to facilitate identification, by the same observer to minimise errors. LMs were identified based on previous studies (*e.g.*, [82-86]) and particularities of the species being studied. In the event of a sampled specimen having multiple muscle scars, the most prominent scar was used to mark the LMs. LMs 1-4 are points marked on the muscle scars (#1 and #2 on the anterior muscle scar, and #3 and #4 on the posterior muscle scar); LMs 5-8 are points marked on the hinge teeth; LMs #9, and #10 are both placed on the hinge; LM #11 is placed on the tip of the ligament and representing the top extremity of

the shell; LMs #12, #13, and #14 are points marked on the extremities of the shell, with LM #14 being the left extremity, #13 the right extremity and #12 the bottom extremity; LM #15 was obtained by counting the 10th rib of the shell and placing the point where that rib connects with the pallial line; LM #16 represents the intersection between the pallial line and an imaginary line that connected LM #11 with LM #12. The 16 LMs considered in this study were classified as follows: LMs numbers 1-4, and 15 as Type I; LMs 5-10 as Type II; LMs 11-14 as Type III; and 16 as pseudo or semi-landmark (Fig. 2.3).

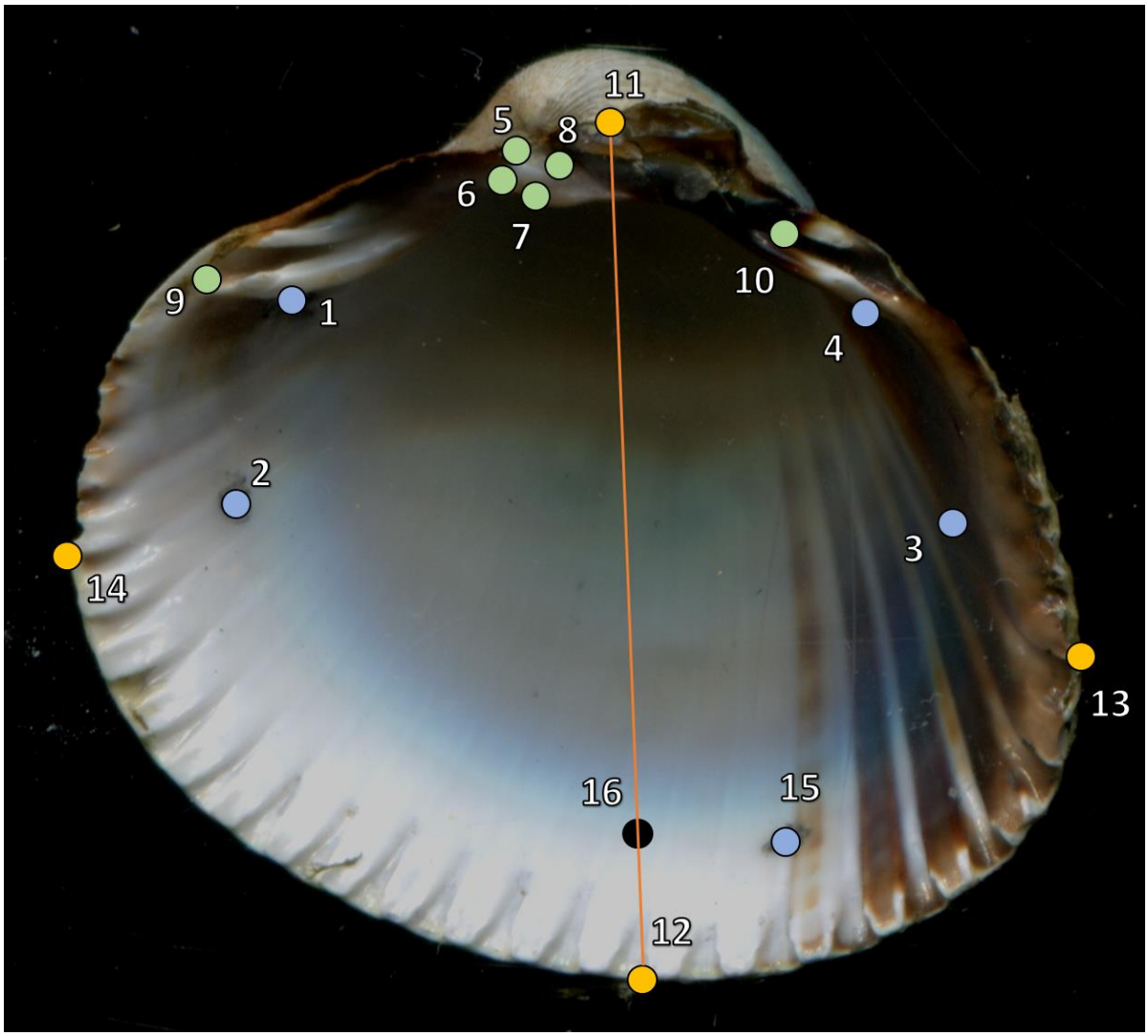


Fig. 2. 3 - Right valve shell of a *Cerastoderma edule* specimen analysed in this study with all landmark placements (in blue: Type I; in green: Type II; in yellow: Type III; and in black: pseudo or semi-landmark) and a representation of the imaginary line used to determine landmark #16 in orange.

All marked valves were digitised using a flat scanner (EPSON Perfection V300 Photo), to avoid potential camera distortions associated with zoom levels, and singled out, one by one, using the GIMP program version 2.10.30 [178].

The TPS series (specifically tpsDig2 version 2.32) [179] and RStudio [173] (including the package "geomorph" [180-183]; "dplyr" [184]; "tidyr" [185]; "Hmisc" [186]; "factoextra" [187]; "ggplot2" [174]; "gridExtra" [177]; "Morpho" [188]; and "randomForest" [189]) were used to obtain the shape variables and to perform the statistical analysis. Using tpsDig2 [179], the coordinates of 16 LMs for each specimen's right valve were obtained by the same observer.

In order to minimize the sum-of-squared distances, a Generalised Procrustes Analysis (GPA) [190] was used to translate, rotate and rescale the LM configurations to a common centroid as reference (to unit centroid size) (*e.g.*, [102, 118, 190, 191]). The centroid size (Csize), calculated by the square root of the sum-of-squared Euclidean distances of a set of LMs from their geometric centroid [99, 108, 130, 191; 192], was used to represent size [94, 104, 192]. With it, the GPA performed removed all information unrelated to shape [97, 190, 193], adjusting for size [104; 118] and eliminating variation due to differences in scale and orientation [83, 86, 131, 191]. To be able to better visualise the data and have an idea of how it varies, mean variation of Csize between species and aquatic systems was represented using RStudio [173] and with a standard error of 95%.

A mean LM configuration called consensus or reference/average configuration was obtained as a result of the GPA, and used to visualise the direction and magnitude of shape variation in the morphometric space [83, 116, 126, 190, 192].

An analysis of variance (ANOVA) was performed using RStudio [173]. A Procrustes ANOVA was performed using the "geomorph" package [180-183] with random distributions of F-statistics to calculate z-scores and P-values [194] and assess statistical hypotheses describing patterns of shape variation and covariation for a set of Procrustes-aligned coordinates [194] (annex I). Model selection procedure started with a full model, which contained all interactions possible with Csize, species, and aquatic systems, to a simpler model, from which the interaction between species and Csize, between aquatic system and Csize, and between those three factors were eliminated. The criterion used for the elimination of interactions between aquatic systems and Csize, and between the three factors from the model was if the interactions were statistically significant or not (P-value ≤ 0.05). Models with and without each term were compared in each step using an ANOVA. For the interaction between species and Csize, it was decided that the z-scores were too small in comparison with the other scores, therefore it was eliminated from the model. The full model was compared to the simpler model to assess model equity/equality. Thus, the simpler model, containing the factors Csize, species, aquatic system, and the interaction between species and aquatic system, was selected as the final model to use for the analysis. The dataset was tested for normality. These values were based on 1,000 random permutations.

A Principal Component Analysis (PCA) (analogous to the Relative Warp Analysis; annex II) was done using the "factoextra" package [187] to decompose the data into axes in which shape is partitioned as independent components of the shape variation in the original dataset. This analysis is designed in a way where the axes represent different information and the first axis represents most variation in the dataset, followed by the second, and so on [94, 130]. To observe patterns of variation within the dataset, the axes that reflect maximum variation and covariation were examined [109, 118, 119, 127, 130, 132]. When patterns seemed relevant in PCA, a plot of shape differences between a reference and target specimen(s) was performed, using RStudio [173] ("geomorph" package [180-183]), in order to verify those patterns, with vector lengths multiplied by a factor of two to improve observation.

A trajectory analysis (TA) was performed on RStudio [173] with the package "geomorph" [180-183]. A trajectory is defined by a sequence of points in the data space [181]. These trajectories can be quantified for various attributes (size, orientation, and shape), and comparisons of these attributes enable the statistical comparison of shape change trajectories [181, 195-198]. The TA performed here represented the description and comparison of geometric attributes of change between species and aquatic systems [197, 198]. With the function used for this data set, it was possible to quantify multivariate trajectories from a set of observations and assess variation in the attributes of the trajectories, namely angle, distance and direction [181, 199], using the basic arguments of a pairwise

function with an argument for trajectory points [180, 181, 199]. In the context of this work, the function was used to indicate differences between trajectory magnitudes (*i.e.*, the path length of trajectories) and to summarise the analysis as trajectory correlations, which are angles between trajectory directions (major axes of variation) [180, 181].

With this data, a pairwise analysis could have been used to test pairwise differences between least-squares means, *i.e.*, Euclidean distance between least-squares means [199]. However, for this study, it was decided to not use a pairwise analysis since the information that could be gathered by performing it is part of the information obtained with a TA [198].

In addition, a morphological disparity (MD) analysis was carried out, to measure morphological diversity [129]. High MD represents a high diversity of morphologies (*i.e.*, shapes or body plans) and is likely associated with high levels of functional and ecological diversity [200]. Disparity analyses (DA) can be based on multiple characteristics [200]. For example, in this study, a Csize-based descriptor (mathematical descriptor) and a model-based descriptor, were used to describe MD [200]. Different types of descriptors capture and emphasise different aspects of morphology, levels of trait correlation, and scales of change [201]. Even for the same set of sampled taxa, disparity patterns may or may not be consistent across different types of descriptors [201]. None of the descriptors is better than the other, and, as most disparity datasets are multidimensional, it is often beneficial to use more than one measure to summarise different aspects of variation [200], thus why the use of two different kind of descriptors here (mathematical descriptor and model-based descriptor). The function (from package "geomorph" [180-183]) used for this analysis estimated MD and performed pairwise comparisons to identify differences between groups [180, 181]. MD was estimated as the Procrustes variance, overall or for groups, using randomised residuals of a linear model fit, through permutation [180, 181]. The function was performed using landmark-based geometric morphometrics (GM) data where Procrustes variance is calculated for the whole dataset (first descriptor) and using the final model (second descriptor) in order to compare MDs in a more detailed way [180, 181].

For classification, both a canonical variates analysis (CVA) and a random forest (RF) were applied. As mentioned earlier, it has been argued that the use of CVA is not appropriate for this kind of data [133]. However, since most of previous works use this analysis (*e.g.*, [106, 119, 129, 136]), it was important to include it in this work in order to have comparison grounds with past and future works, in spite of not being an ideal methodology for the kind of data used here. The CVA done in this work was performed on RStudio [173] with the package "Morpho" [188] in order to assess possible classification of the cockles shells' localisation using landmark-based GM data. With the cross-validated classification results in frequency, the proportion of each class error (*i.e.*, total of wrong predictions/classifications divided by the total of observed specimens for a specific class) was calculated and converted to a percentage (multiplied by 100).

A RF can be performed as a classification method as an alternative to CVA. RF classifier was performed on RStudio [173] with the package "randomForest" [189] in order to assess if landmark-based GM data of cockles shells were reliable predictors of their geographic origin and species differentiation. The dataset was randomly sub-divided into a training set (70%) and a test/validation set (30%) following previous works (*e.g.*, [139, 141, 144]). The RF algorithm was set on 500 randomly built and uncorrelated trees and 5 splits. Three RF models were built, (1) for species discrimination; (2) for aquatic systems and (3) for both (species and aquatic systems). Internal cross-validation was achieved by applying the model to the test/validation set, calculating the probability of correct classification (PCC) *i.e.*, accuracy. PCC or accuracy was calculated by the package [189] as all correctly

classified observations divided by all observations (for test or validation set). The proportion of each class error (*i.e.*, total of wrong predictions divided by the total of observed specimens for a specific class) was recorded and then converted to a percentage (multiplied by 100). For more statistical information on RF see [142].

2.4. Others

Scientific references for this work were found using B-ON (https://www.b-on.pt/) and Web Of Science (https://www.webofknowledge.com) as search engines. For each search, species (*Cerastoderma*, *edule*, Cardium, *glaucum*) and method (morphometric, morphometry, TPSdig, and geometric morphometrics) were used as keywords to a maximum of 150 results per search. In addition to those, some references related to the context of the work were searched later on Google Scholar (https://scholar.google.com/) through different keywords (*e.g.*, random forest, bivalve morphotype, symmetry, allometry, etc…), following the logical thought process of this work and its results/discussion. Finally, some references used in this work were suggested by specialists in the field/supervisors.

3. Results

A total of 358 *Cerastoderma edule* and 146 *Cerastoderma glaucum* were collected in the five studied aquatic systems, summing 504 specimens (Tab. 3.1).

Tab. 3. 1 - Number of specimens (*Cerastoderma edule* and *Cerastoderma glaucum*) collected in the five studied systems (the Sado and Tagus estuaries, the Albufeira and Óbidos lagoons, and Ria de Aveiro).

Number of specimens	Ria de Aveiro	Óbidos lagoon	Tagus estuary	Albufeira lagoon	Sado estuary	Total
C. edule	41	39	166	66	46	358
C. glaucum	14	11	68	19	34	146
Total	55	50	236	85	80	504

3.1. Linear morphometrics

Shell heights of *C. edule* varied between 20.350 ± 2.173 mm (mean and SE 95%) in the Tagus estuary and 28.604 ± 1.766 mm in Ria de Aveiro, whereas length varied between 22.877 ± 2.324 mm in the Tagus estuary and 31.798 ± 2.105 mm in Ria de Aveiro and widths, between 16.140 ± 1.988 mm in the Tagus estuary and 22.682 ± 1.425 mm in Ria de Aveiro (Fig. 3.1; annex III). Shell means of the linear measurements between the Óbidos lagoon, the Sado estuary and the Albufeira coastal lagoon of *C. edule* showed very little variation on average (≤ 1 mm) (Fig. 3.1; annex III). For *C. glaucum* mean height varied from 21.500 ± 1.845 mm in the Tagus estuary to 29.519 ± 1.660 mm in Ria de Aveiro, the mean length varied from 24.206 ± 2.051 mm in the Tagus estuary to 33.324 ± 1.929 mm in Ria de Aveiro, while mean width varied between 17.456 ± 1.648 mm in the Tagus estuary and 23.811 ± 1.621 mm in Ria de Aveiro (Fig. 3.1; annex III).

The shells of *C. edule* collected in Ria de Aveiro were the roundest ones, followed by the Tagus estuary, the Sado estuary, the Óbidos lagoon and the Albufeira lagoon (Fig. 3.1). *C. glaucum*'s shell

roundness varied more than *C. edule*'s, but the variation between aquatic coastal systems showed a similar pattern, with the shells from the Tagus estuary being rounder, followed by Ria de Aveiro, the Sado estuary, the Albufeira lagoon, and, finally, the Óbidos lagoon (Fig. 3.1). Note also that the sites with the largest and smallest shells, were also the roundest, showing the roundedness was not related with size (Fig. 3.1).

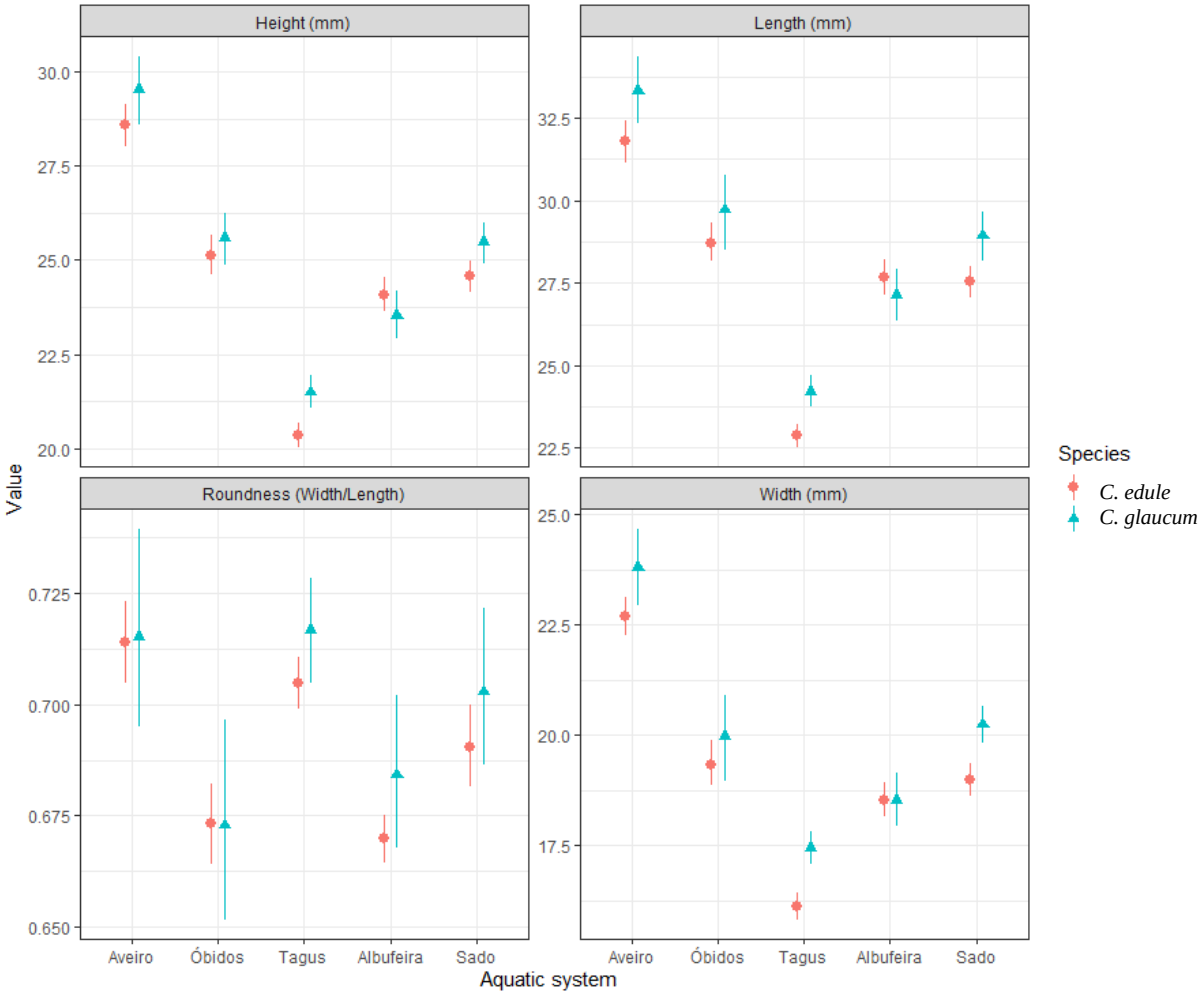


Fig. 3. 1 - Mean and respective standard error (95%) for height, length and width (in millimetres) of *Cerastoderma edule* (red) and *Cerastoderma glaucum* (blue) at the five studied systems. Shell roundness (width divided by length) (mean and SE95%) of *Cerastoderma edule* and *Cerastoderma glaucum* at the studied coastal systems (Ria de Aveiro, the Sado and Tagus estuaries, the Óbidos and Albufeira coastal lagoons).

Within the Tagus estuary, and only *C. edule* specimens specifically (Fig. 3.2), shell roundness varied between sampling stations, *e.g.*, shells from E.26 and E.29 (downstream) appeared rounder and with more variability than the other stations. Additionally, all stations localised downstream (26, 29, 30, 31, and at the beach) presented shells with a mean roundness equal or superior to 0.71, while all the shells from the upstream's stations (8, 12, and 14) displayed roundness between 0.69 and 0.70 (Fig. 3.2). Finally, *C. edule*'s shell harvested in stations 23, 23A, 23B, and 23C had roundness varying from 0.68 to 0.69 (Fig. 3.2). Stations 23, 23A, 23B, and 23C had roundnesses more similar to the upstream stations, but still, they had the less round shells of all the stations (Fig. 3.2).

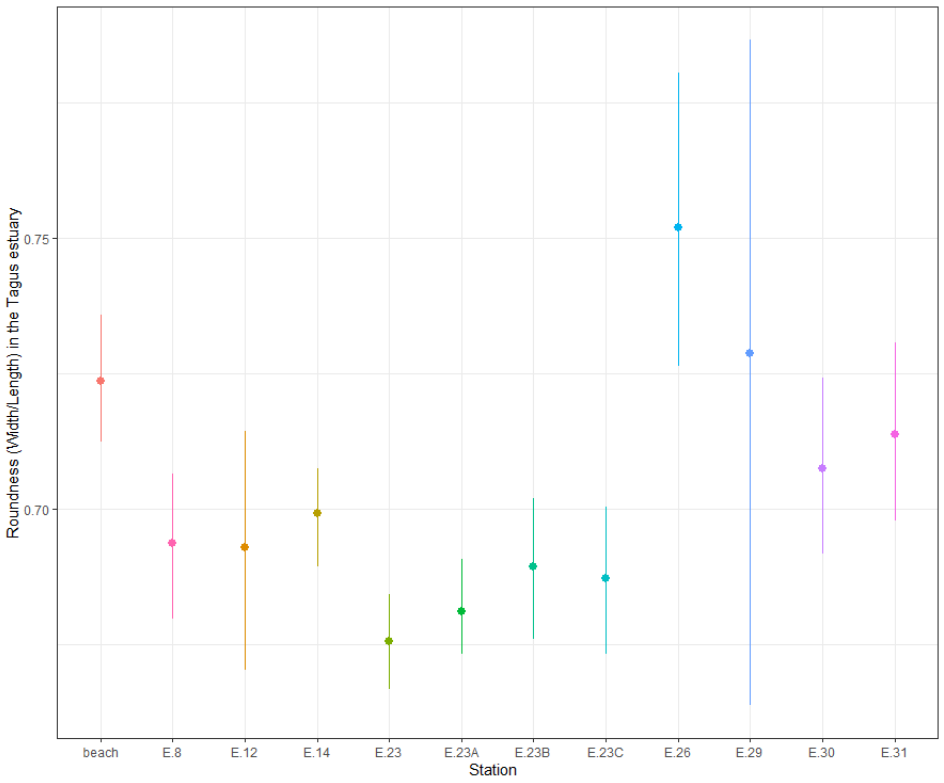


Fig. 3. 2 - Shell roundness (width divided by length) for *Cerastoderma edule* and the different sampling stations in the Tagus estuary (mean and SE95%).

3.2. Geometric morphometrics

The dispersion around landmarks (LMs) position varied across the LM number. LMs numbers 15, 16, 12, and 13 showed higher dispersion, while LMs numbers 5-8, and 11 (corresponding to the hinge teeth and umbo zone) had smaller variation (Fig. 3.3).

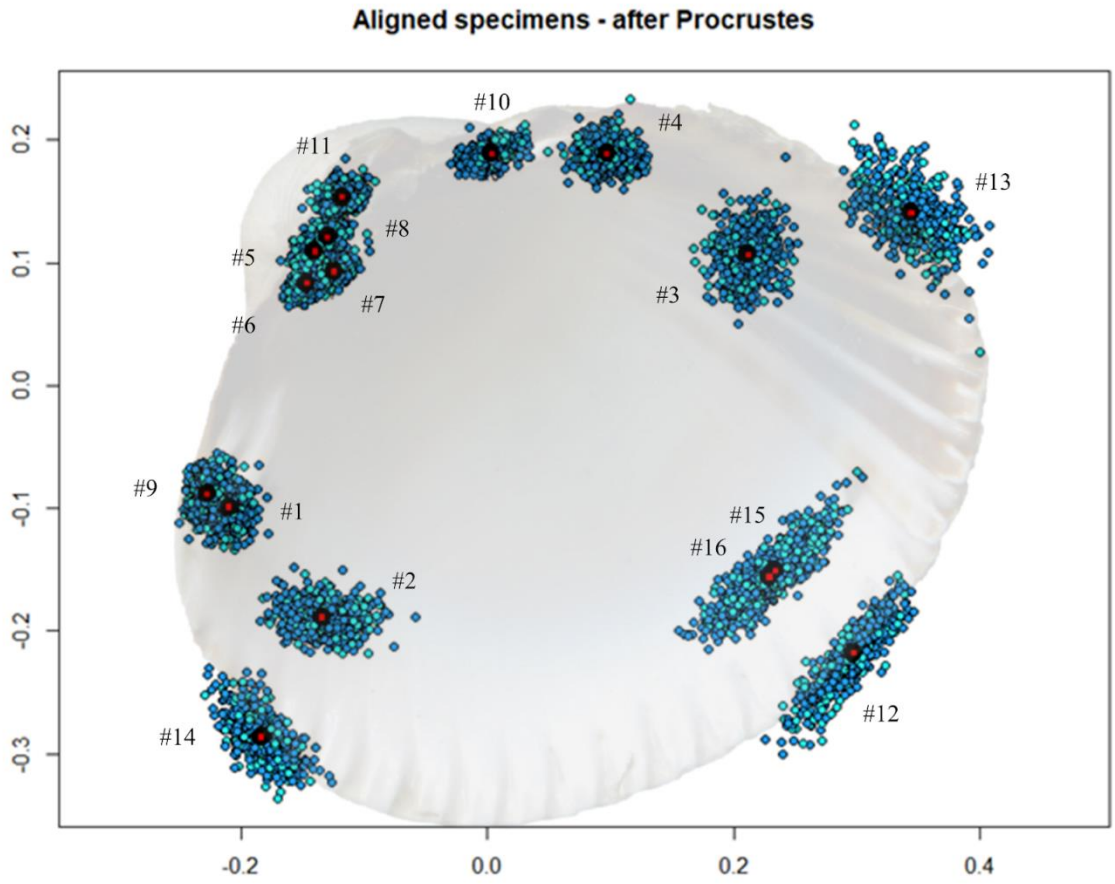


Fig. 3. 3 - Consensus configuration, i.e., average configuration of landmarks (red balls) with the variation around each numbered landmark (blue points represent each specimen, darker blue for *Cerastoderma edule* and light blue for *Cerastoderma glaucum*). In the background, a representative shell was added to help understand shell shape and position.

In terms of shape size, *C. edule* specimens were always slightly smaller than *C. glaucum* (Fig. 3.4) (measured using centroid size, which represents overall shape size) in all sampled aquatic systems, except in the Albufeira lagoon, where the opposite was observed. Specimens from Ria de Aveiro were the largest on average, followed by shells from the Óbidos lagoon and the Sado estuary, then the Albufeira lagoon, and, lastly, the smallest centroid size was found in the Tagus estuary (Fig. 3.4). Additionally, differences between aquatic systems were much higher than differences between species (Fig. 3.4).

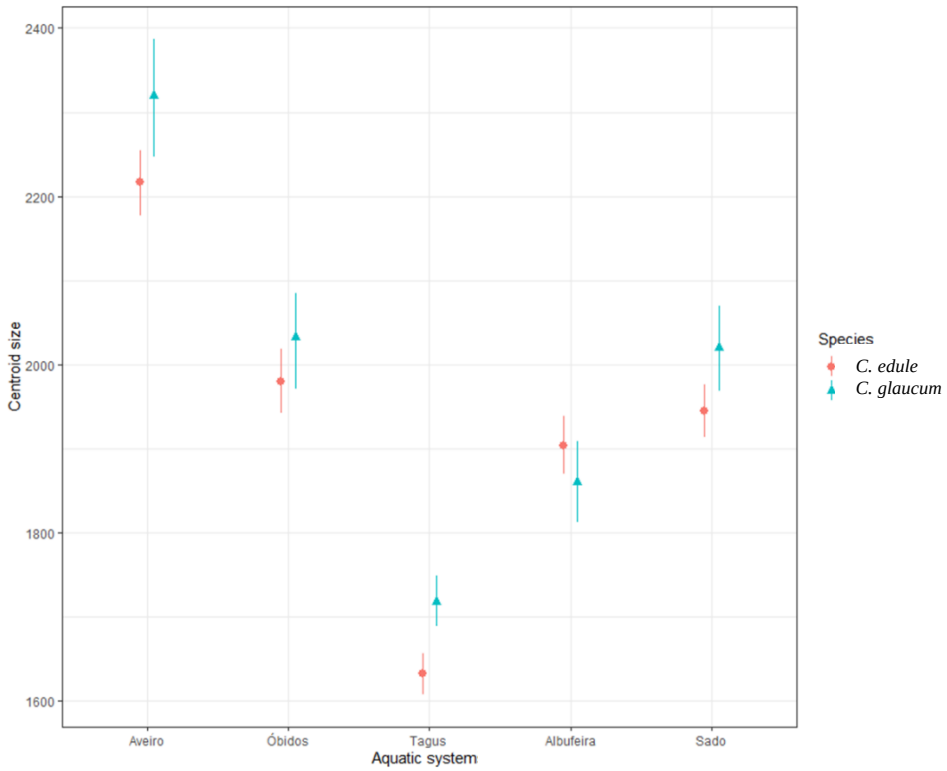


Fig. 3. 4 - Variation of centroid size between species and aquatic systems (mean and SE95%). *Cerastoderma edule* in red and *Cerastoderma glaucum* in blue.

After model selection, the final and best linear model for this analysis (containing the factors centroid size (Csize), species, aquatic system, and the interaction between species and aquatic system) showed shell shape significantly changed with specimen size (F-value = 46.789 and P-value = 0.001), between the two species (F-value = 24.088 and P-value = 0.001), and between the aquatic system (F-value = 41.372 and P-value = 0.001) (P-value ≤ 5%) (Tab. 3.2). It also indicated that the interaction between species and aquatic systems was statistically significant (P-value ≤ 5%) (F-value = 7. 105 and P-value = 0.001), thus differences in shape between species varied across the systems (Tab. 3.2).

Tab. 3. 2 - Results of the simpler model of the factors Centroid size, species, aquatic system, and the interaction between species and aquatic system.

	Degrees of freedom	Sum-of-squares	Mean squares	R-Squared statistic	F-statistic	Z-test	P-value
Centroid Size	1	0.241	0.241	0.062	46.789	6.240	0.001
Species	1	0.123	0.123	0.032	24.088	5.082	0.001
Systems	4	0.846	0.211	0.218	41.372	7.652	0.001
Species:Systems	4	0.145	0.036	0.038	7.105	6.750	0.001
Residuals	493	2.520	0.005	0.651			
Total	503	3.873					

The first axis of the Principal component analysis (PCA) (1st dimension or Principal component 1 *i.e.*, PC1) explained 45.3% whereas the second axis explained 13.7% of the variation and covariation in shell shape. *C. edule* showed a positive and higher score than *C. glaucum* in both the first and second PCs (Principal components) (Fig. 3.5). These differences were reflected in changes in position of all LMs except for numbers 1, 10, and 15 (Fig. 3.6).

Fig. 3. 5 - Principal component analysis (PCA) of the shape variables representing variation and covariation of the two studied species (with 95% confidence ellipses) within the configuration set distributed in two axes, the first axis (PC1 explaining 45.23% of the variation and covariation) and the second axis (PC2 explaining with 13.7% of the variation and covariation), without aquatic system differentiation. In red *Cerastoderma edule* and in blue *Cerastoderma glaucum*.

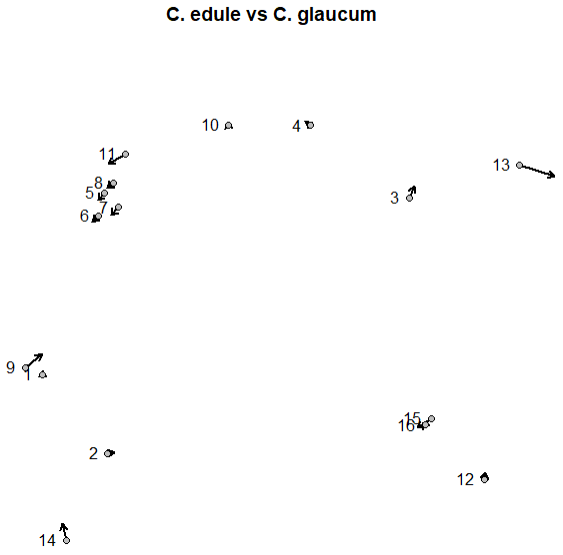


Fig. 3. 6 - Visual representation of the differences in the average landmark position between the two bivalve species *Cerastoderma edule* and *Cerastoderma glaucum* (vectors amplified by two).

When accounting for shape variation by systems (ignoring species differentiation) (Fig. 3.7), specimens from the Sado and Tagus estuaries appear to show more similar shell shapes, followed by Óbidos and Aveiro, whereas Albufeira showed more extreme shapes in PC1 (Fig. 3.7). Since the Sado estuary had almost no variation in PC2 (or second dimension), the variation observed in LMs (numbers 1-4, 9, 10, 12-16) was mainly linked to PC1, with LMs #12 and #16 recording biggest variation (Fig. 3.7 and Fig. 3.8A). Additionally, Ria de Aveiro was the only system in the fourth quadrant (PC1 > 0 and PC2 < 0) in Fig. 3.7 and its variation from the overall mean was observed in all LMs except for numbers 4 and 10 (Fig. 3.8B). The two groups at the extremes of the first component were the Albufeira lagoon and the Tagus estuary, whereas the extremes of the second dimension were Ria de Aveiro and the Albufeira lagoon (Fig. 3.7). Looking at the shape differences between each extreme, the Albufeira lagoon and the Tagus estuary's specimens varied in all LMs, and the Albufeira lagoon and Ria de Aveiro's specimens differed in all LMs except number 15 (corresponding to the only Type I LM placed on the paleal line) (Fig. 3.8C and Fig. 3.8D).

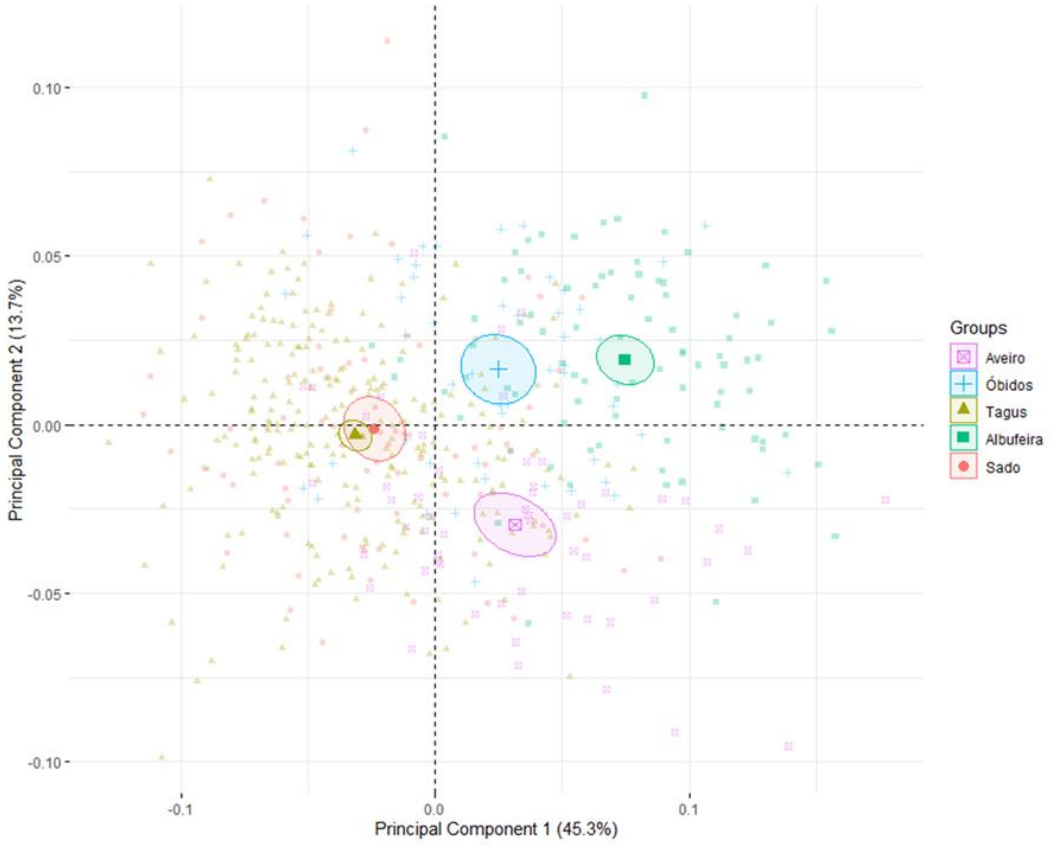


Fig. 3. 7 - Principal component analysis (PCA) of the shape variables representing variation and covariation of the five studied aquatic systems (with 95% confidence ellipses) within the configuration set distributed in two axes, the first axis (PC1 explaining 45.3% of the variation and the second axis (PC2) explaining 13.7% of the variation).

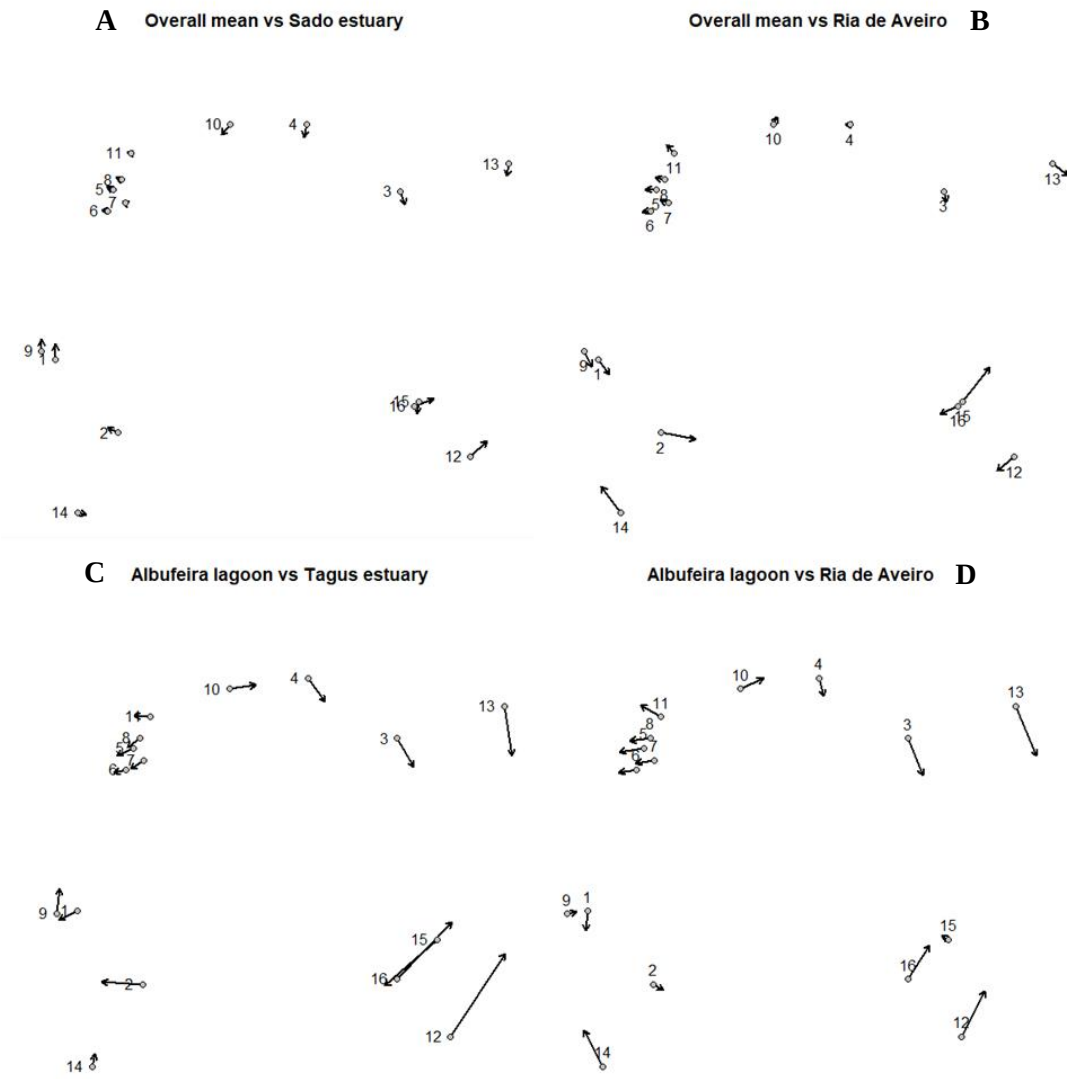


Fig. 3. 8 - Visual representation of the differences in the average landmark position between two groups with vector length multiplied by a factor of two. A - Landmark variation between the average landmark configurations of the overall mean versus the Sado estuary. B - Landmark variation between average landmark configurations of the overall mean versus Ria de Aveiro. C - Landmark variation between the average landmark configurations of the Albufeira lagoon and the Tagus estuary. D - Landmark variation between the average landmark configurations of the Albufeira lagoon and Ria de Aveiro.

In the PCA representation of the two species for each studied system overlaid (Fig. 3.9), it is clear that the two groups at the extremes of the first dimension were *C. edule* from the Albufeira lagoon and *C. edule* from Tagus. When looking at the LM variation between them, all differed substantially except LM number 14 (Fig. 3.10A). *C. glaucum* from Ria de Aveiro and *C. edule* from the Óbidos lagoon appeared at the extremes of the second dimension with changes in all LMs minus numbers 2, 4, 10, and 15 (Fig. 3.9 and Fig. 3.10B). Both species from the Sado and Tagus estuaries had similar variations (Fig. 3.9). *C. edule* from the Albufeira lagoon were isolated from the other groups and, when observing the LM variation from the overall mean, all LMs except from #14 varied (Fig. 3.9 and Fig. 3.10C). The only group of *C. edule* observed with the second axis negative was the *C. edule* from Ria de Aveiro, and this variation was reflected in all LMs aside #4, but especially in LMs localised in the bottom of the shell (Fig. 3.9 and Fig. 3.10D). The Albufeira lagoon's *C. glaucum* appeared very similar to the Óbidos lagoon's species (Fig. 3.9). *C. edule* specimens from Ria de Aveiro and the Sado estuary showed opposite patterns (quadrants), showing variation in all LMs (Fig. 3.9 and Fig. 3.10E). Similarly, *C. glaucum* from the Sado and Tagus estuaries presented opposite variation patterns from the Óbidos

lagoon's *C. edule* and the Albufeira lagoon's *C. glaucum*, with visible changes in all 16 LMs (Fig. 3.9 and Fig. 3.10F). The only aquatic system with PC1 positive and PC2 negative (fourth quadrant) for both species was Ria de Aveiro (Fig. 3.9). *C. glaucum* from the Óbidos lagoon had the second dimension very close to zero, meaning the variation observed in LMs numbers 2-4, 9, 10, 12-14, and 16 was mostly associated with PC1 (Fig. 3.9 and Fig. 3.10G). Similarly, Ria de Aveiro's *C. glaucum* was also close to having zero variation, but in the first dimension, showing variation in all LMs aside from numbers 3, 15, and 16 that was linked to PC2 (Fig. 3.9 and Fig. 3.10H). Finally, only the Óbidos and Albufeira lagoons displayed both species in the second quadrant (PC1 and PC2 positive), reflecting variations from the overall mean in all LMs (Fig. 3.9 and Fig. 3.10I).

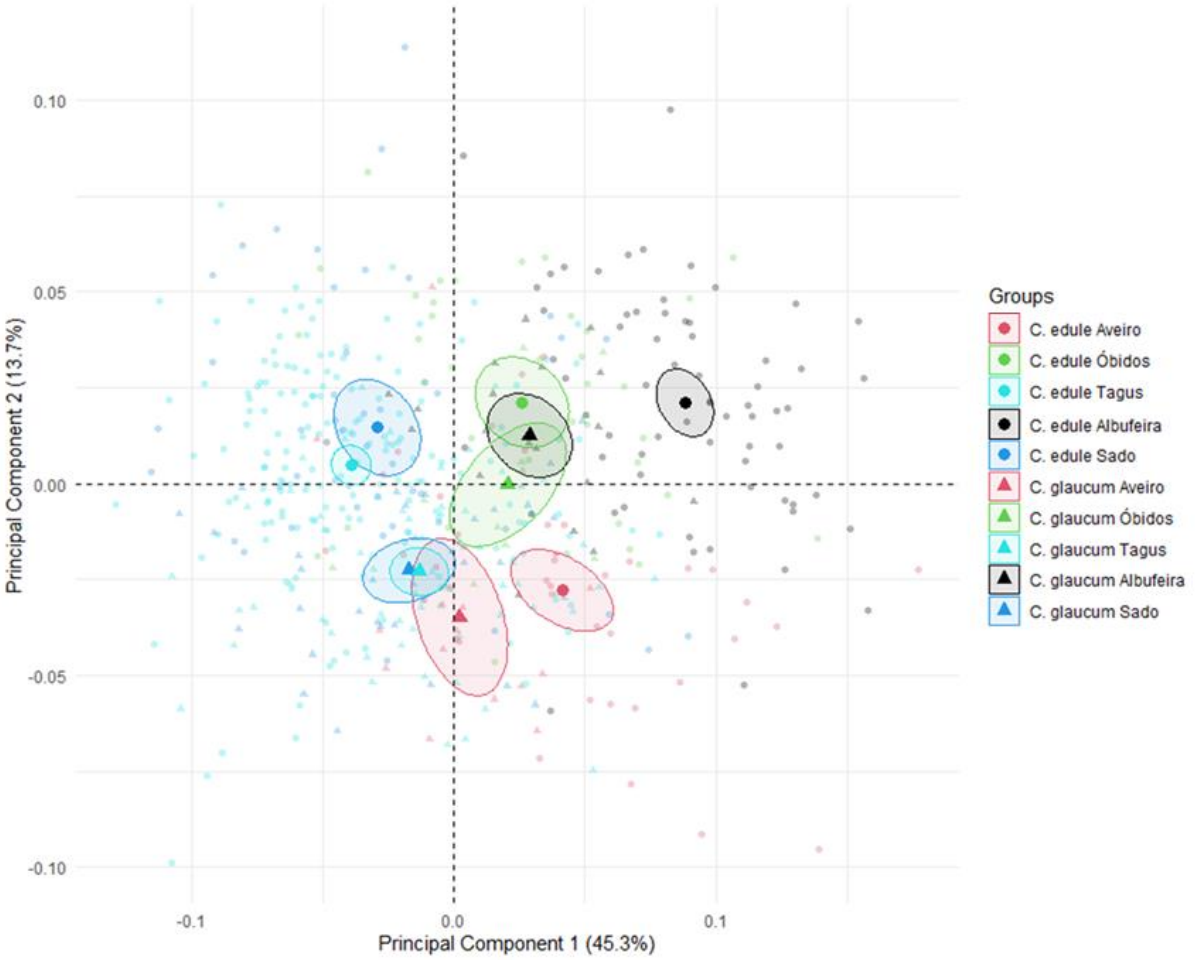
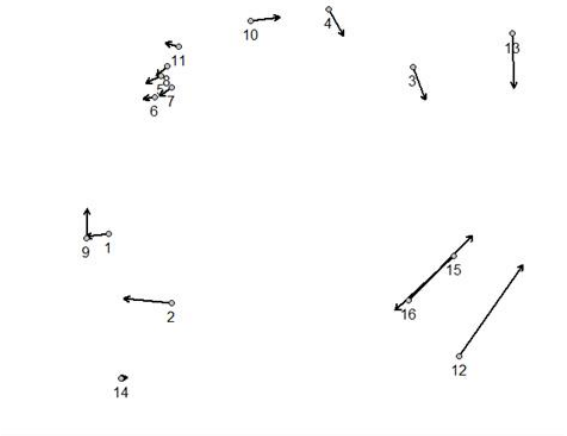
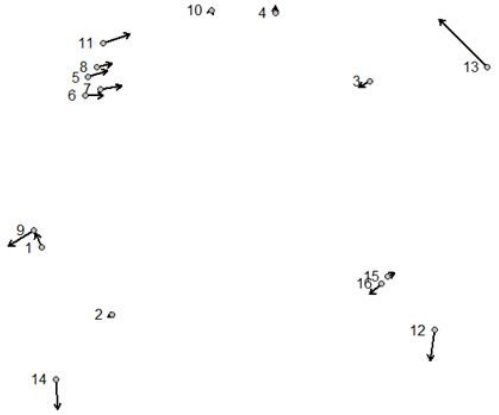


Fig. 3. 9 - Principal component analysis (PCA) of the shape variables representing variation and covariation of the two species and the five studied aquatic systems (with 95% confidence ellipses) within the configuration set distributed in two axes, the first axis (PC1 explaining 45.3% of the variation) and the second axis (PC1 explaining 13.7% of the variation).

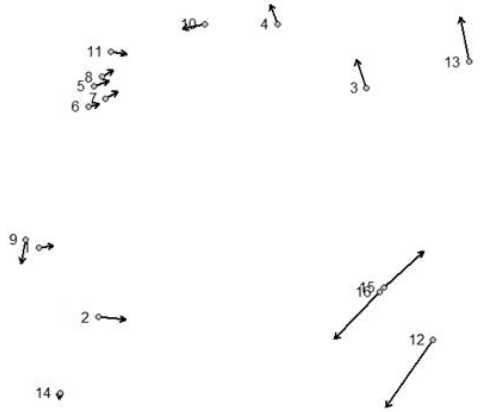
A C. edule Albufeira lagoon vs C. edule Tagus estuary



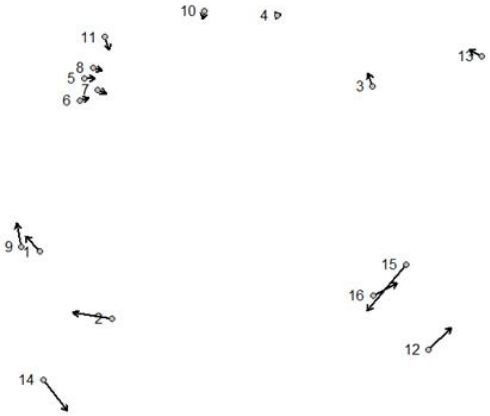
C. glaucum Ria de Aveiro vs C. edule Óbidos lagoon **B**



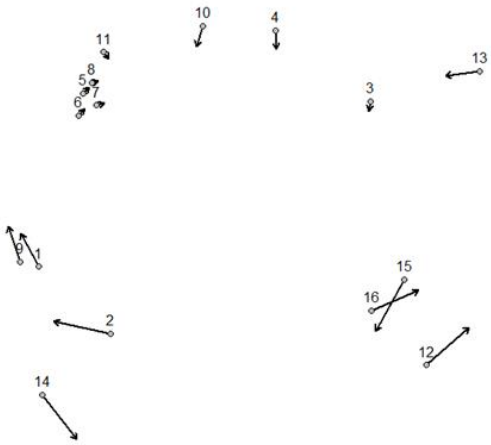
C Overall mean vs C. edule Albufeira lagoon



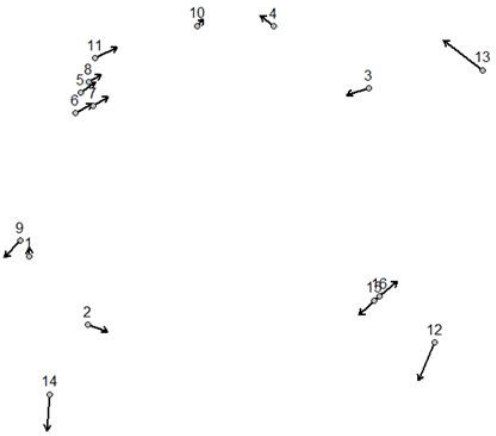
C. edule Ria de Aveiro vs all C. edule except Ria de Aveiro **D**



E C. edule Ria de Aveiro vs C. edule Sado estuary



C. glaucum Sado & Tagus estuaries vs
C. edule Obidos & C. glaucum Albufeira lagoons **F**



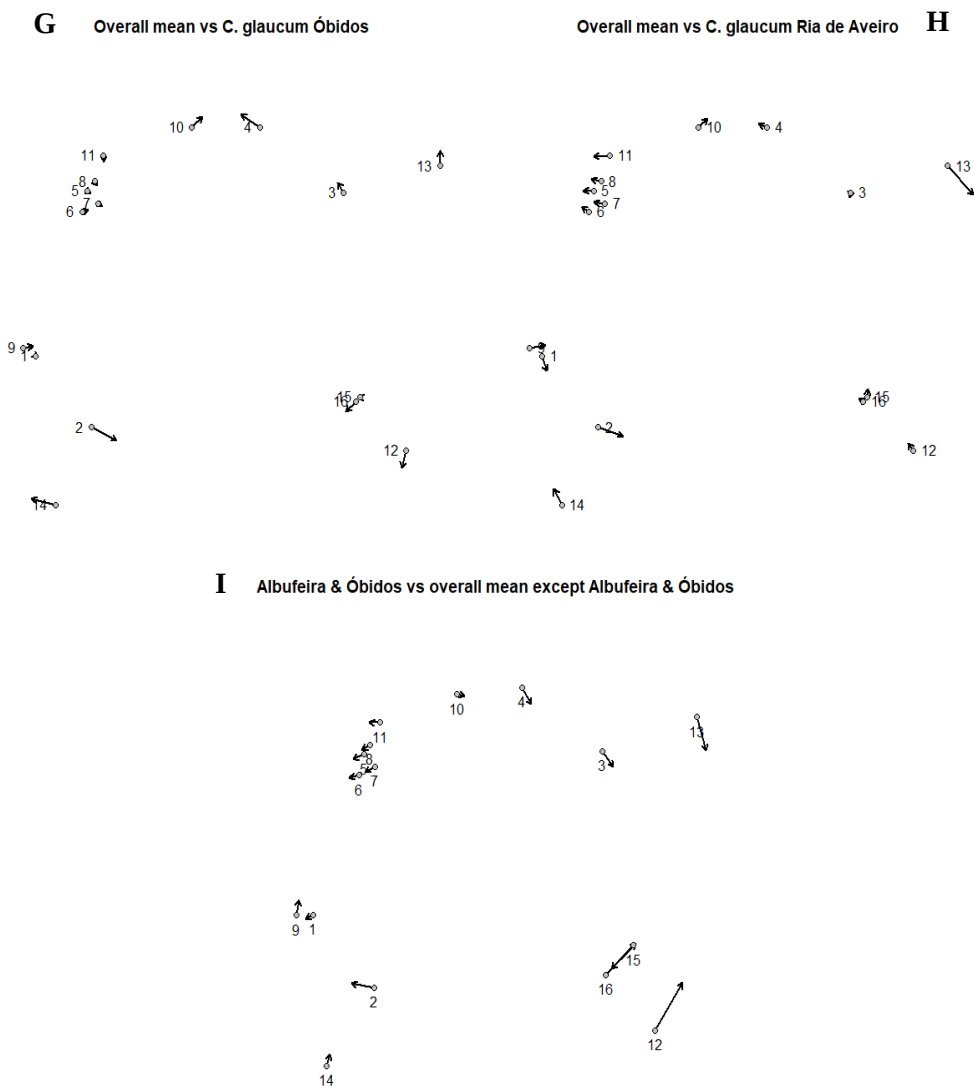


Fig. 3. 10 - Visual representation of the differences in the average landmark position between two groups with vector lengths (multiplied by a factor of two). A - Landmark variation between the average landmark configurations of the *Cerastoderma edule* from the Albufeira lagoon and the Tagus estuary. B - Landmark variation between the average landmark configurations of Ria de Aveiro's *Cerastoderma glaucum* versus the Óbidos lagoon's *Cerastoderma edule*. C - Landmark variation between the average landmark configurations of the overall mean versus the Albufeira lagoon's *Cerastoderma edule*. D - Landmark variation between the average landmark configurations of the *Cerastoderma edule* from Ria de Aveiro and the overall mean from all *Cerastoderma edule* specimens except for Ria de Aveiro. E - Landmark variation between the average landmark configurations of the *Cerastoderma edule* from Ria de Aveiro and from the Sado estuary. F - Landmark variation between the average landmark configurations of the *Cerastoderma glaucum* from the Sado and Tagus estuaries versus *Cerastoderma edule* from the Óbidos lagoon and *Cerastoderma glaucum* from the Albufeira lagoon. G - Landmark variation between the average landmark configurations of the overall mean and the Óbidos lagoon's *Cerastoderma glaucum*. H - Landmark variation between the average landmark configurations of the overall mean and Ria de Aveiro's *Cerastoderma glaucum*. I - Landmark variation between the average landmark configurations of the specimens from Albufeira and Óbidos versus the overall mean excluding samples from the Albufeira and Óbidos lagoons.

From the representation of PC2 (13.7%) and PC3 (8.3%), overlaying the 95% confidence ellipses of the two species for each studied aquatic system (Fig. 3.11), the following patterns were observed: *C. glaucum* from Ria de Aveiro and *C. edule* from the Albufeira lagoon appeared at the extremes of the second dimension, this variation was translated into changes in all LMs except #2 (3.12A); Ria de Aveiro's *C. edule* and the Albufeira lagoon's *C. glaucum* were observed at the extremes of the third dimension, meaning variation in all LMs minus #16 (3.12B); both species from the Sado and Tagus estuaries showed similar shape variations; specimens from Ria de Aveiro and from the Albufeira lagoon

were the groups with the most divergent shape variation, this translated into changes in all LMs even if small (#15 and #2 for *C. edule*, #15 and #4 for *C. glaucum*) (3.12C and 3.12D); all *C. glaucum* (except form the Óbidos lagoon) presented a bigger PC3 than their *C. edule* counterparts associated to shape variation for all LMs save numbers 1, 2, 4, 10 and 15 (3.12E); all *C. edule* showed similar shape patterns except for the group of *C. edule* from Ria de Aveiro which is the only group with the second dimension negative, representing changes in all LMs except number 4 (3.12F); in PC3, the groups that showed variation close to zero were *C. glaucum* from Ria de Aveiro and *C. edule* from the Tagus estuary, the later indicated variation in LMs numbers 2, 4, 10, and 12-16 (3.12G); the Albufeira lagoon's *C. glaucum* is the only group with a positive value in the second dimension, showing shape differences from the all the other groups of *C. glaucum* in all LMs (3.12H); finally, *C. glaucum* from the Óbidos lagoon had similar variation patterns with the other groups of *C. edule*.

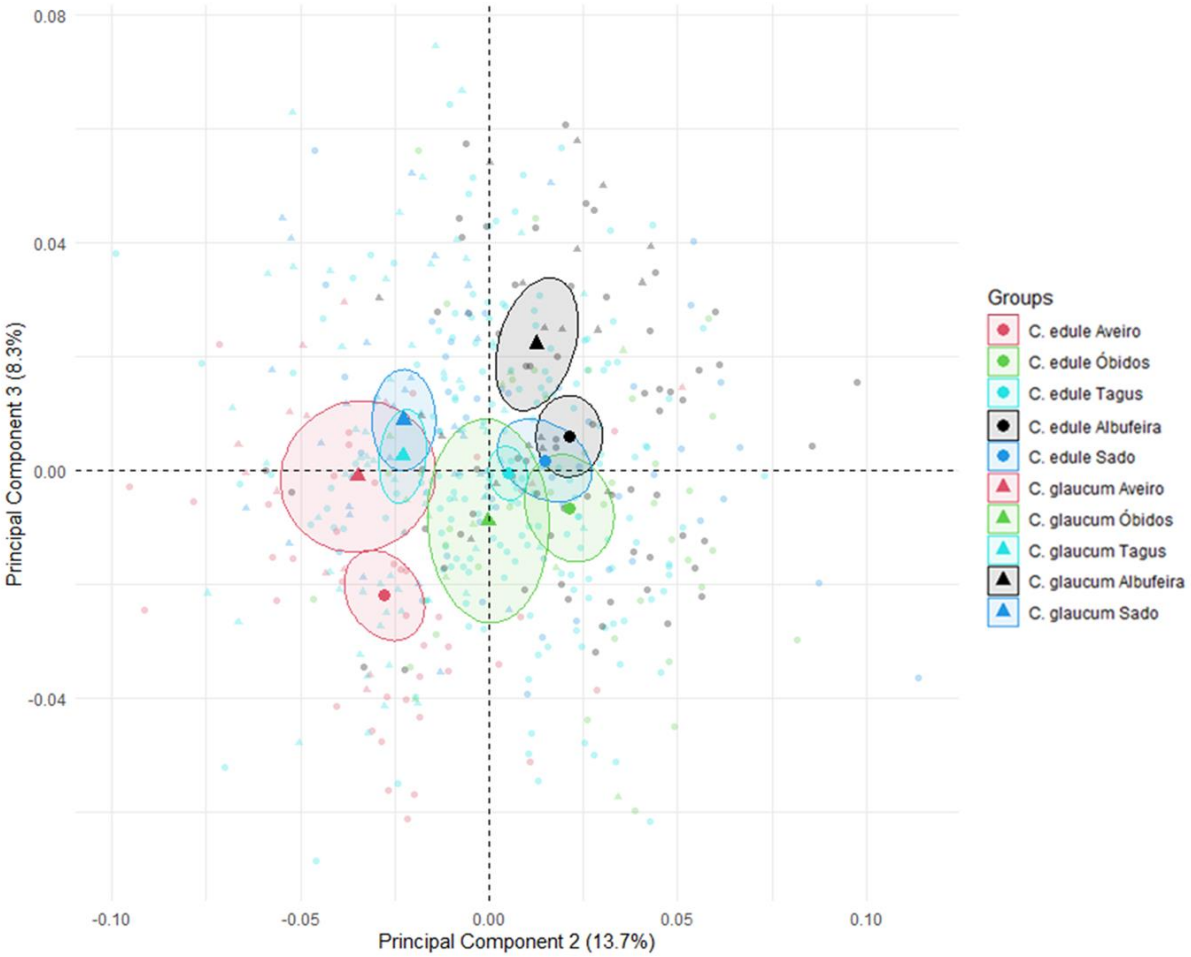
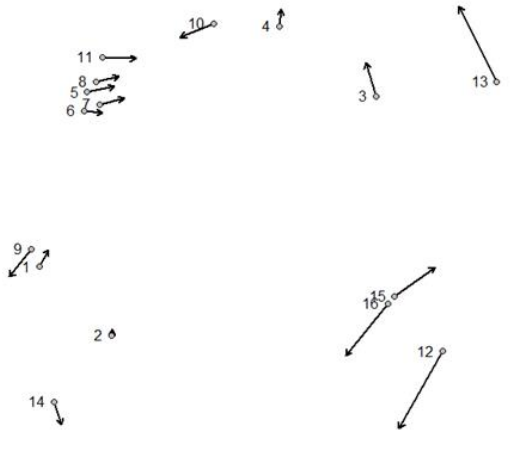
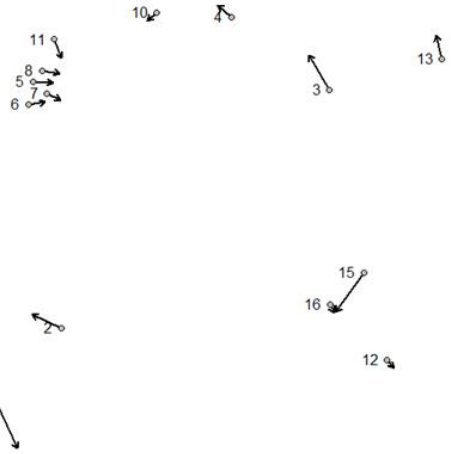


Fig. 3. 11 - Principal component analysis (PCA) of the shape variables representing variation and covariation of the two species and the five studied aquatic systems (with 95% confidence ellipses) within the configuration set distributed in two axes, the second axis (PC2 explaining 13.7% of variation) and the third axis (PC3 explaining 8.3% of variation).

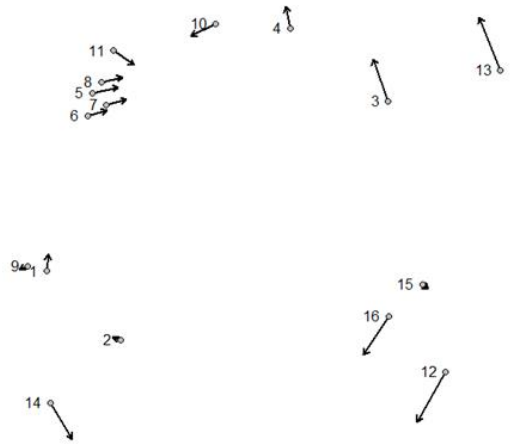
A C. glaucum Ria de Aveiro vs C. edule Albufeira lagoon



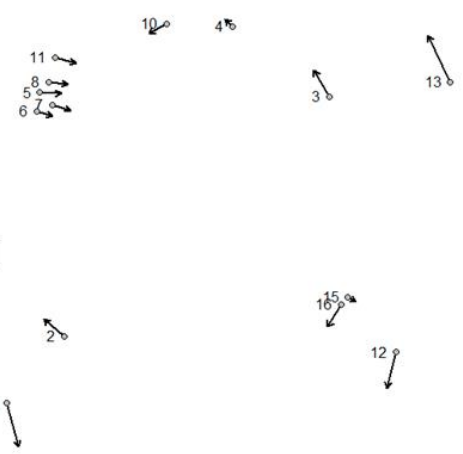
C. edule Ria de Aveiro vs C. glaucum Albufeira lagoon **B**



C C. edule Ria de Aveiro vs C. edule Albufeira lagoon

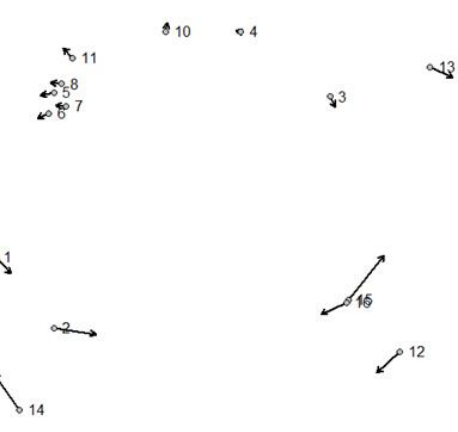


C. glaucum Ria de Aveiro vs C. glaucum Albufeira lagoon **D**



E all C. glaucum except Óbidos vs all C. edule except Óbidos

all C. edule except Aveiro vs C. edule Ria de Aveiro **F**



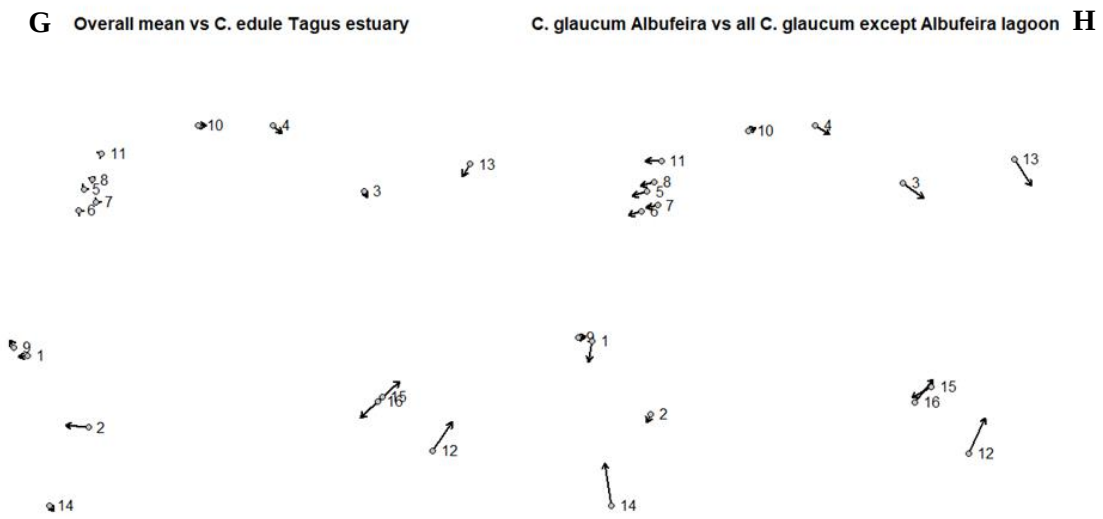


Fig. 3. 12 - Visual representation of the differences in the average landmark position between two groups with vector length amplified by two. A - Landmark variation between the average landmark configurations of the *Cerastoderma glaucum* from Ria de Aveiro and *Cerastoderma edule* from the Albufeira lagoon. B - Landmark variation between the average landmark configurations of Ria de Aveiro's *Cerastoderma edule* versus the Albufeira lagoon's *Cerastoderma glaucum.* C - Landmark variation between the average landmark configurations of the *Cerastoderma edule* from Ria de Aveiro and from the Albufeira lagoon. D - Landmark variation between the average landmark configurations of the *Cerastoderma glaucum* from Ria de Aveiro and from the Albufeira lagoon. E - Landmark variation between the average landmark configurations of the overall mean for each species and all aquatic systems except the Óbidos lagoon. F - Landmark variation between the average landmark configurations of the overall mean of *Cerastoderma edule* except from Ria de Aveiro versus those specimens. G - Landmark variation between the average landmark configurations of the overall mean versus the Tagus estuary's *Cerastoderma edule*. H - Landmark variation between the average landmark configurations of the *Cerastoderma glaucum* from the Albufeira lagoon versus the overall mean except for those specimens.

Patterns observed in the PC3 and PC4, explaining 8.3% and 7.4% of variation respectively and overlaying the 95% confidence ellipses of the two species for each studied aquatic system (Fig. 3.13), were: the two groups observed at the extremes of the fourth dimension were the Sado estuary's *C. edule* and the Óbidos lagoon's *C. glaucum*, representing changes in all LMs (Fig. 3.14A); all *C. edule* had a positive value in the PC4 except for *C. edule* from the Albufeira lagoon, this translated into shape variation in all LMs minus #14 (Fig. 3.14B); *C. edule* from the Sado estuary were isolated from the other groups with bigger PC4 than any other group and showed changes from the overall mean in LMs numbers 1-4, 9, 10, 12-16 (Fig. 3.14C); all *C. glaucum* had PC4 < 0, except for *C. glaucum* from the Albufeira lagoon (Fig. 3.12H); the Sado and Tagus estuaries' *C. glaucum* had similar variation patterns but not *C. edule*, the later specimens presented variations in LMs 1, 3, 4, 10, 12-16 (Fig. 3.14D); in PC3 and PC4, the Tagus estuary's *C. edule* showed variation close to zero (Fig. 3.12G); and, all *C. edule* displayed variation in PC4 bigger than their *C. glaucum* counterparts, except for the specimens in the Albufeira lagoon, meaning changes in all LMs except #10 and #16 (Fig. 3.14E).

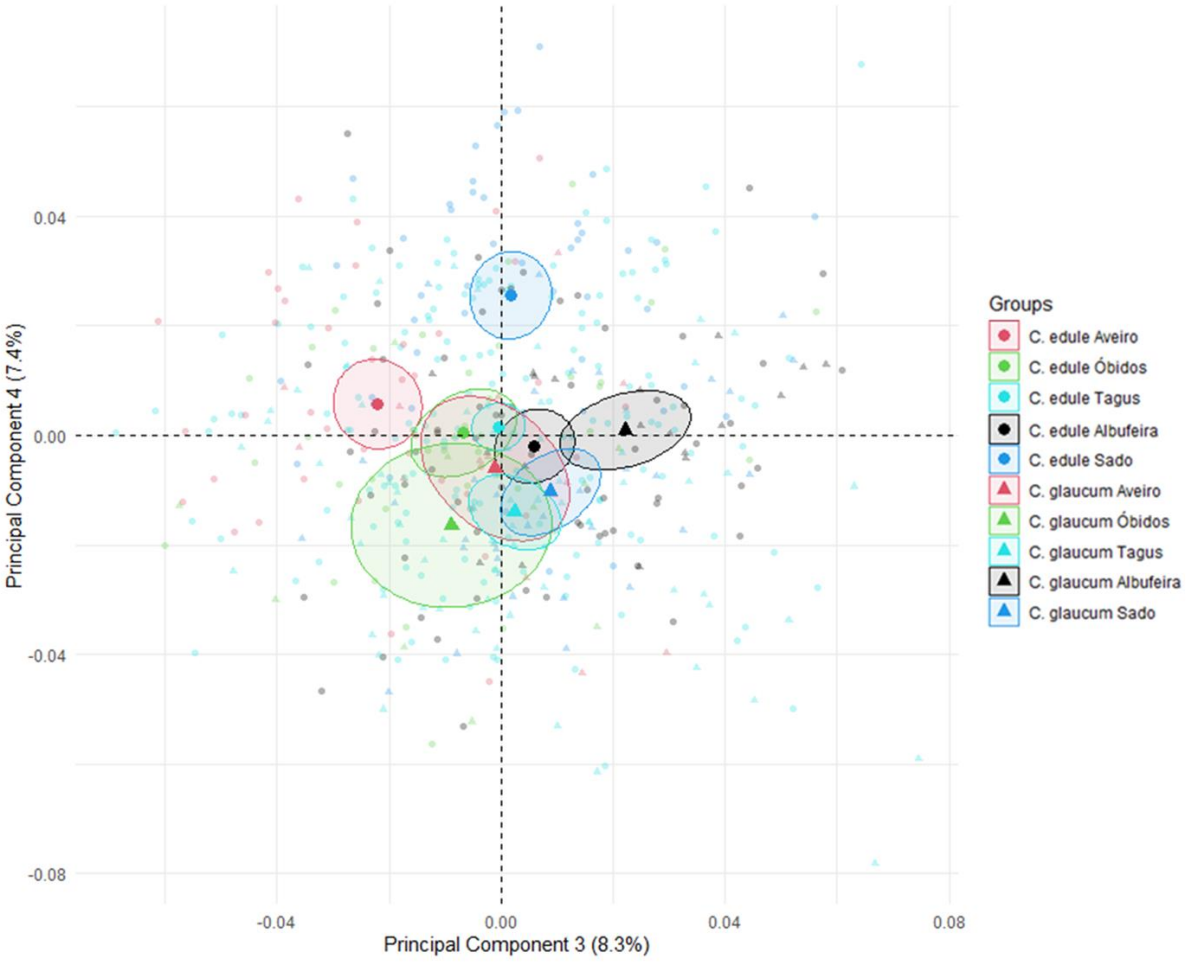


Fig. 3. 13 - Principal component analysis (PCA) of the shape variables representing variation and covariation of the two species and the five studied aquatic systems (with 95% confidence ellipses) within the configuration set distributed in two axes, the third axis (PC3 explaining 8.3% of the variation) and the fourth axis (PC4 explaining 7.4% of the variation).

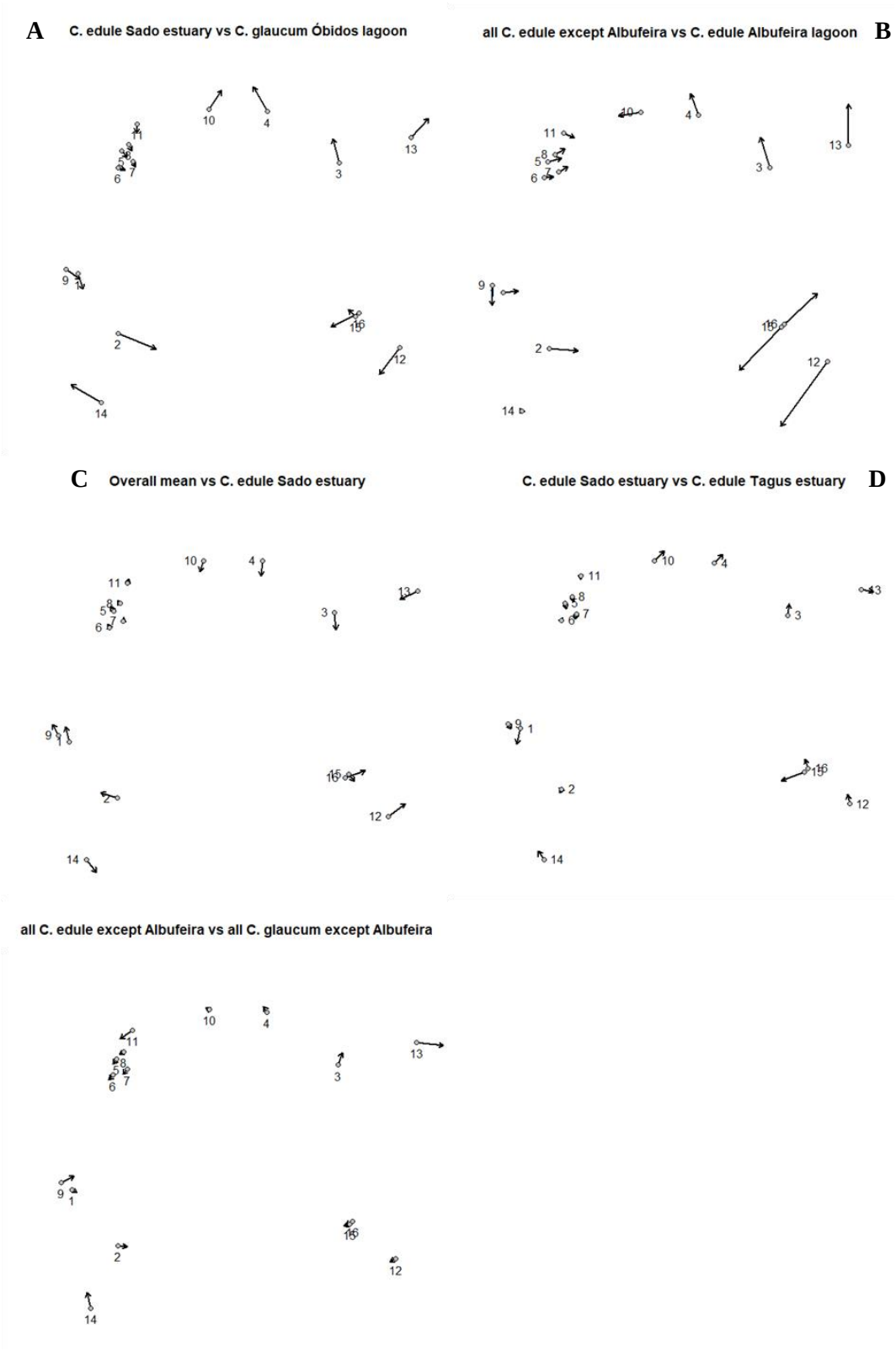


Fig. 3. 14 - Visual representation of the differences in the average landmark position between two groups with vector length multiplied by two. A - Landmark variation between the average landmark configurations of the *Cerastoderma edule* from the Sado estuary and *Cerastoderma glaucum* from the Óbidos lagoon. B - Landmark variation between the average landmark configurations of the overall mean of all *Cerastoderma edule* except from the Albufeira lagoon versus those specimens. C - Landmark variation between the average landmark configurations of the overall mean and the Sado *estuary's Cerastoderma edule*. D - Landmark variation between the average landmark configurations of the Sado and Tagus estuaries' *Cerastoderma edule*. E - Landmark variation between the average landmark configurations of the overall mean for each species and all aquatic systems except the Albufeira lagoon.

The trajectory analysis (TA) showed that the path distances (magnitude) that were statistically significantly different were: the Tagus estuary and the Albufeira lagoon (d-value = 0.023 and P-value = 0.006), and the Albufeira lagoon and the Óbidos lagoon (d-value = 0.031 and P-value = 0.003) (Fig. 3.15). Additionally, the trajectory correlations *i.e.* angular differences between trajectory principal axes that were statistically significantly different were: the Sado estuary and the Albufeira lagoon (angle-value = 95.876 and P-value = 0.001), the Sado estuary and Ria de Aveiro (angle-value = 82.927 and P-value = 0.004), the Tagus estuary and the Albufeira lagoon (angle-value = 114.207 and P-value = 0.001), the Tagus estuary and Ria de Aveiro (angle-value = 103.311 and P-value = 0.001), and the Albufeira lagoon and the Óbidos lagoon (angle-value = 71.845 and P-value = 0.021) (Fig. 3.15).

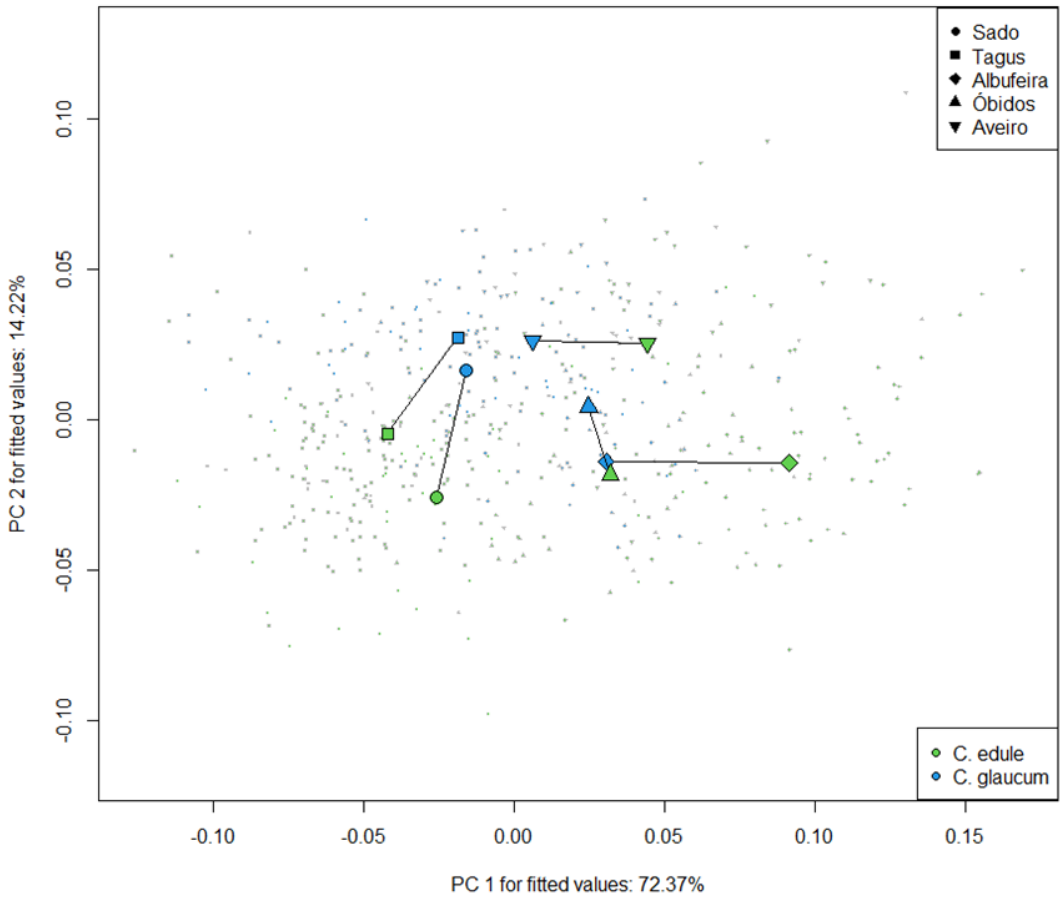


Fig. 3. 15 - Visual representation of the results of the trajectory analysis, where aquatic systems are trajectories and species are trajectory points.

Results of morphological disparity (*i.e.*, diversity of morphologies) (MD) using the mathematical descriptor (Csize-based) showed that *C. edule* from Albufeira was the group that displayed a bigger morphological variability compared to the other groups, and *C. glaucum* from Óbidos the smallest (Fig. 3.16). *C. edule* from Ria de Aveiro and from the Óbidos lagoon yielded similar MD (<0.001 difference) (Fig. 3.16). Additionally, all *C. glaucum* had smaller MD than their *C. edule* counterparts (Fig. 3.16). Furthermore, the only *C. glaucum* that showed a bigger MD than a group of *C. edule* (from the Tagus estuary) was from Ria de Aveiro (Fig. 3.16).

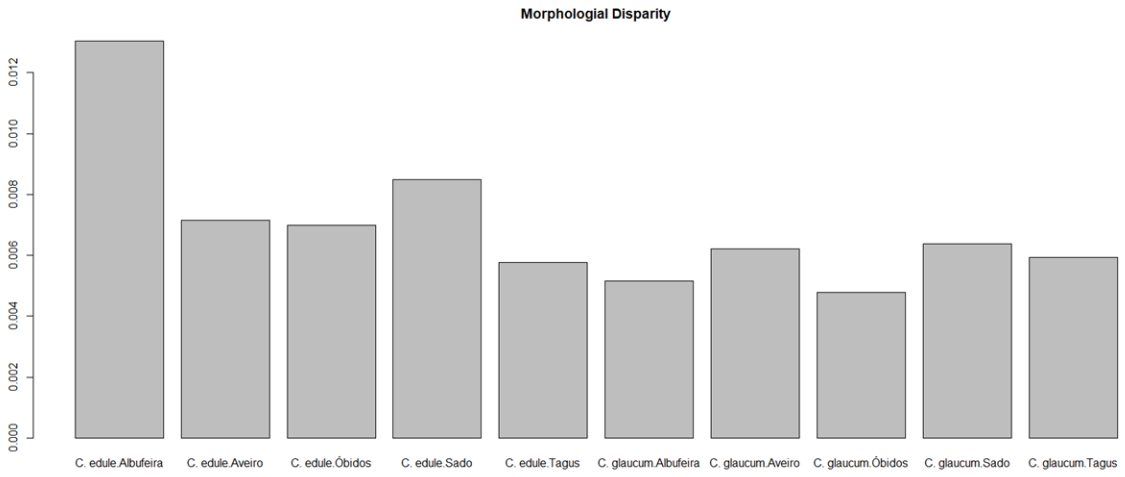


Fig. 3. 16 - Visual representation of the morphological disparity displaying Procrustes variance calculated from geometric morphometric landmark data (Centroid size based) as mathematical descriptor (first descriptor).

Using the model-based descriptor (second descriptor) to calculate MD, *C. edule* from Ria de Aveiro displayed the biggest MD compared to the other groups, while *C. glaucum* from Albufeira showed the smallest MD (Fig. 3.17). *C. edule* from the Óbidos lagoon and the Sado and Tagus estuaries seem to have similar MD (<0.001 difference) (Fig. 3.17). The aquatic system with the closest MD between species was the Tagus estuary (Fig. 3.17). Furthermore, all *C. glaucum* (except in the Tagus estuary's) presented smaller MD than all *C. edule* and all *C. glaucum* had smaller MD than their *C. edule* counterparts (Fig. 3.17).

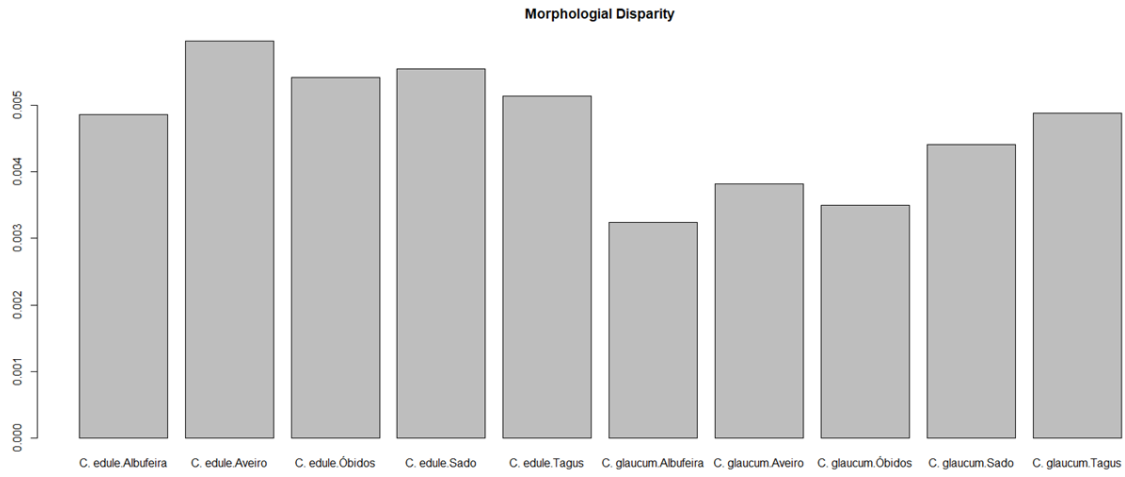


Fig. 3. 17 - Visual representation of the morphological disparity displaying Procrustes variance calculated using the final model (second descriptor).

The Canonical variates analysis (CVA) using the LM data classified aquatic systems with an accuracy of 83.13% and from the Tagus and Sado estuaries to Ria de Aveiro to the Óbidos lagoon, and to the Albufeira lagoon alongside the first canonical axis (62.6%) (Fig. 3.18). Alongside the second canonical axis (20.6%) systems were classified from the Albufeira lagoon and the Tagus and Sado estuaries to the Óbidos lagoon, and to Ria de Aveiro (Fig. 3.18). Class error for Ria de Aveiro was equal

to 18.18%, for the Óbidos lagoon was equal to 10.00%, for the Tagus estuary was equal to 19.23%, for the Albufeira lagoon was equal to 11.77%, and for the Sado estuary was equal to 18.75% (Tab. 3.3).

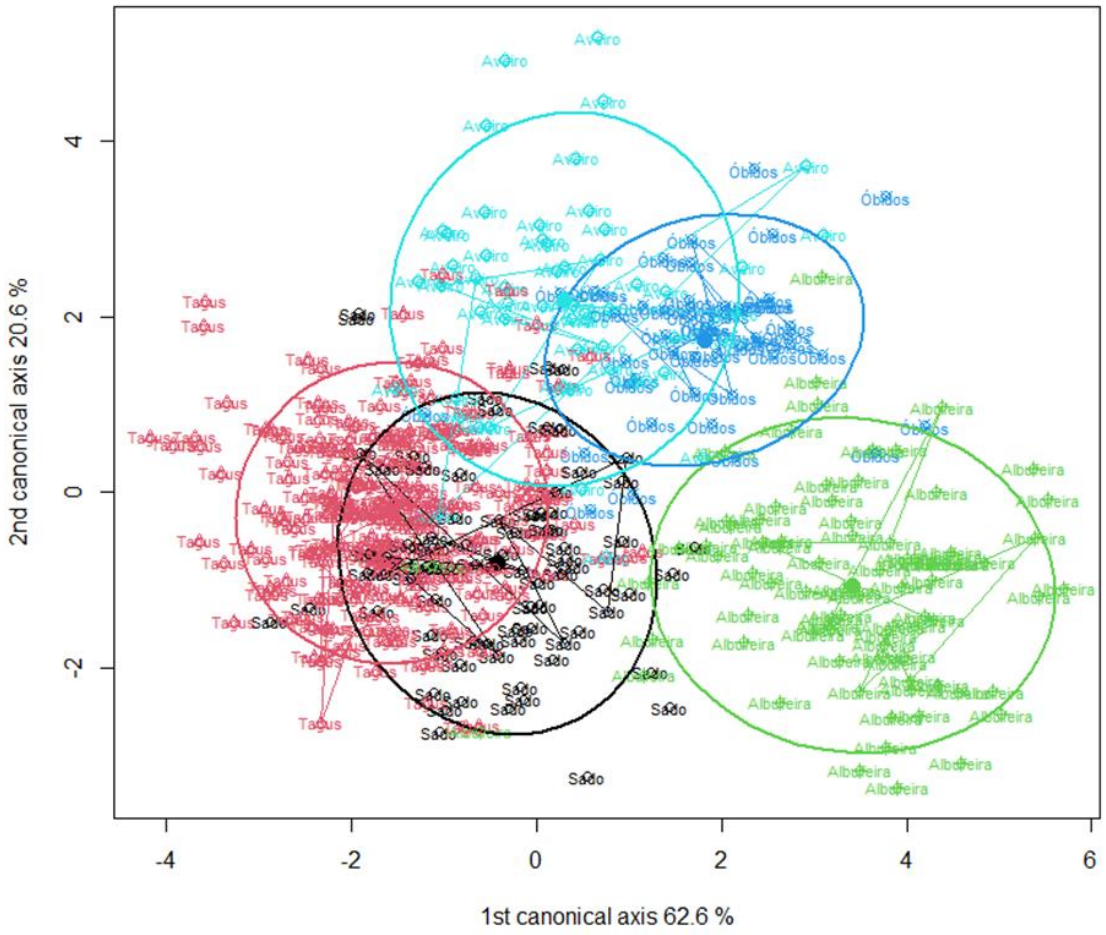


Fig. 3. 18 – Visual representation of the canonical variates analysis (CVA) based on landmark data for the five systems (Ria de Aveiro, the Óbidos lagoon, the Tagus estuary, the Albufeira lagoon, and the Sado estuary).

Tab. 3. 3 Confusion matrix of the canonical variates analysis based on landmark data for the five aquatic systems (Ria de Aveiro, the Óbidos lagoon, the Tagus estuary, the Albufeira lagoon, and the Sado estuary) with class error (wrong classifications/predictions divided by the total of observed specimens for that specific system) in percentage (for the cross-validated classification results in frequency).

Observed \ Predicted	Sado	Tagus	Albufeira	Óbidos	Aveiro	Outliers	Class error in %
Sado	65	10	0	0	4	1	18.750
Tagus	35	189	0	2	7	1	19.233
Albufeira	4	0	75	6	0	0	11.765
Óbidos	0	2	2	45	1	0	10.000
Aveiro	4	3	0	3	45	0	18.182

The random forest (RF) classifier built for species showed an accuracy (PCC - probability of correct classification) of 88.81%. Class error for *C. edule* was equal to 3.96% and for *C. glaucum* was equal to 28.57% (Tab. 3.4).

Tab. 3. 4 - Confusion matrix of Random forest classifier built for species (*Cerastoderma edule* and *Cerastoderma glaucum*) with class error (wrong predictions divided by the total of observed specimens for that specific species) in percentage calculated for test/validation set.

Observed \ Predicted	*C. edule*	*C. glaucum*	Class error in %
C. edule	97	4	3.960
C. glaucum	12	30	28.571

The RF classifier built for aquatic systems showed an accuracy (PCC) of 74.13%. Class error for Ria de Aveiro was equal to 46.67%, for the Óbidos lagoon was equal to 56.25%, for the Tagus estuary was equal to 4.84%, for the Albufeira lagoon was equal to 16.00%, and for the Sado estuary was equal to 56.00% (Tab. 3.5).

Tab. 3. 5 - Confusion matrix of Random forest classifier built for aquatic systems (the Sado estuary, the Tagus estuary, the Albufeira lagoon, the Óbidos lagoon, and Ria de Aveiro) with class error (total of wrong predictions divided by the total of observed specimens for that specific system) in percentage calculated for test/validation set.

Observed \ Predicted	Sado	Tagus	Albufeira	Óbidos	Aveiro	Class error in %
Sado	11	13	0	0	1	56.000
Tagus	2	59	1	0	0	4.839
Albufeira	0	1	21	2	1	16.000
Óbidos	0	5	2	7	2	56.250
Aveiro	0	6	1	0	8	46.667

The RF classifier built for species and aquatic systems showed an accuracy (PCC) of 67.13% (Tab. 3.6). Class error for *C. edule* (*C.e*) from Ria de Aveiro was equal to 45.46%, for *C. edule* (*C.e*) from the Óbidos lagoon was equal to 41.67%, for *C. edule* (*C.e*) from the Tagus estuary was equal to 6.97%, for *C. edule* (*C.e*) from the Albufeira lagoon was equal to 4.76%, and for *C. edule* (*C.e*) from the Sado estuary was equal to 50.00% (Tab 3.6). Class error for *C. glaucum* (*C.g*) from Ria de Aveiro was equal to 75.00%, for *C. glaucum* (*C.g*) from the Óbidos lagoon was equal to 100%, for *C. glaucum* (*C.g*) from the Tagus estuary was equal to 31.58%, for *C. glaucum* (*C.g*) from the Albufeira lagoon was equal to 100.00%, and for *C. glaucum* (*C.g*) from the Sado estuary was equal to 81.82% (Tab. 3.6).

Tab. 3. 6 - Confusion matrix of Random forest classifier built for species and aquatic systems (*Cerastoderma edule* (*C.e*) from the Sado estuary, *Cerastoderma edule* (*C.e*) from the Tagus estuary, *Cerastoderma edule* (*C.e*) from the Albufeira lagoon, *Cerastoderma edule* (*C.e*) from the Óbidos lagoon, *Cerastoderma edule* (*C.e*) from Ria de Aveiro, *Cerastoderma glaucum* (*C.g*) from the Sado estuary, *Cerastoderma glaucum* (*C.g*) from the Tagus estuary, *Cerastoderma glaucum* (*C.g*) from the Albufeira lagoon, *Cerastoderma glaucum* (*C.g*) from the Óbidos lagoon, and *Cerastoderma glaucum* (*C.g*) from Ria de Aveiro) with class error (total of wrong predictions divided by the total of observed specimens in specific combination of species and systems) in percentage calculated for test/validation set.

Observed \ Predicted	*C. e* Albufeira	*C. e* Aveiro	*C. e* Óbidos	*C. e* Sado	*C. e* Tagus	*C. g* Albufeira	*C. g* Aveiro	*C. g* Óbidos	*C. g* Sado	*C. g* Tagus	Class error in %
C. edule Albufeira	20	1	0	0	0	0	0	0	0	0	4.762
C. edule Aveiro	1	6	0	0	4	0	0	0	0	0	45.455
C. edule Óbidos	1	2	7	0	2	0	0	0	0	0	41.667
C. edule Sado	0	0	0	7	6	0	0	0	1	0	50.000
C. edule Tagus	1	0	0	1	40	0	0	0	0	1	6.977
C. glaucum Albufeira	0	1	1	0	2	0	0	0	0	0	100.000
C. glaucum Aveiro	0	1	0	0	0	0	1	0	0	2	75.000
C. glaucum Óbidos	0	0	0	0	1	1	0	0	0	2	100.000
C. glaucum Sado	0	1	0	0	1	0	0	0	2	7	81.818
C. glaucum Tagus	0	2	0	0	4	0	0	0	0	13	31.579

4. Discussion

Cerastoderma edule and *Cerastoderma glaucum* are two sibling species of cockles that co-exist in most areas [12, 13]. The cockles' wide distribution and considerable morphological similarities challenges population discrimination [1, 11, 12, 23, 46, 48, 49]. This fact negatively affects the traceability of this genus [89], which basically revolves around fatty acid profiles or geochemical profiles (*e.g.*, [202, 203]). This study used a quick and inexpensive technique based on geometric morphometrics (GM) methods and successfully discriminated between two species (88.81% accuracy), as well as the aquatic systems *i.e.*, the provenance of the specimens (83.13% accuracy for the Canonical variates analysis (CVA) and 74.13 % accuracy for Random forest (RF)), thus being a promising traceability method. Shape analysis using landmark-based GM methods was used as traceability technique, not requiring highly skilled technicians nor costly reagents and equipment.

4.1 Linear morphometrics

Measurements like height, length, and width fall into the category of linear morphometrics [204]. They are easy to collect, low cost and simple to understand and interpret [204]. However, such method is under the influence of sample strategy, which in this work was not similar in all locations, thus, it was a priority to have size-free methods, such as roundness and landmark-based GM [68, 204].

Samples from the Sado estuary were collected by hand in the shallow intertidal areas, while professional harvesters collected specimens at the subtidal areas of Ria de Aveiro, the Albufeira and Óbidos lagoons and samples from the Tagus estuary were collected using both approaches. Such differences in sampling methods could influence the results of this study if only linear morphometrics had been used, and therefore cannot be used to draw conclusions on the population size variation between aquatic systems. For linear morphometrics, the size of the sample could be an additional factor to consider, since the number of specimens collected in each coastal system was unbalanced, with lesser specimens collected in the Óbidos lagoon and a higher number of specimens collected in the Tagus estuary.

The study of linear morphometrics and shell roundness of the cockles species *C. edule* and *C. glaucum* showed that specimens from Ria de Aveiro had rounder and bigger shells on average for both species, compared to the other coastal systems, whereas specimens from the Tagus estuary had smaller shells. Cockles' population structure is often affected by high mortality events [205]. These might be associated to food limitation; density; oxygen depletion and organic loadings; temperature and salinity; parasites, pathogens and commensals; toxicants and other persistent pollutants; predation, and changes in sediment, suspended solids, topography and bathymetry [11-13, 28, 205]. Population size structure might also be related to fishing pressure [28].

The measurements of *C. edule* from the Sado estuary and the Albufeira and Óbidos lagoons were very similar, especially the Sado estuary and the Albufeira lagoon, although different sampling methods were used. However, the same does not happen with *C. glaucum*, even if the sampling method was the same within each system. This difference between species could be related to the size of the sample, as *C. glaucum* specimens were always less abundant than their *C. edule* counterparts.

Although the Tagus estuary is a very variable aquatic system in terms of environmental conditions [206], compared to the other systems, its specimens had lower roundness variability. Still, there was

some variability in the shell roundness of *C. edule* at different stations of the Tagus estuary, with three major groups identified: upstream (stations 8, 12, and 14); downstream (stations 26, 29, 30, 31, and at the beach); and intermediate (stations 23, 23A, 23B, and 23C). The upstream stations have a lower salinity influence (~31 for upstream stations versus ~33 for the other stations (NIPOGES, unpublished data)), and are subject to more turbulence. Despite downstream and intermediate stations being located closer to the ocean, these are mostly sheltered from wave influence in the Tagus estuary. The evidently higher variability in stations 26 and 29, especially in station 29, can be related to the size of the sample. Station 29 had the most discrepancies between shells and had almost three times fewer specimens sampled than station 23B, which is the station with fewer specimens after station 29. Stations 23, 23A, 23B, and 23C suffer less wave influence since they are located in an inner bay, which could explain the fact of having shells with lower mean roundness than other stations.

Despite roundness being a useful method for sample size-free observations, it has multiple statistical problems associated with ratios [207]. Roundness is a ratio (width divided by length) and, according to literature, ratios are not the most accurate representation of biological data and have complex mathematical properties [207]. Adding to the fact that factors affecting length and width might have a repercussion on the roundness results.

Linear measurements can allow easy discrimination between different morphologies without the laborious tasks of collecting many specimens, digitizing landmarks (LMs), and performing multivariate statistical analyses [83]. However, as mentioned before, linear morphometrics are rarely sufficient to quantify shape information about an organism [83, 86, 98, 100-104], reinforcing the need to use GM approaches.

4.2 Geometric morphometrics

Compared to classic morphometric methods, GM methods are more effective to capture information about the shape of an organism and have a greater power to discriminate between species [83, 109, 117]. GM methods are, not only successfully applied to various studies, but also a valuable approach to analyse shape variation in bivalves (*e.g.*, [61, 79, 82, 83, 86, 192, 208]). Furthermore, in spite of its relative simplicity, the LM methodology applied in this study has shown to be an effective and accurate method to study shape changes in various contexts (*e.g.*, [61, 68, 82-84, 113, 116, 192, 208]). In this work, the analysis of shape changes between centroid size (Csize), species, aquatic system, and the respective interactions, using GM based on LMs applied to bivalve shells, found significant differences between the shape of the two species and between different aquatic systems. The best model had a common allometric effect for both species and coastal systems and displayed shape changes between species and systems. This analysis exposed significant shape differences using only GM data. Such accomplishment would not be achievable using only linear morphometrics, as it was found in previous literature in general or using other species [83, 109, 117].

On the downside, the application of landmark-based GM approaches to some biological structures can be challenging, especially in cases where there is ambiguity in defining homologous points [98]. For example, some of the specimens may have had multiple muscle scars, thus failing the assumption of LMs homology for not being comparable among specimens [86, 109]. This can be perceived as an obstacle to using LM methods, despite its use not truly requiring biological homology [98, 209]). Moreover, even though all LMs were marked by the same observer, human bias is always a constraint

to be considered. Nevertheless, landmarks-based methods contain the richest geometric information available in any object, in spite of the potential subjectivity in the location [109].

The variation of Csize between species and aquatic system showed that specimens from Ria de Aveiro were larger, which was also confirmed by the linear measurements. Likewise, the shells from the Tagus estuary were the smallest of all aquatic coastal systems, and *C. edule* from the Óbidos lagoon, the Sado estuary, and the Albufeira lagoon had similar sizes. As mentioned before, although size differences might be related to sampling strategies, it could also be related to the cockles' population structure which is often affected by high mortality events [205]. This might also be related to fishing pressure, which is increasing in Ria de Aveiro [28], affecting the size structure of *Cerastoderma* spp., which population has been reported to be decreasing [153, 154]. Knowing this, it would be expected that in Ria de Aveiro's specimens would be smaller. However, this was not observed in the current study, with the biggest specimens collected in Ria de Aveiro, possibly associated with lower competition with other sympatric species, since the average bivalve density in Ria de Aveiro is near 10 specimens/m^2, while an average of 60 specimens/m^2 was recorded in the Tagus estuary, supporting the occurrence of bigger specimens in the former system (NIPOGES, unpublished data). Bivalve landings in the Tagus estuary indicate a much less relevant cockle harvesting activity (no landings between 2015 and 2020 in the Tagus estuary versus 98,947 tonnes in Ria de Aveiro for the same period) [54, 163] (Source: *Direção-Geral de Recursos Naturais, Segurança e Serviços Marítimos* - DGRM) and recent monitoring data showed a good recruitment of the species in the Tagus estuary, with a high number of specimens under the minimum landing size (25 mm in length) (NIPOGES, unpublished data). The higher density of bivalves might be associated with food limitation, therefore explaining the smaller sized shells collected in this study.

Shape changes during growth (allometry) occur due to some parts of the body growing more than others (disproportionate growth) [210]. The variability around each LM position relative to the consensus configuration (mean location of each LM) provided a better understanding of the shape variation of the whole dataset. In this study, LMs numbers 13, 14, 15, and 16 showed notably more variation/dispersion when compared to others. In bivalve shells, parts farther apart from the umbo develop more than areas around the teeth [6], which may partially explain the higher variability observed in the LMs' locations for these areas.

The variation observed in LMs #15 and #16 (including the inversion of its position), both placed on the pallial line, was the biggest and most evident, not only through the variability around the consensus configuration, but also throughout this study (*e.g.*, visualisation of variation and covariation distributed within PC1). This inversion might be related to the Type of LM (Type I for #15 and pseudo-landmark for #16). The variation observed in LM #15, was considered an accurate way of discrimination, since it is a Type I LM, therefore, a representation of a true biological homology [109, 113]. However, since LM #16 is a pseudo-landmark and is constructed from an imaginary line linking two other LMs (#11 and #12) and the pallial line, any noticeable variation in those two LMs could be and was reflected in LM #16 and, therefore, not directly linked to group differentiation. This highlights the need to have LMs that reflect primary homologies (Type I and Type II), as much as possible, as it has been referred to in previous literature [109, 113].

Additionally, this inversion might occur if a specimen has fewer and broader ribs [8, 12, 171], and, therefore, the junction of the 10th rib and the pallial line where LM #15 is placed, would be farther apart from the top of the shell, thus making the position of LM #15 exceed the location of LM #16. The number of ribs has been reported to be correlated with salinity [8, 12, 39, 171]. This hypothesis should

be verified in future works, as it can become a way to easily and accurately discriminate groups along the salinity gradient of aquatic systems.

The Principal component analysis (PCA) conducted to visualise the shape change patterns in the dataset found in the linear model, showed that PC1 (explaining 45.3% of the variation) appeared to be linked to the inversion of LMs #15 and #16, while in previous works represented variation related with allometry, the main source of shape change in animals (*e.g.*, [94, 98, 109, 191]). This result is in accordance with the absence of interaction between species and systems and Csize (linear model results). While LM #16, is most probably not related to group differentiation, LM #15, a Type I LM, has an ecologically and biologically relevant influence and, as seen throughout this study, might be a way of discriminating between groups. The relative position of #16/#12 may be related with shell symmetry around the main axis (umbo to LM #12, perpendicular to the anterior-posterior axis). A body is symmetric with respect to a given plane if it is carried into itself by reflection in said plane [211]. In this case, the body (right valve) can be reflected through its midplane, the plane separating the body's two similar halves [212]. Bilateral symmetry in bivalves relates to right/left valve symmetry [6, 62, 213], but there is also a symmetry associated within a valve (the right one in this case), as it is being referred to here. Moreover, it has been described in literature that symmetry (that being bilateral or within valves) in bivalves is related to the thickness of the shell and is a feature that influences burrowing capacity [6]. On the other hand, the symmetry of *Pisidium subtruncatum* (pill clam) is influenced by sediment type [73]. In this work, it was observed that the closer the LM #16 was to the midplane of the valve (umbo to LM #12), the more symmetric the valve was and the less pointed. This might help differentiate between systems or sediment type. The relation between bivalve shell symmetry, both between and within vales and sediment type versus burrowing capacity should be further investigated in future works.

Moreover, considering that allometry might not be linked to PC1 because of the different biological accuracies in LMs #15 and #16, the next, logical association would be with PC2 (explaining 13.7% of the variation), *i.e.*, the next component in the PCA explaining more variation [94, 130]. As expected, while looking at the representation of the two species within the configuration set, there is a distinct separation between the two species, mostly linked to the second dimension.

Additionally, when observing differences in the average LM position between both species, the inversion of LMs numbers 15/16 does not occur, adding to the fact that this shape change is not linked to differences between species.

Differences between the two bivalve species were found to be related to the largest changes in LM numbers 9, 11, 13, and 14 positions. LMs #13 and #14 represent the extremities of the shell, LM #11 the tip of the ligament (the area that connects the two shells) and LM #9 located on the left end of the shell hinge. Since all these LMs are Type III (except for number 9, which is Type II), they provide less reliable shape information than Type I LMs since their position is more subjective [113]. The position variation of LMs 13 and 14 showed that *C. glaucum* shells were more pointed on average (#13 varied downwards the shell, away from the umbo while #14 varied upwards the shell, closer to the umbo). Variability in LM #11 position, as shown by the shape change vectors between the two species, is mostly horizontal variation (direction left-right on the shell *i.e.*, anterior-posterior axis). This might be related to the fact that *C. glaucum* specimens are known to have thicker ligaments than *C. edule* specimens [8, 171]. Additionally, differences in LM #9, observed in the shape change vector, were directed towards the umbo and shell centre. This indicates that the hinge is shorter and located more towards the interior of the shell for *C. glaucum*, which has never been reported before, to the author's best knowledge. Further emphasising that GM methods can detect subtle changes in shape that would not be seen using

other methodologies [61, 83, 86, 193]. Aside from this, most findings mentioned before, support *C. glaucum*'s shell being thicker, as the changes in LM positions were mostly facing inwards, which is in agreement with the thicker shells reported in previous works for this species [6, 12, 171]. Further works are required to test this hypothesis. Moreover, most changes in the LM positions between the two species (from LM #9, #11, #13, and #14) were possible to interpret in face of the geometric properties of the shell and previous knowledge of the species.

Moreover, the variance of some components of the PCA is considered negligible, and the variation in the data can be described by the few first (independent) principal components (PCs) [94], hence the use of only the first four PCs. The PCA is an informative and useful representation, however, it might not display bivariate plots of this kind of data without loss of information, and thus these projections can be misleading [201]. This fact justifies the use of a linear model of the GM data for an analysis of shape as the most accurate depiction of said data. In this work, PCA and its PCs were simply used to visualise the variance detected in the data as it was also done in previous literature [119, 127, 132, 191].

RF is one of the best-performing learning algorithms and most successful methods to handle classification problems [138-140, 147-150]. The RF classifier for species showed that it was possible to predict species differentiation with a high accuracy of 88.81% (for test/validation set), with landmark-based GM data. Furthermore, a very low percentage of *C. edule* specimens were wrongly classified (class error of only 3.96%, the lowest in this work). These results suggest that species differentiation through machine learning methods and using GM data can easily be done. And, since RF is one of the most successful classification methods (*e.g.*, [141, 144, 147-150]), these results provide good prospects to distinguish *C. edule* and *C. glaucum* using this technique.

Although shape differences between species were significant, shape changes across aquatic systems were larger than between species. *C. edule* and *C. glaucum* are two very similar species, and in fact, their identification using external morphological aspects is dubious [11, 12]. A morphotype is a characteristic morphological form of an organism or a particular group of organisms belonging to the same taxa [214]. They are common in bivalves [77, 78, 82, 103, 117] and, are sometimes used to categorise sub-populations (*e.g.*, [103]). *Cerastoderma* spp. have shown distinct morphotypes in previous works and have a history of issues in species differentiation [215, 216]. For example, *C. glaucum* showed two morphotypes that were once considered to be different species, as concluded by genetic evidence, that named it *Cerastoderma* spp. species complex (*C. lamarcki* and *C. glaucum* alongside with *C. edule*) (*e.g.*, [215; 216]). However, recently, *C. lamarcki* has been accepted as synonym of *C. glaucum*, and therefore, both are currently considered to be one species, with two or more morphotypes [217]. This demonstrates that these species have a high capacity to produce morphotypes. GM only identifies shape changes, and does not provide any evidence on species differentiation *per se*. However, as in this particular case, the main objective was to differentiate between aquatic systems, and not between species, since both species are harvested and traded with a single identity (*Cerastoderma* spp.). These results do not support the discrimination between *C. edule* and *C. glaucum* as different species or simply different morphotypes. Future works involving genetic analysis might be required to specifically address this aspect.

As mentioned earlier in this study, shape changes across aquatic systems were larger than between species. This might also be evidence that the influence of the environmental conditions on the shell shape is larger than differences between species. Cockles can live up to 10 years in some habitats but more commonly, up to six years or down to two, during the occurrence of high mortality events [1, 46]. During that period the shell can accumulate environmental information in the shell. Bivalve shells have

been used as environmental proxies or indicators (*e.g.*, [218-220]). For example, the shell of *Arctica islandica* has been shown to be an excellent proxy to reconstruct paleoclimates (paleothermometer *i.e.*, determining past temperatures) [221-223]. Another example is the shell of the dog cockle *Glycymeris glycymeris* which is a geochemical proxy to study the Iberian Upwelling System [224, 225]. Those studies show the sensitivity of bivalves to environmental changes, which are recorded in the shell.

The CVA showed an accuracy of 83.13% (using the cross-validated classification results in frequency) in the classification of the specimens according to the aquatic systems where they have been collected. This classification showed a relationship between the shell shape gradient and the aquatic system size from the biggest (the Tagus estuary) to the smallest (the Albufeira lagoon), along the first canonical axis. A latitudinal gradient seemed to be associated with the second canonical axis, from south (the Albufeira lagoon and both estuaries, which have small latitudinal difference (about 30 km^2)) to the northern ones (the Óbidos lagoon and then Ria de Aveiro, further north). However, as mentioned before, it has been argued that ordinations based on CVA do not preserve the Procrustes geometry [133]. Despite this, such methodology has been often used in previous works (*e.g.*, [106, 119, 129, 136]), and thus, it was performed in order to provide comparison grounds with past and future studies.

The RF classifier (for aquatic systems only) showed that landmark-based GM data supported the identification of the aquatic system where the specimens were collected with a high accuracy of 74.13% (using the validation set). Despite shape differences across aquatic systems being larger than between species, RF results suggest that systems differentiation through machine learning methods and using GM data can easily be done but is harder than species differentiation. This might be due to the fact that there are more classes associated with the variation (five systems instead of two species). Moreover, as mentioned before, RF is one of the most successful classification methods (*e.g.*, [141, 144, 147-150]), therefore, the results of this study provide good prospects to distinguish between aquatic systems using this technique.

An accuracy of 67.13% (using the validation set) was obtained when aquatic systems and species were considered in the RF analysis (only a 7.00% difference from the previous case). The confusion matrix showed that the RF algorithm always mistook between systems instead of species within a system, except for *C. glaucum* from the Tagus estuary, although shape changes across aquatic systems were larger than between species. Predicting species and systems accurately was less efficient than only one of these two factors (less levels/classes), although these results are promising to distinguish between species and aquatic systems using this technique.

Ria de Aveiro was described as morphologically different from other systems through the trajectory analysis (TA) and its trajectory correlations. It was also the system showing the largest specimens (results from linear morphometrics and from the analysis of Csize variation). In spite of this, there was always a size overlap between systems, and thus, it was possible to compare the shape of the shell between systems without linking it directly to the size of the shell in that specific system. PCA's visualisation showed that this system's shape variation was observed on the extremity of PC2 (the PC which explains the most variation after PC1). This group seemed to have the most differentiated shapes, therefore, it would be expected that this group's shell shape would be easy to distinguish from other groups. However, when looking at the RF's confusion matrix for systems, Ria de Aveiro had the third highest class error (46.67%). This might be due to Ria de Aveiro's shells having bigger shape variability, as displayed in the morphological disparity (MD) results calculated with the second descriptor for *C. edule* (described further in the section about MD). It is important to mention that these results might be related to sample size, as it will be further explained later.

Differences in the average LM position between the overall mean versus Ria de Aveiro showed variation in all LMs except for #4 and #10 placed on the posterior muscle scar and the hinge, respectively. Furthermore, vectors displayed from LMs #1, #2 (Type I), and #9 (Type II) showed that, on average, Ria de Aveiro's specimens have their anterior shell area more distant from the umbo than other specimens studied, suggesting a way to discriminate these shells from other systems. Moreover, an unexpected variation was associated with LM #15 and #16, both placed on the pallial line, with vectors directing away from each other, when, in most parts of this study, these vectors inverted. This described that the 10th rib of Ria de Aveiro's specimens was located more towards the umbo than specimens from the other systems, possibly meaning that Ria de Aveiro's *Cerastoderma* spp. had narrower ribs compared to the average of the other systems. Since salinity might influence rib size, number, and growth [8, 34, 39, 171], having narrower ribs might be related to different salinity values for Ria de Aveiro. Additionally, [226], showed that *C. edule*'s distribution in Ria de Aveiro is heavily influenced by freshwater discharge, which decreases salinity. In fact, multiple factors are known to influence salinity in estuaries, namely the extent of the dry season, river discharge, tidal cycle, geomorphologic configuration, etc. [151, 227]. In recent years, Ria de Aveiro hydrodynamics have been changing due to modifications in geomorphologic configuration and are, to a large extent, a consequence of anthropogenic activities on the lagoon morphodynamics [227]. Such factors might also be impacting bivalve shell shape, and therefore, our results.

Differences in the average LM position between the Albufeira lagoon and Ria de Aveiro (both considered the most different systems from the others) showed that all LMs differed except #15 (Type I LM, therefore has a true biological homology origin [109, 113]). As mentioned earlier, the position of this LM relatively to #16 (which can be considered as a reference point), is an indirect indicator of the ribs' size and/or number, which would be one of the greatest differences (mostly seen in PC1). However, these differences are not notable between the Albufeira and Ria de Aveiro, meaning that there was no difference in the relative location of the 10th rib between these two groups. Thus, ribs of both species might grow in similar ways in these two zones, as opposed to the other systems. Since, as mentioned earlier, salinity might influence rib size and number [8, 12, 39, 171], having similar LM #15, might be related to being exposed to similar salinity values in these two systems. Between factors that influence salinity, tidal amplitude is the factor most resembling for these two systems (tidal amplitude of 0.6 - 3.2 m for Ria de Aveiro and 0.55 - 3.86 m for the Albufeira lagoon) [151, 227-229]. However, as mentioned before, hydrodynamics in Ria de Aveiro have been changing in recent years [227], and the same has been observed with the Albufeira lagoon's hydrodynamics [230]. Every year the connection to the ocean of the Albufeira lagoon is artificially open to avoid eutrophication [230]. The opening's location changes over the years and is done with mechanical resources such as backhoe loaders or mechanical shovels [230]. This anthropogenic geomorphologic configuration modification leads to recurrent hydrodynamics changes in the Albufeira lagoon, for example, the discharge increases, influencing salinity and sediment settlement [230]. Both salinity and sediment are known to influence bivalves' shell morphology [6]. Therefore, due to these hydrodynamic changes, it is difficult to discuss similarities between both systems. The absence of variation in LM #15 between these two groups might also explain why they were so different from the others considering the importance LM #15 had on PC1. Aside from this, all LMs from the whole upper area of the shell (around the umbo area) showed an expansion of the shell for Ria de Aveiro's specimens compared to the ones from the Albufeira lagoon. LM #3 (Type I) showed the highest variation between these two groups, followed by LMs #5 to #8 (Type II). LM #3 was placed on the anterior muscle scar, and here, it represented a longer muscle scar for Ria de Aveiro's specimens. LM #5 to #8 were placed on the hinge teeth of the shell and, since vectors varied upwards the shell, it showed that Ria de Aveiro had shells with higher height. Furthermore, LM #5 to #8 tended

to not vary as much as the others throughout this analysis except for the case of those two systems, therefore, it could be a good approach to discriminate between both systems.

This variation in LM 5-8 was observed again when looking in more detail and comparing *C. glaucum* from Ria de Aveiro and *C. edule* from the Óbidos lagoon, the two groups on the extremities of PC2 (when observing the variation and covariation of both species and the five systems). Highlighting that it could serve as a good approach to discriminate both groups from each other. Furthermore, the RF's confusion matrix (for species and systems) showed that these two groups were never confused with each other. Still comparing *C. glaucum* from Ria de Aveiro and *C. edule* from the Óbidos lagoon, vectors LMs #1 and #3 (Type I), are the most accurate and secure method of differentiation between these two groups, despite the vectors being short [109, 113]. Their shape differences could reflect bigger shells for Ria de Aveiro' *C. glaucum* and is in accordance with earlier findings. However, with the length and direction of the vector from LM #9 (Type II), a longer hinge can be deduced, thus not corroborating the hypothesis of smaller shells and earlier conclusions about *C. glaucum* having a smaller hinge. Nonetheless, as mentioned before, observations from Type II LMs are not as biologically accurate as Type I's [109, 113].

Additionally, differences in the average LM position between *C. edule*'s overall mean versus Ria de Aveiro's *C. edule* (and earlier findings) showed variation in LMs #1 and #2 (among others) placed on the anterior muscle scar, its vector directing downwards the shell away from the umbo. Therefore, a good method to differentiate Ria de Aveiro's *C. edule* from other *C. edule* would be identifying which specimens had an anterior muscle scar placed downwards the shell, farther from the umbo. The differences in the average LM position between the overall mean and Ria de Aveiro's *C. glaucum* and the one versus Ria de Aveiro's *C. edule* presented were presumably associated with species differentiation. Those differences were congruent throughout this study and were reflected in a thicker ligament for *C. glaucum*, as expected for this species [6, 12].

In the RF classification for systems, the Óbidos lagoon had the highest class error, with 56.25%, being the hardest system to discriminate from others.

Differences in the average LM position between the overall mean and the Óbidos lagoon's *C. glaucum*, vectors from LMs numbers 2-4 (Type I, reflecting primary homologies [109, 113]) suggested that the posterior muscle scar was closer to the umbo than other groups and that the specimens had longer muscle scars on average, especially the anterior, indication that it might be an excellent method to discriminate this group from others. Furthermore, the lack of variation in LMs 5 to 8 might be a good method to differentiate this between system's species, since earlier findings suggested that such variation could be linked to this system's *C. edule*. However, the absence of variation in LM #11 was unexpected since *C. glaucum* are known to have thicker ligaments [171], which was generally seen in this study. Furthermore, it was possible to observe that *C. edule* from the Óbidos lagoon and *C. glaucum* from the Albufeira lagoon were similar in shape. This might be related to the fact that the Albufeira lagoon's *C. glaucum*'s height, length, and width were the only measurements which did not follow the other systems' pattern of means (higher values on average than *C. edule* counterparts).

Moreover, when observing the differences in the average LM position between two similar groupings, the *C. glaucum* from the Sado and Tagus estuaries versus *C. edule* from the Óbidos lagoon and *C. glaucum* from the Albufeira lagoon, something unexpected was found. Despite LM number 16's variation being associated multiple times with LM #12, here, their vector directions were opposing, suggesting instead, an association with LM #11 (tip of the ligament). This and the fact that the estuaries displayed shorter anterior muscle scar (LM #1 and #2) hinted at a way to discriminate these groups from

others. It was also implied that the estuaries' specimens had thicker shells on average (LM numbers 5-14).

Despite the Tagus estuary not particularly standing out from the other systems throughout this study, the RF classifier showed the second lowest class error (4.84%) in this work for this system. This might be related to sample size, since the Tagus estuary had the biggest sample (almost 3.4 times bigger than the next big sample, the Albufeira lagoon), the classifier could get a lot of its predictions right just by guessing the Tagus estuary continuously. Looking at the RF classifier's confusion matrices for systems and for species and systems, the Sado estuary's shells were highly confused with the ones from the Tagus estuary, suggesting that these two systems have very similar shells. This was also observed through CVA and PCA but was not displayed with the linear morphometrics analysis, emphasising the advantage of the use of GM over linear morphometrics [83, 86, 102-104]. Such similarities might be related to similar geomorphological properties. *Cerastoderma* spp.'s shell is solely composed of aragonite [36], thus similar concentrations of this component (found in both estuaries [231]) could lead to similar shells in these two systems. Or the fact that the Tagus and Sado estuaries are the two biggest estuaries in Portugal and there is only a small difference in latitude between each (about 30 km^2) [160, 164]. However, changes in the hydrodynamics of both estuaries (mostly due to anthropogenic pressure) have been reported in multiple studies (*e.g.*, [206, 232-235]), therefore, it is difficult to speculate about which parameters might influence the results of this study. Throughout this study, both estuaries appeared to have very similar shaped shells except when analysing PC4 and *C. edule*'s shells. Therefore, despite possibly finding other ways to discriminate the Tagus estuary's *C. edule*, it was still useful to analyse the differences in the average LM position between the Sado and Tagus estuaries' *C. edule*. Unsurprisingly, most variation in average LM position was small. The most notable ones were reflecting a smaller posterior muscle scar (associated with LMs #3 and #4), a smaller shell (LM #12 pointing inwards the shell), a hinge further away from the umbo (LM #9), and broader ribs (associated with LM #15), all for the Tagus estuary's specimens, offering a way to discriminate *C. edule* from both systems despite them having similar shaped shells in general.

Moreover, something that was associated earlier with the Tagus estuary's *C. edule* was verified again when observing the differences in the average LM position between the overall mean versus the Tagus estuary's *C. edule*. That is, LM #16 and #12's vectors directing opposite ways and shorter anterior muscle scar (LM #1 and #2) was seen again with the differences in the average LM position between the overall mean versus the Tagus estuary's *C. edule* highlighting this as a good way to discriminate this group from others.

Additionally, when comparing *C. edule* specimens from the Albufeira lagoon and the Tagus estuary (the groups observed at the extremes of PC1), the most distinct vector was from LM #14 (on the extremity of the shell, Type III), where almost no variation was observed. Therefore, between the Tagus estuary and the Albufeira lagoon, it is possible, and probably easier, to differentiate species using LM #14. Furthermore, when looking at the RF's confusion matrix (for species and systems), the classifier mistook the shell of these groups only once, further displaying that these shells are very different from each other.

The Albufeira lagoon was described as morphologically different from the other systems through TA and its trajectories magnitudes and trajectory correlations. In addition, this system had the second lowest RF class error (16.00%), therefore emphasising the fact that the Albufeira lagoon is easily distinguishable from the other systems. However, the Albufeira lagoon did not particularly stand out from the other groups when looking at Csize variation and linear morphometrics. It was even mentioned

that this group had very similar values to other groups. With the PCA's visualisation, it was possible to see that this system's shape variation was observed on the extremity of PC2 (the PC which explains the most variation after PC1) and its specimens were isolated from other groups. Despite the Albufeira lagoon's *Cerastoderma* spp. having similar shell size compared to other systems, they seemed to have the most differentiated shapes (alongside Ria de Aveiro). Therefore, it would be expected that this group's shell shape would be easy to distinguish from the other groups, through GM, something that would not be possible only with linear morphometrics [83, 109, 117].

As mentioned earlier, LM #5 to #8 tended to not vary as much as the others throughout this analysis except for the case of Ria de Aveiro and the Albufeira lagoon and could be a good approach to discriminate between both systems. This was validated again when looking at differences in the average LM position between *C. glaucum* from the Sado and the Tagus estuary versus *C. edule* from the Óbidos lagoon and *C. glaucum* from the Albufeira lagoon.

The differences mentioned before about Ria de Aveiro and the Albufeira lagoon could be related to the fact that the Albufeira lagoon is a semi-enclosed lagoon, providing different environmental conditions for the growth of *Cerastoderma* spp. (*e.g.*, wave and air exposure, salinity, temperature) [168, 169]. However, the Óbidos lagoon, similarly to the Albufeira lagoon, is a semi-enclosed coastal lagoon, and while results from the variation of Csize and from linear measurements resemble the Albufeira lagoon's, results from PCA showed that these two systems' shells are different in shape, despite similar in size. This might be related to the fact that the Albufeira lagoon is deeper and is closed to the ocean for longer periods, creating stratification in the water column and, with lower renewal by exchanges with the seawater [236]. Furthermore, the Óbidos lagoon has a higher salinity amplitude (from 2.1 to 38.5 versus 14.3 to 36.3 for Albufeira) and smaller tidal amplitude (1.0 to 2.0 m versus 0.55 to 3.86 m for Albufeira) [168, 229]. As mentioned earlier, salinity is a crucial factor in the growth of cockles (including its ribs) [8, 12, 34, 39, 171] and multiple factors can influence salinity (for instance, tidal cycles) [227]. The recent hydrodynamics of both lagoons have been studied [230, 237]. As mentioned earlier, for the Albufeira lagoon, recurrent hydrodynamics changes happen every year and might affect bivalves' shell morphology [230]. For the Óbidos lagoon, similar hydrodynamics changes and consequences happen due to dredging the channels [237]. For this reason, it is difficult to discuss the observed differences between lagoons. In addition, since groups from both lagoons were the only ones with a positive variation and covariation seen in PC1 and PC2 for all ten groups (*i.e.*, two species in five aquatic systems), the analysis of its differences in average LM position compared with the other systems was useful in order to find patterns to discriminate these systems from the others. The findings of this study implied broader ribs in specimens from both lagoons (LM #15). However, despite the patterns observed being a good means to discriminate the lagoons' specimens from the other systems, it was not the case for the distinction between each lagoon.

The RF classifier with species and systems simultaneously easily distinguished *C. edule* specimen from the Albufeira lagoon (smallest class error (4.76%)). This was expected since the Albufeira lagoon was considered the most distinct system alongside Ria de Aveiro, as mentioned before. Moreover, when observing the differences in the average LM position between the overall mean versus the Albufeira lagoon's *C. edule*, it was unsurprising to find variation in all LMs except #14 (Type III). Most importantly, all Type I LMs (numbers 1-4 (muscle scars) and 15 (pallial line)) and LMs #5 to #8 and #10 (Type II, on the hinge) showed significant vector length or/and direction that might reflect smaller shaped shells in this system. Therefore, it might be possible to differentiate *C. edule*'s shell from the Albufeira lagoon as smaller than the other groups on average. Additionally, these variations were verified when comparing average LM configuration of all *C. edule* specimens except from the Albufeira

lagoon versus those specimens. It was also seen that the ligament tip (LM #11) in those specimens was placed more towards the interior of the shell. This could be related to a thicker shell and might explain why those specimens were the only *C. edule* group with an average width higher than its *C. glaucum* counterpart and might represent a means to discriminate that group from the others. As mentioned before, thicker shells have been linked to sediment type [6]. Geomorphologic configuration modifications have been reported annually in the Albufeira lagoon to avoid stratification in the water column, promoting exchanges with the seawater and bigger sediment export [230]. Therefore, it might be useful to further analyse the relationship between shell morphology, sediment type, and geomorphologic configuration in future works alongside the recent hydrodynamics of this system.

In this study, it was considered relevant to analyse the representation of the differences in average LM position between the overall mean for each species and all systems except the Albufeira lagoon. The vectors representing those variations were expected to be small and it was verified that they were associated to Ria de Aveiro. This fact could most likely be because Ria de Aveiro is the other system that was considered most distinct from the other groups. The most notable changes in average LM position were already observed throughout this study for Ria de Aveiro's specimens.

Furthermore, looking into more detail, when comparing vectors from the representation of each specimen from Ria de Aveiro and from the Albufeira lagoon, group discrimination could be associated with LM #2 and #4 (both Type I, placed on the muscle scars), which hinted to a smaller anterior and posterior muscle scar exclusively associated with *C. glaucum* from the Albufeira lagoon. Other variations were common for both species, therefore, exclusive to each aquatic system, most already analysed before, thus confirming earlier claims about Ria de Aveiro and the Albufeira lagoon's samples.

In the case of the Sado estuary, the RF classifier built for systems showed the second highest values for class error (56.00%), this system's shell being highly confused with shells from the Tagus estuary, as mentioned before. Furthermore, the Sado estuary's shells never stood out throughout the different analyses of this study (*e.g.*, PCA, CVA, etc…), therefore, any differences seen in LM variation were small (in comparison to other systems) and difficult to discuss.

The visual representation of the average LM position between the overall mean versus the Sado estuary was interesting to analyse. The variation associated with LMs numbers 1-4 (Type I), placed on the adductor muscle scars, showed vectors located on the posterior muscle scars (#3 and #4) turning inwards and downwards the shell, while the vectors located on the anterior muscle scars (#1 and #2) were directing upwards the shell (towards the umbo). This could be a way to discriminate the Sado estuary's shell from the other four aquatic systems. In addition, LMs numbers 9 and 10 (Type II, reflecting primary homologies [113]), which are placed on the hinge, also varied towards the interior of the shell, indicating that the Sado estuary's shell had smaller hinge (which has never been reported before, to the author's best knowledge), therefore might be thicker than shells from the other studied systems. These facts were verified when observing LM variation of *C. edule* shells from the Sado estuary compared with the overall mean, aside from showing a pallial line (associated with LMs #15 and #16) further away from the umbo.

The Sado estuary's *C. edule* were isolated from the other groups when observing variation distributed in PC4 and, when looking at the representation of the differences in average LM position between the overall mean and the Sado estuary's *C. edule*, it was possible to visualise which LMs were associated with such isolation. They related to a smaller hinge (LM #9 and #10) and a posterior area (LM numbers 3, 4, 10, and 13) directed more towards the interior of the shell, signifying a smaller length or thicker shells for *C. edule* from the Sado estuary. Shell thickness has been reported to be linked to

predation and to influence burrowing ability (as mentioned earlier) [6]. The higher thickness of the shells of *C. edule* from the Sado estuary might relate to higher predation pressure or to sediment type. Therefore, might be useful to further study these relationships in future works in order to better differentiate this system from others.

Furthermore, the differences in the average LM position between *C. edule* from Ria de Aveiro and from the Sado estuary corroborated that a good method to differentiate Ria de Aveiro's *C. edule* from other *C. edule* would be identifying which specimens had an anterior muscle scar placed downwards the shell, farther from to the umbo. Beyond that, it was found that vector length and direction hinted at a smaller hinge (LM #5 to #10) for the Sado estuary's *C. edule* and generally smaller sized shell (LM numbers 3-11, and 13) (verified when looking at Csize variation). In addition, it was suggested that a possible way to discriminate the Sado estuary's *C. edule* from the other groups (including the Sado estuary's *C. glaucum*) would be analysing differences in LMs #10 (on the hinge) and #4 (on the posterior muscle scar).

4.2.1 Disparity analysis

A disparity analysis (DA) was also selected as a statistical approach to identify shape differences. DA are a current and widespread description of morphological diversity [201, 129]. It is acknowledged that, for multidimensional datasets, the use of multiple descriptors is beneficial [200], and, since different types of descriptors capture and emphasise different disparity patterns that may or may not be consistent even when using the same set of sampled taxa [200, 201], it was unsurprising to find different results with the DA when using the mathematical descriptor (Csize-based) and when using the model-based descriptor. Such differences in results are most probably due to the use of a linear model where there is a common allometric model for all species and coastal systems. That is, theoretical morphospaces constructed from model-based descriptors are thought of as independent from the empirical sample of specimens studied and capable of producing non-existent morphologies [201]. Additionally, as discussed before, the data is influenced by biological phenomena such as allometry [200], and differences in allometric patterns can have a large impact on disparity [129]. By fixing parameters via the use of the model, interactions were supposedly counterbalanced, and the disparity increased [129]. This was not observed in this study.

Results for the first descriptor (Csize-based) showed that *C. edule* from the Albufeira lagoon is the group that has a visibly bigger MD compared to the other groups, while *C. glaucum* from the Óbidos lagoon showed the smallest MD. It is expected that a healthier stock shows a higher shape diversity *i.e.*, larger disparity [200], since these would be ecologically and functionally more diverse [200, 238]. Knowing this, it would be logical to deduce that *C. edule* from the Albufeira lagoon would be the most diverse and healthiest stock of all groups analysed, followed by *C. edule* from the Sado estuary.

However, when looking at the results of the MD calculated using a model-based descriptor (second descriptor), *C. edule* from Ria de Aveiro is the group with higher MD and *C. edule* from the Albufeira lagoon becomes the group of *C. edule* with less MD, contradicting earlier findings. The logical assumption would be that *C. edule* from Ria de Aveiro is the healthiest stock among groups and *C. glaucum* from the Albufeira lagoon would be the least diverse and healthy. It would be sensible to infer that this fact might be due to the size of the sample, nonetheless, Albufeira is not the system with less specimens of *C. glaucum* sampled. The system with smaller sample size is *C. glaucum* from the Óbidos lagoon which is the group with less MD only in the case where MD is calculated with the first descriptor.

Furthermore, according to [201], methods which use eigenvalues, like it is the case here, are relatively insensitive to sample size. This can be confirmed when looking at the Tagus estuary's sample size for *C. edule*, which is about 2.5 times bigger than the next big sample (the Albufeira lagoon's *C. edule*). If sample size were to be significant in this analysis, it would be logical to assume that *C. edule* from the Tagus estuary to have the highest MD amongst all groups, which is not the case. *C. edule* from the Tagus estuary is the group of *C. edule* with less MD. Following this train of thought and knowing that samples from the Tagus estuary were collected in multiple stations (12) in different areas of the estuary, it becomes clear that, according to this part of the analysis, *C. edule* from the Tagus estuary is the less diverse and healthy stock amongst *C. edule* samples. This might be supported by the fact that this system is subjected to a strong anthropogenic pressure (*e.g.*, [50, 239]). Yet, when looking at *C. glaucum* from the Tagus estuary the same cannot be deduced, particularly when analysing MD through the second descriptor, where these specimens become the group with highest MD.

Additionally, when comparing the two descriptors and *C. glaucum* specimens, both lagoons presented the less MD, therefore, are expected to be ecologically and functionally less diverse and have a less healthy stock [200, 238]. These values being alike might be linked to the fact that both these systems are semi-enclosed coastal lagoons, and thus, have similar environmental conditions. As mentioned before, the recent hydrodynamics of both lagoons have been studied [230, 237]. Geomorphologic configuration modifications in both estuaries have been happening [230, 237] affecting hydrodynamics in these systems. This can influence bivalves' shell morphology, thus, further works should be done on the subject in the future.

Moreover, it is important to note that disparity analyses are a projection of the morphospace (just as PCA) and not the morphospace itself [201]. It is a useful and informative tool to represent high-dimensional spaces, however, these projections can be misleading since there is always loss of information when displayed in such ways as they were in this study [201].

4.3 For future reference

The results obtained in this work support a set of recommendations to be considered when carrying out future works in GM.

Species identification should be confirmed with genetic analysis to avoid misinterpretations, since, for this study, the initial identification of the species was based solely on external morphological characters. The morphological similarities of the two studied species and their high phenotypic variation make their differentiation difficult, which is why a genetic analysis might be helpful [11, 45, 48, 68]. Nevertheless, previous research indicated that GM could differentiate between populations in a way that genetic analysis could not [68].

It would be interesting to expand the study to other aquatic systems in order to have further contrasting comparisons that provide additional and more accurate differences between aquatic systems.

To further analyse the population size structure, it is essential to use the same sampling methods for all aquatic systems and species (for the linear morphometrics analysis), which was not possible here due to the origin of the specimens. However, GM methods are relatively independent from sampling methodologies and size influences [68, 204], therefore, being the best methods used in this study.

The integration of environmental data in the analysis (linear model, PCA, and CVA) to study the patterns and interactions between such factors and the systems and the shape of *Cerastoderma* spp. would be an interesting aspect to add to the current work. Although the final proof of the relationship between a variable and shape can only be obtained under laboratory conditions, where most variables can be controlled.

It would also be interesting to use more advanced GM methods, for example, the use of sliding LMs, three-dimensional data [95, 132, 240, 241] or LM directly on the images. The cockles' shells are in three-dimensions, therefore the two dimensions coordinates used for the LM analysis in this study might not capture all variability. It is possible that depth between LMs might create a source of error [118]. The techniques can also be complementary, for example, the use of both landmark- and contour-based methods since these provide different information [61, 101]. Moreover, GM is an evolving discipline, providing increasingly powerful techniques for characterisation and comparison of shape [61, 241] therefore, more advanced techniques might have come to light without the author's knowledge.

A last example of recommendation is to avoid the use of semi or pseudo-landmarks for GM analysis and prioritise LMs with a true biological homology origin since these represent morphologically equivalent structures (primary homologies) most accurately [109, 113].

5. Final considerations

Cerastoderma spp. are important bivalve species, not only economically but also ecologically. They occur in sympatry in most areas, including the five systems studied in this work. Their wide distribution and considerable morphological similarities challenges population discrimination and, therefore, their traceability. Traceability is a highly important concept for fisheries and permits to determine/confirm the origins of individuals for which the harvest location is not clear.

In the case of the *Cerastoderma* genus, traceability techniques have been focused on fatty acid profiles or geochemical profiles. This work aimed at finding an adequate, easy, and low cost approach for population discrimination that is as effective as other methods. Such approach could be regarded as an important contribution for the development of traceability techniques for *Cerastoderma* spp. or other bivalve species. The use of GM provided accurate results to discriminate cockle populations from different aquatic systems on the Portuguese coast, without the need for highly skilled laboratory technicians nor costly reagents and equipment, achieving the objectives of this work.

Besides being rapid and simple, GM methods can provide fundamental information for a wide range of contexts such as biology, ecology, evolution, or, like it is the case here, fisheries, in particular stock/population traceability. The use of GM analysis addresses subtle shape variations by quantifying and visualising shape changes. Therefore, it is an excellent approach for population discrimination, particularly in cases where sibling species are present in systems with distinct characteristics.

The findings of this study support the use of landmark-based GM analysis. This methodology allowed the study in great detail of the shape shell variation of *C. edule* and *C. glaucum* removing any information unrelated to shape (*e.g.*, scale, orientation, etc…) through Generalised Procrustes analysis (GPA). The final model displayed that differences in shape between systems and species were significant. These differences were quantified through the trajectory analysis (TA) and its trajectory correlations and magnitudes. Through the visualisation of the Principal component analysis (PCA), its

principal components (PCs), the consensus configuration, and the visualisation of LM differences between groups, it was possible to understand in which concrete aspects shell shape changes occurred. These landmark-based GM methods permitted the description of complex structures like the shell of a bivalve and allowed the differentiation between aquatic systems and species' shell shapes with great detail.

In addition, it was possible to accurately classify species and aquatic systems through the use of machine learning *i.e.*, Random forest (RF), using landmark-based GM data only. The findings of this work provide good prospects for future studies with the objective of distinguishing between *C. edule* and *C. glaucum* and aquatic systems using this technique.

In summary, the landmark-based GM methods used in this work successfully allowed the study in great detail of the shape shell variation of two bivalve species with commercial interest (*C. edule* and *C. glaucum*) and permitted the differentiation of five different geographical groups (Ria de Aveiro; the Óbidos coastal lagoon; the Tagus estuary; the Albufeira coastal lagoon; the Sado estuary) distributed along the Portuguese coast.

Finally, the methodologies used in this work could be extended to other bivalve populations and aquatic systems, being an accurate, easy, and low cost discrimination approach. Such procedures could be beneficial for the traceability of other populations and a valuable asset in fisheries in general.

6. References[1]

1. Carss DN, Brito AC, Chainho P, Ciutat A, de Montaudouin X, Fernández et al. Ecosystem services provided by a non-cultured shellfish species: The common cockle *Cerastoderma edule*. Marine Environmental Research. 2020; 158. doi: 10.1016/j.marenvres.2020.104931.
2. Gili J, Coma R. Benthic suspension feeders: their paramount role in littoral marine food webs. Trends in Ecology & Evolution. 1998;13(8):316–321. doi: 10.1016/S0169-5347(98)01365-2
3. Turja R, Höher N, Snoeijs P, Baršienė J, Butrimavičienė L, Kuznetsova T, et al. A multibiomarker approach to the assessment of pollution impacts in two Baltic Sea coastal areas in Sweden using caged mussels (*Mytilus trossulus*). Science of The Total Environment. 2014;473-474: 398–409. doi: 10.1016/j.scitotenv.2013.12.038
4. Seed R. Shell growth and form in the Bivalvia. In: Rhoads DC. Lutz RA. editors. Skeletal growth of aquatic organisms. New York: Plenum Publishing Corp; 1980. pp. 23–67.
5. Funk A, Reckendorfer, W. Environmental Heterogeneity and Morphological Variability in *Pisidium subtruncatum* (Sphaeriidae, Bivalvia). International Review of Hydrobiology. 2008;93:188–199. doi: 10.1002/iroh.200710969
6. Stanley SM. Relation of shell form to life habits in the Bivalvia. Geological Society of America. 1970;125.
7. Yusseppone MS, Bianchi VA, Castro JM, Luquet CM, Sabatini SE, Ríos de Molina MC, et al. Long-term effects of water quality on the freshwater bivalve *Diplodon chilensis* (Unionida: Hyriidae) caged at different sites in a North Patagonian river (Argentina). Ecohydrology. 2020;13: e2181. doi: 10.1002/eco.2181
8. Boyden CR. A Comparative Study of the Reproductive Cycles of the Cockles *Cerastoderma Edule* and *C. Glaucum*. Journal of the Marine Biological Association of the United Kingdom. 1971;51(03):605. doi: 10.1017/s0025315400014995

[1] – References made accordingly to the PLoS ONE scientific journal, ISSN: 1932-6203 (presented in an internationally accepted manner, according to the scientific area in which the final work is included).

9. Brock V. The geographical distribution of *Cerastoderma* [*Cardium*] *edule* (L.) and *C. Lamarcki* (Reeve) in the baltic and adjacent seas related to salinity and salinity fluctuations. Ophelia. 1980;19(2):207–214. doi: 10.1080/00785326.1980.10425517
10. Brock V. Does displacement of spawning time occur in the sibling species *Cerastoderma edule* and *C. lamarcki*?. Marine Biology. 1982;67:33–38. doi: 10.1007/BF00397091
11. Machado MM, Costa AM. Enzymatic and morphological criteria for distinguishing between *Cardium edule* and *C. glaucum* of the Portuguese coast. Marine Biology. 1994;120(4):535–44. doi: 10.1007/BF00350073
12. Mariani S, Piccari F, De Matthaeis E. Shell morphology in *Cerastoderma* spp. (Bivalvia: Cardiidae) and its significance for adaptation to tidal and non-tidal coastal habitats. Journal of the Marine Biological Association of the United Kingdom. Cambridge University Press; 2002;82(3):483–90. doi: 10.1017/S0025315402005751
13. Leontarakis PK, Xatzianastasiou LI, Theodorou JA. Biological Aspects of the Lagoon Cockle, *Cerastoderma glaucum* (Poiret 1879), in a Coastal Lagoon in Keramoti, Greece in the Northeastern Mediterranean. Journal of Shellfish Research. 2008;27(5):1171–5. doi: 10.2983/0730-8000-27.5.1171
14. Sanchez-Salazar ME, Griffiths CL, Seed R. The effect of size and temperature on the predation of cockles *Cerastoderma edule* (L.) by the shore crab *Carcinus maenas* (L.) Journal of Experimental Marine Biology and Ecology. 1987;111:181–193. doi: 10.1016/0022-0981(87)90054-2
15. Beukema JJ, Dekker R. Decline of recruitment success in cockles and other bivalves in the Wadden Sea: possible role of climate change, predation on postlarvae and fisheries. Marine Ecology Progress Series. 2005;287:149e167. doi: 10.3354/meps287149
16. Beukema JJ, Dekker R. Annual cockle *Cerastoderma edule* production in the Wadden Sea usually fails to sustain both wintering birds and a commercial fishery. Marine Ecology Progress Series. 2006;309:189–204. doi: 10.3354/meps309189
17. Crespo D, Verdelhos T, Dolbeth M, Pardal MA. Effects of the over harvesting on an edible cockles (*Cerastoderma edule* Linaeus, 1758) population on a southern European estuary. Fresenius Environmental Bulletin. 2010;19(12):2801–2811
18. Genelt-Yanovskiy E, Poloskin A, Granovitch A, Nazarova S, Strelkov P. Population structure and growth rates at biogeographic extremes: A case study of the common cockle, *Cerastoderma edule* (L.) in the Barents Sea. Marine Pollution Bulletin. 2010;61(4–6):247–53. doi: 10.1016/j.marpolbul.2010.02.021
19. Whitton TA, Jenkins SR, Richardson CA, Hiddink JG. Aggregated prey and predation rates: juvenile shore crabs (*Carcinus maenas*) foraging on post-larval cockles (*Cerastoderma edule*). Journal of Experimental Marine Biology and Ecology. 2012;432–433:29–36. doi: 10.1016/j.jembe.2012.07.014
20. Dunham A, Gurney-Smith H, Plamondon N, Yuan S, Pearce CM. Aquaculture potential of the basket cockle (*Clinocardium nuttallii*). Part 1: Effects of stocking density on first year grow-out performance in intertidal and off-bottom suspended culture. Aquaculture Research. 2013;44(8):1236–53. doi: 10.1111/j.1365-2109.2012.03125.x
21. Morgan E, O' Riordan RM, Culloty SC, Riordan RMO, Culloty SC. Climate change impacts on potential recruitment in an ecosystem engineer. Ecology Evolution. 2013;3(3):581–94. doi: 10.1002/ece3.419
22. André C, Lindegarth M, Jonsson PR, Sundberg P. Species identification of bivalve larvae using random amplified polymorphic DNA (RAPD): differentiation between *Cerastoderma edule* and *C. lamarki*. Journal of the Marine Biological Association of the United Kingdom. 1999;79:563–565. doi: 10.1017/S0025315498000691

23. Reise K. Metapopulation structure in the lagoon cockle *Cerastoderma lamarcki* in the northern Wadden Sea. Helgoland Marine Research. 2003;56;252–258. doi: 10.1007/s10152-002-0125-z
24. Kater BJ, Geurts van Kessel AJM, Baars JJMD. Distribution of cockles *Cerastoderma edule* in the eastern Scheldt: habitat mapping with abiotic variables. Marine Ecology Progress Series. 2006;318:221–227. doi: 10.3354/meps318221
25. Gam M, de Montaudouin X, Bazairi H. Population dynamics and secondary production of the cockle *Cerastoderma edule*: a comparison between Merja Zerga (Moroccan Atlantic Coast) and Arcachon Bay (French Atlantic Coast). Journal of Sea Research. 2010;63:191–201. doi: 10.1016/j.seares.2010.01.003
26. Lobo J, Costa PM, Caeiro S, Martins M, Ferreira AM, Caetano, M, et al. Evaluation of the potential of the common cockle (*Cerastoderma edule* L.) for the ecological risk assessment of estuarine sediments: bioaccumulation and biomarkers. Ecotoxicology. 2010;19(8):1496–1512. doi: 10.1007/s10646-010-0535-7
27. Nilin J, Pestana JLT, Ferreira NG, Loureiro S, Costa-Lotufo LV, Soares AM. Physiological responses of the European cockle *Cerastoderma edule* (Bivalvia: Cardidae) as indicators of coastal lagoon pollution. Science of The Total Environment. 2012;435:44–52. doi: 10.1016/j.scitotenv.2012.06.107
28. Maia F, Barroso CM, Gaspar MB. Biology of the common cockle *Cerastoderma edule* (Linnaeus, 1758) in Ria de Aveiro (NW Portugal): Implications for fisheries management. Journal of Sea Research. 2021;171:102024. doi: 10.1016/j.seares.2021.102024
29. Jensen KT. Dynamics and growth of the cockle, *Cerastoderma edule*, on an intertidal mud-flat in the Danish Wadden sea: Effects of submersion time and density. Netherlands Journal of Sea Research. 1992;28(4):335-345. doi: 10.1016/0077-7579(92)90035-D
30. Montaudouin X, Bachelet G. Experimental evidence of complex interactions between biotic and abiotic factors in the dynamics of an intertidal population of the bivalve *Cerastoderma edule*. Oceanologica Acta. 1996;19(3–4):449–463.
31. Dabouineau L, Ponsero A. Synthesis on biology of Common European Cockle *Cerastoderma edule*; 2011 [cited 2021 May 13]. Database: Archive ouverte HAL [Internet]. Available from: https://hal.archives-ouvertes.fr/hal-00581394
32. Beukema JJ. Biomass and species richness of the macrobenthic animals living on a tidal flat area in the Dutch Wadden Sea: Effects of a severe winter. Netherlands Journal of Sea Research. 1979;13(2):203-223. doi: 10.1016/0077-7579(79)90003-6
33. Strasser M, Reinwald T, Reise K. Differential effects of the severe winter of 1995/96 on the intertidal bivalves *Mytilus edulis*, *Cerastoderma edule* and *Mya arenaria* in the Northern Wadden Sea. Helgoland Marine Research. 2001;55:190–197. doi: 10.1007/s101520100079
34. Mahony KE, Egerton S, Lynch SA, Blanchet H, Goedknegt MA, Groves E, et al. Drivers of growth in a keystone fished species along the European Atlantic coast: The common cockle *Cerastoderma edule*. Journal of Sea Research. 2022;179:102148. doi: 10.1016/j.seares.2021.102148
35. Ong EZ, Briffa M, Moens T, Van Colen C. Physiological responses to ocean acidification and warming synergistically reduce condition of the common cockle *Cerastoderma edule*. Marine Environmental Research. 2017;130:38–47. doi: 10.1016/j.marenvres.2017.07.001
36. Cubillas P, Köhler S, Prieto M, Chaïrat C, Oelkers EH. Experimental determination of the dissolution rates of calcite, aragonite, and bivalves. Chemical Geology. 2005;216(1-2):59–77. doi: 10.1016/j.chemgeo.2004.11.009
37. Barnes RSK. The intertidal lamellibranchs of Southampton water, with particular reference to *Cerastoderma edule* and *C. glaucum*. Journal of Molluscan Studies. 1973;40(5):413–433. doi: 10.1093/oxfordjournals.mollus.a065239

38. Thomas KD. Population studies on molluscs in relation to environmental archaeology. In: Brothwell DR, Thomas KD, Clutton-Brock J. editors. Research Problems in Zooarchaeology. University College London Institute of Archaeology Publications, Vol. 20, Occasional Publication nº3; 2016. pp. 9–18.
39. Rygg B. Studies on *Cerastoderma edule* (L.) and *Cerastoderma glaucum* (Poiret). Sarsia. 1970;43(1):65–80. doi: 10.1080/00364827.1970.10411169
40. Boyden R, Russell PJC. The Distribution and Habitat Range of the Brackish Water Cockle (*Cardium* (*Cerastoderma*) *glaucum*) in the British Isles. Journal of Animal Ecology. 1972;41(3):719–734. doi: 10.2307/3205
41. Brock V. Habitat selection of two congeneric bivalves, *Cardium edule* and *C. glaucum* in sympatric and allopatric populations. Marine Biology. 1979;54:149–156. doi: 10.1007/BF00386594
42. Verdelhos T, Marques J, Anastácio P. The impact of estuarine salinity changes on the bivalves *Scrobicularia plana* and *Cerastoderma edule*, illustrated by behavioral and mortality responses on a laboratory assay. Ecological Indicators. 2015;52:96–104. doi: 10.1016/j.ecolind.2014.11.022
43. Boyden CR. The Behaviour, Survival and Respiration of the Cockles *Cerastoderma Edule* and *C. Glaucum* in Air. Journal of the Marine Biological Association of the United Kingdom. 1972;52(03):661. doi: 10.1017/s0025315400021640
44. Russell PJC. Biological studies on *Cardium glaucum*, based on some Baltic and Mediterranean populations. Marine Biology. 1972;16(4):290–296. doi: 10.1007/BF00347752
45. Tarnowska K, Wołowicz M, Chenuil A, Féral JP. Comparative studies on the morphometry and physiology of European populations of the lagoon specialist *Cerastoderma glaucum* (Bivalvia). Oceanologia. 2009;51(3):437–58. doi: 10.5697/oc.51-3.437
46. Malham SK, Hutchinson TH, Longshaw, M. A review of the biology of European cockles (*Cerastoderma* spp.). Journal of the Marine Biological Association of the United Kingdom. 2012;92(07):1563–1577. doi: 10.1017/s0025315412000355
47. Kingston PF. Studies on the reproductive cycles of *Cardium edule* and *C. glaucum*. Marine Biology. 1974;28:317–323. doi: 10.1007/BF00388500
48. Brock V. Morphological and biochemical criteria for the separation of *Cardium glaucum* (Bruguiere) from *Cardium edule* (L.). Ophelia. 1987;17:207–214. doi: 10.1080/00785326.1978.10425484
49. Cardoso JFMF, Witte JI, van der Veer HW. Differential reproductive strategies of two bivalves in the Dutch Wadden Sea. Estuarine, Coastal and Shelf Science. 2009;84:37–44. doi: 10.1016/j.ecss.2009.05.026
50. Freire P, Rilo A, Ceia R, Nogueira MRM, Catalao J, Taborda R et al. Tipificação das zonas marginais estuarinas. O caso do estuário do Tejo. Actas das 2as Jornadas de Engenharia Hidrográfica; 2012 Jun 20-22; Lisbon, Portugal. Lisbon: Instituto Hidrográfico; 2012. p. 319–322. doi: 10.13140/2.1.3597.5369
51. Martinez-Castro C, Vázquez E. Reproductive cycle of the cockle *Cerastoderma edule* (Linnaeus 1758) in the ria de Vigo (Galicia, Northwest Spain). Journal of Shellfish Research. 2012;31(3):757–767. doi: 10.2983/035.031.0320
52. Food and Agriculture Organization of the United Nations [Internet]. c2022 [cited 2022 Oct 13]. Available from: http://www.fao.org/fishery/statistics/global-aquaculture-production/en
53. Gamito SJ. The benthic macrofauna of some water reservoirs of salt-pans from Ria Formosa (Portugal). Scientia Marina. 1989;53(2–3):145–158
54. Lista de Espécies [Internet]. Instituto Português do Mar e da Atmosfera. c2019 [cited 2021 May 12]. Available from: https://www.ipma.pt/pt/bivalves/docs/files/Lista_de_especies_em_29_05_2019.pdf

55. Instituto Nacional de Estatística - Estatísticas da Pesca: 2019 [Internet]. Lisboa: INE. c2020 [cited 2021 May 12]. Available from https://www.ine.pt/xurl/pub/43569029
56. Instituto Nacional de Estatística - Estatísticas da Pesca: 2021 [Internet]. Lisboa: INE. c2022 [cited 2022 Sep 23]. Available from https://www.ine.pt/xurl/pub/36828280
57. Dickie LM, Boudreau PR, Freeman KR. Influences of stock and site on growth and mortality in the blue mussel (*Mytilus edulis*). Canadian Journal of Fisheries and Aquatic Sciences. 1984;41:134–140. doi: 10.1139/f84-013
58. Fuentes J, Reyero I, Zapata C, Alvarez G. Influence of stock and culture site on growth rate and mortality of mussels (*Mytilus galloprovincialis* Lmk.) in Galicia, Spain. Aquaculture. 1992;105:131–142. doi: 10.1016/0044-8486(92)90125-5
59. McCoy MW, Bolker BM, Osenberg CW, Miner BJ, Vonesh JR. Size correction: comparing morphological traits among populations and environments. Oecologia. 2006;148:547–554. doi: 10.1007/s00442-006-0403-6
60. Shields JL, Barnes P, Heath DD. Growth and survival differences among native, introduced and hybrid blue mussels (*Mytilus* spp.): Genotype, environment and interaction effects. Marine Biology. 2008;154:919–928. doi: 10.1007/s00227-008-0985-0
61. Rufino MM, Vasconcelos P, Pereira F, Fernández-Tajes J, Darriba S, Méndez J, et al. Geographical variation in shell shape of the pod razor shell *Ensis siliqua* (Bivalvia: Pharidae). Helgoland Marine Research. 2012;67(1):49–58. doi: 10.1007/s10152-012-0303-6
62. De Noia M, Telesca L, Vendrami DLJJ, Gokalp HK, Charrier G, Harper EM, et al. Population genetic structure is unrelated to shell shape, thickness and organic content in european populations of the soft-shell Clam *Mya arenaria*. Genes (Basel). 2020;11(3):1–20. doi: 10.3390/genes11030298
63. Mirzoeva A, Zhukov O. Conchological variability of *Anadara kagoshimensis* (Bivalvia: Arcidae) in the northern part of the Black–Azov Sea basin. Biologia (Bratisl). 2021;76(12):3671–84. doi: 10.1007/s11756-021-00844-4
64. Seed, R. Factors Influencing Shell Shape in the Mussel *Mytilus Edulis*. Journal of the Marine Biological Association of the United Kingdom. 1968;48(3):561-584. doi: 10.1017/S0025315400019159
65. Brown RA, Seed R, O'Connor RJ. A comparison of relative growth in *Cerastoderma* (=*Cardium*) *edule*, *Modiolus modiolus*, and *Mytilus edulis* (Mollusca: Bivalvia). Journal of Zoological Society of London. 1976;17:297–315. doi: 10.1111/j.1469-7998.1976.tb02298.x
66. Eagar RMC, Stone NM, Dickson PA. Correlations between shape, weight and thickness of shell in four populations of *Venerupis romboides* (Pennant), Journal of Molluscan Studies. 1984;50(1):19–38. doi: 10.1093/oxfordjournals.mollus.a065839
67. Baker AM, Bartlett C, Bunn SE, Goudkamp K, Sheldon F, Hughes JM. Cryptic species and morphological plasticity in long-lived bivalves (Unionoida: Hyriidae) from inland Australia. Molecular Ecology. 2003;12:2707–2717. doi: 10.1046/j.1365-294X.2003.01941.x
68. Sousa R, Freire R, Rufino M, Méndez J, Gaspar M, Antunes C, et al. Genetic and shell morphological variability of the invasive bivalve *Corbicula fluminea* (Müller, 1774) in two Portuguese estuaries. Estuarine, Coastal and Shelf Science. 2007;74(1–2):166–74. doi: 10.1016/j.ecss.2007.04.011
69. Costa C, Aguzzi J, Menesatti P, Antonucci F, Rimatori V, Mattoccia M. Shape analysis of different populations of clams in relation to their geographical structure. Journal of Zoology. 2008;276(1):71–80. doi: 10.1111/j.1469-7998.2008.00469.x
70. Caill-Milly N, Bru N, Barranger M, Gallon L, D'amico F. 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. 2014;33(2):355–72. doi: 10.2983/035.033.0205

71. Aguirre ML, Richiano S, Álvarez A, Farinati EA. Reading shell shape: implications for palaeoenvironmental reconstructions. A case study for bivalves from the marine Quaternary of Argentina (south-western Atlantic). Historical Biology. 2016;28(6):753–73. doi: 10.1080/08912963.2015.1026898
72. Aguirre ML, Perez SI, Sirch YN. Morphological variability of Brachidontes Swainson (Bivalvia, Mytilidae) in the marine Quaternary of Argentina (SW Atlantic). Palaeogeography, Palaeoclimatology, Palaeoecology. 2006;239(1–2):100–25. doi: 10.1016/j.palaeo.2006.01.019
73. Funk A, Reckendorfer W. Environmental Heterogeneity and Morphological Variability in *Pisidium subtruncatum* (Sphaeriidae, Bivalvia). International Review of Hydrobiology. 2008;93(2):188–99. doi: 10.1002/iroh.200710969
74. Johnson EH. Experimental tests of bivalve shell shape reveal potential tradeoffs between mechanical and behavioral defenses. Scientific Reports. 2020;10(1):19425–36. doi: 10.1038/s41598-020-76358-x
75. Eagar RMC. Shape and function of the shell: a comparison of some living and fossil bivalve mollusc. Biological Reviews. 1978;53:169–210. doi: 10.1111/j.1469-185X.1978.tb01436.x
76. Gaspar MB, Santos MN, Vasconcelos P, Monteiro C. Shell morphometric relationships of the most common bivalve species (Mollusca: Bivalvia) of the Algarve coast (southern Portugal). Hydrobiologia. 2002;477:73–80. doi: 10.1023/A:1021009031717
77. Scholz H, Hartman JH. Paleoenvironmental reconstruction of the Upper Cretaceous Hell Creek Formation of the Williston Basin, Montana, USA: implications from the quantitative analysis of unionid bivalve taxonomic diversity and morphologic disparity. Palaios. 2007;22(1):24–34. doi: 10.2110/palo.2005.p05-059r
78. Zieritz A, Aldridge DC. Identification of ecophenotypic trends within three European freshwater mussel species (Bivalvia: Unionoida) using traditional and modern morphometric techniques, Biological Journal of the Linnean Society. 2009;98(4):814–825. doi: 10.1111/j.1095-8312.2009.01329.x
79. Costa C, Menesatti P, Aguzzi J, D'Andrea S, Antonucci F, Rimatori V, et al. External Shape Differences between Sympatric Populations of Commercial Clams *Tapes decussatus* and *T. philippinarum*. Food Bioprocess Technology. 2010;3(1):43–8. doi: 10.1007/s11947-008-0068-8
80. Gardner JPA, Thompson RJ. Influence of genotype and geography on shell shape and morphometric trait variation among North Atlantic blue mussel (*Mytilus* spp.) populations. Biological Journal of the Linnean Society. 2009;96:875–97. doi: 10.1111/j.1095-8312.2008.01166.x
81. Gordillo S, Márquez F, Cárdenas J, Zubimendi MÁ. Shell variability in *Tawera gayi* (Veneridae) from southern South America: a morphometric approach based on contour analysis. Journal of the Marine Biological Association of the United Kingdom. 2011;91(4):815–22. doi: 10.1017/S0025315410000391
82. Morais P, Rufino MM, Reis J, Dias E, Sousa R. Assessing the morphological variability of *Unio delphinus* Spengler, 1783 (Bivalvia: Unionidae) using geometric morphometry. Journal of Molluscan Studies. 2014;80(1):17–23. doi: 10.1093/mollus/eyt037
83. Rufino MM, Gaspar MB, Pereira AM, Vasconcelos P. Use of shape to distinguish *Chamelea gallina* and *Chamelea striatula* (Bivalvia: Veneridae): Linear and geometric morphometric methods. Journal of Morphology. 2006;267(12):1433–40. doi: 10.1002/jmor.10489
84. Márquez F, Robledo J, Peñaloza GE, Van der Molen S. Use of different geometric morphometrics tools for the discrimination of phenotypic stocks of the striped clam *Ameghinomya antiqua* (Veneridae) in north Patagonia, Argentina. Fisheries Research. 2010;101(1–2):127–31. doi: 10.1016/j.fishres.2009.09.018

85. Nerlović V, Korlević M, Mravinac B. Morphological and Molecular Differences Between the Invasive Bivalve *Ruditapes philippinarum* (Adams & Reeve, 1850) and the Native Species *Ruditapes decussatus* (Linnaeus, 1758) from the Northeastern Adriatic Sea. Journal of Shellfish Research. 2016;35(1):31–9. doi: 10.2983/035.035.0105
86. Uba KIN. Determining Shell Shape Differences in the Horse Mussels *Modiolus philippinarum* (Hanley 1843) and *Modiolus modulaides* (Röding 1798) by Morphometric Analysis. Philippine Journal of Science. 2021;150(4):743–52. doi: 10.56899/150.04.12
87. Pepin P, Carr SM. Morphological, meristic, and genetic analysis of stock structure in juvenile Atlantic cod (*Gadus morhua*) from the Newfoundland Shelf. Canadian Journal of Fisheries and Aquatic Sciences. 1993;50:1924–1933. doi: 10.1139/f93-215
88. Swain DF, Frank KT, Mailet G. Delineating stocks of Atlantic cod (*Gadus morhua*) in the Gulf of St. Lawrence and Cabot Strait areas using vertebral number. ICES Journal of Marine Science. 2001;58:253–269. doi: 10.1006/jmsc.2000.1007
89. Higgins RM, Danilowicz BS, Balbuena JA, Daníelsdóttir AK, Geffen AJ, Meijer WG, et al. Multi-disciplinary fingerprints reveal the harvest location of cod *Gadus morhua* in the northeast Atlantic. Marine Ecology Progress Series. 2010;404:197–206. doi: 10.3354/meps08492
90. Nielsen EE, Hansen MM, Schmidt C, Meldrup D, Grønkjær P. Population of origin of Atlantic cod. Nature. 2001;413:272. doi: 10.1038/35095112
91. Thorrold SR, Latkoczy C, Swart PK, Jones CM. Natal homing in a marine fish metapopulation. Science. 2001;291:297–299. doi: 10.1126/science.291.5502.297
92. Primmer CR, Koskinen MT, Piironen J. The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. Proceedings of the Royal Society Biological Sciences. 2000;267:1699–1704. doi: 10.1098/rspb.2000.1197
93. Jacquet JL, Pauly D. Trade secrets: renaming and mislabeling of seafood. Marine Policy. 2008;32:309–318. doi: 10.1016/j.marpol.2007.06.007
94. Zuur AF, Ieno EN, Smith GM. Analysing Ecological Data. 1st ed. New York: Springer; 2007.
95. Adams DC, Rohlf FJ, Slice DE. Geometric morphometrics: Ten years of progress following the 'revolution'. Italian Journal of Zoology. 2004;71(1):5–16. doi: 10.1080/11250000409356545
96. Rufino M, Abelló P, Yule AB. Male and female carapace shape differences in *liocarcinus depurator* (decapoda, brachyura): An application of geometric morphometric analysis to crustaceans. Italian Journal of Zoology. 2004;71(1):79–83. doi: 10.1080/11250000409356554
97. Silva IC, Paula J. Is there a better chela to use for geometric morphometric differentiation in brachyuran crabs? A case study using *Pachygrapsus marmoratus* and *Carcinus maenas*. Journal of the Marine Biological Association of the United Kingdom. 2008;88(5):941–53. doi: 10.1017/S0025315408001483
98. Smith UE, Hendricks JR. Geometric morphometric character suites as phylogenetic data: Extracting phylogenetic signal from gastropod shells. Systematic Biology. 2013;62(3):366–85. doi: 10.1093/sysbio/syt002
99. Ruenchit P. State-of-the-Art Techniques for Diagnosis of Medical Parasites and Arthropods. Diagnostics (Basel). 2021;11(9):1545. doi: 10.3390/diagnostics11091545
100. Kotrschal KM, Motta PJ. Correlative, Experimental, and Comparative Evolutionary Approaches in Ecomorphology, Netherlands Journal of Zoology. 1991;42(2-3):400–415. doi: 10.1163/156854291X00414
101. Krapivka S, Toro JE, Alcapán AC, Astorga M, Presa P, Pérez M, et al. Shell-shape variation along the latitudinal range of the Chilean blue mussel *Mytilus chilensis* (Hupe 1854). Aquaculture Research. 2007;38:1770–7. doi: 10.1111/j.1365-2109.2007.01839.x

102. Russo T, Costa C, Cataudella S. Correspondence between shape and feeding habit changes throughout ontogeny of gilthead sea bream *Sparus aurata* L., 1758. Journal of Fish Biology. 2007;71(3):629–56. doi: 10.1111/j.1095-8649.2007.01528.x
103. de Aranzamendi MC, Martínez JJ, Sahade R. Shape differentiation and characterization in the two morphotypes of the Antarctic limpet *Nacella concinna* using Elliptic Fourier analysis of shells. Polar Biology. 2010;33(9):1163–70. doi: 10.1007/s00300-010-0803-2
104. Zelditch, M. L., Swiderski, D. L., & Sheets, H. D. Geometric morphometrics for biologists: a primer. 2nd ed. San Diego, CA: Academic Press; 2012. doi: 10.1016/C2010-0-66209-2
105. Rohlf FJ. Morphometrics. Annual Review of Ecology and Systematics. 1990;21:299–316. Available from: http://www.jstor.org/stable/2097027
106. Rohlf FJ, Marcus LF. A revolution in Morphometrics. Trends in Ecology & Evolution. 1993;8(4):129–132. doi: 10.1016/0169-5347(93)90024-J.
107. Bookstein FL. Biometrics, biomathematics and the morphometric synthesis. Bulletin of Mathematical Biology. 1996;58(2):313–365. doi: 10.1007/BF02458311
108. Bookstein FL. Morphometric Tools for Landmark Data. Morphometric Tools for Landmark Data. 2nd ed. Cambridge, UK: Cambridge University Press; 1997.
109. Claude J. Morphometrics with R. 1st ed. New York: Springer New York, NY; 2008
110. Mitteroecker P, Gunz P. Advances in geometric morphometrics. Evolutionary Biology. 2009;36(2):235–247. doi: 10.1007/s11692-009-9055-x.
111. Minton RL, Norwood AP, Hayes DM. Quantifying phenotypic gradients in freshwater snails: a case study in *Lithasia* (Gastropoda: Pleuroceridae). Hydrobiologia. 2008;605(1):173–82. doi: 10.1007/s10750-008-9332-1
112. Minton RL, Wang LL. Evidence of sexual shape dimorphism in *Viviparus* (Gastropoda: Viviparidae). Journal of Molluscan Studies. 2011;77(3):315–7. doi: 10.1093/mollus/eyr014
113. Palci A, Lee MSY. Geometric morphometrics, homology and cladistics: review and recommendations. Cladistics. 2019;35(2):230–42. doi: 10.1111/cla.12340
114. Cadrin SX. Advances in morphometric identification of fishery stocks. Reviews in Fish Biology and Fisheries. 2000;10:91–112. doi: 10.1023/A:1008939104413
115. Palmer M, Pons G, Linde M. Discriminating between geographical groups of a Mediterranean commercial clam (*Chamelea gallina* (L.): Veneridae) by shape analysis. Fisheries Research. 2004;67(1):93–98. doi: 10.1016/j.fishres.2003.07.006
116. Valladares A, Manríquez G, Suárez-Isla BA. Shell shape variation in populations of *Mytilus chilensis* (Hupe 1854) from southern Chile: a geometric morphometric approach. Marine Biology. 2010;157:2731–8. doi: 10.1007/s00227-010-1532-3
117. Márquez F, Van Der Molen S. Intraspecific shell-shape variation in the razor clam *Ensis macha* along the Patagonian coast. Journal of Molluscan Studies. 2011;77:123. doi: 10.1093/mollus/eyq044
118. Hedrick BP, Dodson P. *Lujiatun Psittacosaurids*: Understanding Individual and Taphonomic Variation Using 3D Geometric Morphometrics. PLoS ONE. 2013;8(8): e69265. doi: 10.1371/journal.pone.0069265
119. Gold MEL, Brochu CA, Norell MA. An Expanded Combined Evidence Approach to the *Gavialis* Problem Using Geometric Morphometric Data from Crocodylian Braincases and Eustachian Systems. PLoS One. 2014;9(9):e105793. doi: 10.1371/journal.pone.0105793
120. Josek T, Allan BF, Alleyne M. Morphometric Analysis of Chemoreception Organ in Male and Female Ticks (Acari: Ixodidae). Journal of Medical Entomology. 2018;55(3):547–552. doi: 10.1093/jme/tjx232

121. Friedman NR, Bennet BL, Fischer G, Sarnat EM, Huang JP, Knowles LLK, et al. Macroevolutionary integration of phenotypes within and across ant worker castes. Ecology and Evolution. 2020;10:9371– 9383. doi: 10.1002/ece3.6623
122. Henriques D, Chávez-Galarza JSG, Teixeira J, Ferreira HJ, Neves C, Francoy TM, et al. Wing Geometric Morphometrics of Workers and Drones and Single Nucleotide Polymorphisms Provide Similar Genetic Structure in the Iberian Honey Bee (*Apis mellifera iberiensis*). Insects. 2020;11(2):89. doi: 10.3390/insects11020089
123. Sumruayphol S, Siribat P, Dujardin J, Dujardin S, Komalamisra C, Thaenkham U. *Fasciola gigantica*, *F. hepatica* and *Fasciola* intermediate forms: geometric morphometrics and an artificial neural network to help morphological identification. PeerJ. 2020;8:e8597. doi: 10.7717/peerj.8597
124. Santos SR, Pessôa LM, Vianna M. Geometric morphometrics as a tool to identify species in multispecific flatfish landings in the Tropical Southwestern Atlantic. Fisheries Research. 2019;213:190–195. doi: 10.1016/j.fishres.2019.01.017
125. Torres SKM, Santos BS. Species Identification Among Morphologically-Similar *Caranx* species. Turkish Journal of Fisheries and Aquatic Sciences. 2020;20:159–169. doi: 10.4194/1303-2712-v20_2_08
126. Costa C, Cataudella S. Relationship between shape and trophic ecology of selected species of Sparids of the Caprolace coastal lagoon (Central Tyrrhenian sea). Environ Environmental Biology of Fishes. 2007;78(2):115–23. doi: 10.1007/s10641-006-9081-9
127. Serb JM, Alejandrino A, Otárola-Castillo E, Adams DC. Morphological convergence of shell shape in distantly related scallop species (Mollusca: Pectinidae). Zoological Journal of the Linnean Society. 2011;163(2):571–84. doi: 10.1111/j.1096-3642.2011.00707.x
128. Cooper WJ, Albertson RC. Quantification and variation in experimental studies of morphogenesis. Developmental Biology. 2008;321(2):295–302. doi: 10.1016/j.ydbio.2008.06.025
129. Zelditch ML, Sheets HD, Fink WL. The ontogenetic dynamics of shape disparity. Paleobiology. 2003;29(1):139–56. doi: 10.1666/0094-8373(2003)029<0139:TODOSD>2.0.CO;2
130. Ubukata T. A morphometric study on morphological plasticity of shell form in crevice-dwelling Pterioida (Bivalvia). Biological Journal of the Linnean Society. 2003;79(2):285–97. doi: 10.1046/j.1095-8312.2003.00144.x
131. Silva IC, Alves MJ, Paula J, Hawkins SJ. Population differentiation of the shore crab *Carcinus maenas* (Brachyura: Portunidae) on the southwest English coast based on genetic and morphometric analyses. Scientia Marina. 2010;74(3):435–44. doi: 10.3989/scimar.2010.74n3435
132. Scalici M, Colamartino M, Spani F, Traversetti L, Persichini T, Maisano M, et al. Integrated early warning systems in marine bivalves reveal detrimental alterations of coastal habitats. Hydrobiologia. 2020;847(11):2573–85. doi: 10.1007/s10750-020-04275-1
133. Mitteroecker P, Bookstein F. Linear Discrimination, Ordination, and the Visualization of Selection Gradients in Modern Morphometrics. Evolutionary Biology. 2011;38(1):100–14. doi: 10.1007/s11692-011-9109-8
134. Wilke ABB, Christe RdO, Multini LC, Vidal PO, Wilk-da-Silva R, de Carvalho GC, et al. Morphometric Wing Characters as a Tool for Mosquito Identification. PLOS ONE. 2016;11(8):e0161643. doi: 10.1371/journal.pone.0161643
135. Lei M, Yu X, Li M, Zhu W. Geographic origin identification of coal using near-infrared spectroscopy combined with improved random forest method. Infrared Physics & Technology. 2018;92:177-182. doi: 10.1016/j.infrared.2018.05.018
136. Kowalewski M. Morphometric analysis of predatory drillholes. Palaeogeography, Palaeoclimatology, Palaeoecology. 1993;102(1–2):69–88. doi: 10.1016/0031-0182(93)90006-5
137. Slice, D. Geometric morphometrics. Annual Review of Anthropology. 2007;36:261–281. doi: 10.1146/annurev.anthro.34.081804.120613

138. Belgiu M, Drăguţ L. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing. 2016;114:24-31. doi: 10.1016/j.isprsjprs.2016.01.011
139. Schonlau M, Zou RY. The random forest algorithm for statistical learning. The Stata Journal. 2020; 20(1);3–29. doi: 10.1177/1536867X20909688
140. Biau G, Scornet E. A random forest guided tour. TEST. 2016;25:197–227. doi: 10.1007/s11749-016-0481-7
141. Oliveira MM, Camanho AS, Walden JB, Miguéis VL, Ferreira NB, Gaspar MB. Forecasting bivalve landings with multiple regression and data mining techniques: The case of the Portuguese Artisanal Dredge Fleet. Marine Policy. 2017;84:110-118. doi: 10.1016/j.marpol.2017.07.013
142. Breiman, L. Random Forests. Machine Learning. 2001;45:5–32. doi: 10.1023/A:1010933404324
143. Franklin J, Miller JA. Mapping Species Distributions – Spatial Inference and Prediction. 1st ed. Cambridge, UK: Cambridge Academic Press; 2010
144. Darr A, Gogina M, Zettler ML. Detecting hot-spots of bivalve biomass in the south-western Baltic Sea. Journal of Marine System. 2014;134:69-80. doi: 10.1016/j.jmarsys.2014.03.003
145. Du P, Samat A, Waske B, Liu S, Li Z. Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features. ISPRS Journal of Photogrammetry and Remote Sensing. 2015;105:38-53. doi: 10.1016/j.isprsjprs.2015.03.002
146. Pal M. Random forest classifier for remote sensing classification. International Journal of Remote Sensing. 2005;26(1):217–222. doi: 10.1080/01431160412331269698
147. Paredes-Villanueva K, Espinoza E, Ottenburghs J, Sterken MG, Bongers F, Zuidema PA. Chemical differentiation of Bolivian *Cedrela* species as a tool to trace illegal timber trade. Forestry: An International Journal of Forest Research. 2018;91(5):603–613. doi: 10.1093/forestry/cpy019
148. Kijewski T, Zbawicka M, Strand J, Kautsky H, Kotta J, Rätsep M, et al. Random forest assessment of correlation between environmental factors and genetic differentiation of populations: Case of marine mussels *Mytilus*. Oceanologia. 2019;61(1):131-142. doi: 10.1016/j.oceano.2018.08.002
149. Doyle D, Gammell MP, Nash R. Morphometric methods for the analysis and classification of gastropods: a comparison using *Littorina littorea*. Journal of Molluscan Studies. 2018;84(2):190–197. doi: 10.1093/mollus/eyy010
150. Cole KE, Yaworsky PM, Hart IA. Evaluating statistical models for establishing morphometric taxonomic identifications and a new approach using Random Forest. Journal of Archaeological Science. 2022;143:105610. doi: 10.1016/j.jas.2022.105610
151. Lopes JF, Lopes CL, Dias JM. Climate Change Impact in the Ria de Aveiro Lagoon Ecosystem: A Case Study. Journal of Marine Science and Engineering. 2019;7(10):352. doi: 10.3390/jmse7100352
152. Matias D, Joaquim S, Matias AM, Moura P, de Sousa JT, Sobral P, et al. The reproductive cycle of the European clam *Ruditapes decussatus* (L., 1758) in two Portuguese populations: Implications for management and aquaculture programs. Aquaculture. 2013;406-407:52–61. doi: 10.1016/j.aquaculture.2013.04.030
153. Maia F, Gaspar M. Caso de Estudo: Apanha de bivalves na Ria de Aveiro Distribuição e abundância dos moluscos bivalves com maior interesse comercial na Ria de Aveiro. Instituto Português do Mar e da Atmosfera (IPMA); 2014 Jan.
154. Maia F, Barroso C, Gaspar M. State and trends in abundance, distribution and population size structure of the common cockle *Cerastoderma edule* (Bivalvia: Cardiidae) in Ria de Aveiro lagoon, Portugal. Congresso Português de Malacologia; 2018 Jun 15-16; Funchal; 2018. doi: 10.13140/RG.2.2.18806.14404

155. Carvalho S, Pereira P, Pereira F, de Pablo H, Vale C, Gaspar MB. Factors structuring temporal and spatial dynamics of macrobenthic communities in a eutrophic coastal lagoon (Óbidos lagoon, Portugal). Marine Environmental Research. 2011:71(2):97–110. doi: 10.1016/j.marenvres.2010.11.005
156. Rebelo CFC, Alves CPF, Moiteiro GC, Ezequiel GMG, Brasão IPC, Vasconcelos JVD, et al. Tourism Through the Gaze of Stakeholders: the Case of Óbidos Lagoon in Portugal. Tourism Planning & Development. 2015;12(4):447–462. doi: 10.1080/21568316.2015.1037928
157. Veiga K, Pedro CA, Ferreira SMF, Gonçalves SC. Monitoring metal pollution on coastal lagoons using *Cerastoderma edule*—a report from a moderately impacted system in Western Portugal (Óbidos Lagoon). Environmental Science and Pollution Research. 2019;26(3):2710–2721. doi: 10.1007/s11356-018-3705-4
158. Martins R, Carneiro M, Rebordão FR. Contribuição Para O Conhecimento Das Artes De Pesca Utilizadas Na Lagoa De Óbidos. Publicações avulsasdo IPIMAR. 2007;16:32
159. Silva CSV. Is It Possible For Native And Introduced *Ruditapes* Species To Live In Sympatry? PhD Thesis. Universidade de Aveiro, Portugal. 2017. Available from: http://hdl.handle.net/10773/22820
160. Garaulet LL. Estabelecimento do bivalve exótico *Ruditapes philippinarum* (Adams & Reeve, 1850) no estuário do Tejo: caracterização da população actual e análise comparativa com a congénere nativa *Ruditapes decussatus* (Linnaeus, 1758) e macrofauna bentónica acompanhante. M.Sc. Thesis. Faculdade de Ciências, Universidade de Lisboa, Portugal. 2011. Available from: http://hdl.handle.net/10451/5602
161. de Pablo H, Sobrinho J, Fonteles C, Neves R, Gaspar MB. The Influence of the River Discharge on Residence Time, Exposure Time and Integrated Water Fractions for the Tagus Estuary (Portugal). Frontiers in Marine Science. 2022;8:734814. doi: 10.3389/fmars.2021.734814
162. Fortunato A, Baptista AM, Luettich RA. A three-dimensional model of tidal currents in the mouth of the Tagus estuary. Continental Shelf Research. 1997;17(14):1689–1714. doi: 10.1016/S0278-4343(97)00047-2
163. Santos CSPC. Estado atual das populações de berbigão (*Cerastoderma* spp.) no estuário do Sado. M.Sc. Thesis, Faculdade de Ciências, Universidade de Lisboa, Portugal. 2019. Available from: http://hdl.handle.net/10451/41456
164. Caeiro S, Martins F, Costa MH, Painho M, Neves R. Metodologia de gestão dinâmica do estuário do Sado. 1º. Congresso sobre Planeamento e Gestão do Litoral dos Países de Expressão Portuguesa; 2001; S. Miguel, Açores, Portugal; 2001. p. 10 Available from: http://hdl.handle.net/10400.1/43
165. Fonseca LC, Andrade F, Pinto P. Contribuição para o conhecimento dos povoamentos bentónicos do estuário do Sado (Setúbal, Portugal). 1º Congresso de Áreas Protegidas; 1987; Lisbon, Portugal; 1989. p. 557-565. doi: 10.13140/RG.2.1.2351.8882
166. Caeiro S, Costa MH, Ramos TB, Fernandes F, Silveira N, Coimbra A, et al. Assessing heavy metal contamination in Sado Estuary sediment: An index analysis approach. Ecological Indicators. 2005;5:151–169. doi: 10.1016/j.ecolind.2005.02.001
167. Cabral S, Carvalho F, Gaspar M, Ramajal J, Sá E, Santos C, et al. Non-indigenous species in soft-sediments: Are some estuaries more invaded than others? Ecological Indicators. 2020;110:105640. doi: 10.1016/j.ecolind.2019.105640
168. Coutinho MTP, Brito AC, Pereira P, Gonçalves AS, Moita MT. A phytoplankton tool for water quality assessment in semi-enclosed coastal lagoons: Open vs closed regimes. Estuarine, Coastal and Shelf Science. 2012;110:134–146. doi: 10.1016/j.ecss.2012.04.007
169. Mendes D, Fortunato AB, Bertin X, Martins K, Lavaud L, Nobre Silva A, et al. Importance of infragravity waves in a wave-dominated inlet under storm conditions. Continental Shelf Research. 2020;192:104026. doi: 10.1016/j.csr.2019.104026

170. Alday M, Cearreta A, Freitas MC, Andrade C. Modern and late Holocene foraminiferal record of restricted environmental conditions in the Albufeira Lagoon, SW Portugal. Geologica Acta. 2013;11:75–84. doi: 10.1344/105.000001754
171. Boyden, C. R. Observations on the shell morphology of two species of cockle *Cerastoderma edule* and *C. glaucum*. Zoological Journal of the Linnean Society. 1973;52(4):269–292. doi: 10.1111/j.1096-3642.1973.tb01885.x
172. R Core Team. R: A language and environment for statistical computing. Version 4.2.0 [software]. 2021 [cited 2022 Apr 28]. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/
173. RStudio Team. RStudio: Integrated Development Environment for R. Version 2022.02.2+485 [software]. 2021 [cited 2022 Apr 28]. RStudio, Inc., Boston, MA. Available from: http://www.rstudio.com/
174. Wickham H. ggplot2: Elegant Graphics for Data Analysis. [software]. 2016 [cited 2022 Apr 28]. Springer-Verlag New York. Available from: https://cran.r-project.org/package=ggplot2
175. Wickham H, Bryan J. _readxl: Read Excel Files_. R package Version 1.4.0 [software]. 2022 [cited 2022 Sep 28]. Available from: https://CRAN.R-project.org/package=readxl
176. Fox J, Weisberg S. An {R} Companion to Applied Regression. 3rd edition. Thousand Oaks CA: Sage; 2019. Available from: https://socialsciences.mcmaster.ca/jfox/Books/Companion/
177. Auguie B. _gridExtra: Miscellaneous Functions for "Grid" Graphics_. R package Version 2.3 [software]. 2017 [cited 2022 Jul 30]. Available from: https://CRAN.R-project.org/package=gridExtra
178. The GIMP Development Team. GIMP. Version 2.10.30 [software]. 2019 [cited 2022 Mar 18]. Free Software Foundation. Available from: https://www.gimp.org
179. Rohlf FJ. tpsDig2. Version 2.32 [software]. 2021 [cited 2022 Mar 18]. Stony Brook, NY: Department of Ecology and Evolution, State University of New York. Available from: http://sbmorphometrics.org/index.html
180. Baken EK, Collyer ML, Kaliontzopoulou A, Adams DC. geomorph v4.0 and gmShiny: enhanced analytics and a new graphical interface for a comprehensive morphometric experience. Methods in Ecology and Evolution. 2021;12:2355-2363.
181. Adams DC, Collyer ML, Kaliontzopoulou A, Baken EK. Geomorph: Software for geometric morphometric analyses. R package Version 4.0.2 [software]. 2021 [cited 2022 Apr 28]. Available from: https://cran.r-project.org/package=geomorph
182. Collyer ML, Adams DC. RRPP: Linear Model Evaluation with Randomized Residuals in a Permutation Procedure. [software]. 2021 [cited 2022 Apr 28]. Available from: https://cran.r-project.org/web/packages/RRPP
183. Collyer ML, Adams DC. RRPP: RRPP: An R package for fitting linear models to high-dimensional data using residual randomization. Methods in Ecology and Evolution. 2018;9(2):1772-1779.
184. Wickham H, François R, Henry L, Müller K. _dplyr: A Grammar of Data Manipulation_. R package Version 1.0.9 [software]. 2022 [cited 2022 Apr 28]. Available from: https://CRAN.R-project.org/package=dplyr
185. Wickham H, Girlich M. _tidyr: Tidy Messy Data_. R package Version 1.2.0 [software]. 2022 [cited 2022 Apr 28]. Available from: https://CRAN.R-project.org/package=tidyr
186. Harrell JrF. _Hmisc: Harrell Miscellaneous_. R package Version 4.7-1 [software]. 2022 [cited 2022 Apr 28]. Available from: https://CRAN.R-project.org/package=Hmisc
187. Kassambara A, Mundt F. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package Version 1.0.7 [software]. 2020 [cited 2022 May 11]. Available from: https://CRAN.R-project.org/package=factoextra

188. Schlager S. Morpho and Rvcg – Shape Analysis in R. In Zheng G, Li S, Szekely G, editors. Statistical Shape and Deformation Analysis. Academic Press; 2017. pp. 217–256.
189. Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002;2(3):18-22. Available from: https://CRAN.R-project.org/doc/Rnews/
190. Rohlf FJ, Slice DE. Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. Systematic Biology. 1990;39:40–59. doi: 10.2307/2992207
191. Russo T, Pulcini D, O'Leary Á, Cataudella S, Mariani S. Relationship between body shape and trophic niche segregation in two closely related sympatric fishes. Journal of Fish Biology. 2008;73(4):809–28. doi: 10.1111/j.1095-8649.2008.01964.x
192. Neubauer TA, Harzhauser M, Mandic O. Phenotypic evolution in a venerid bivalve species lineage from the late Middle Miocene Central Paratethys Sea: a multi-approach morphometric analysis. Biological Journal of the Linnean Society. 2013;110(2):320–34. doi: 10.1111/bij.12120
193. Márquez F, González-José R, Bigatti G. Combined methods to detect pollution effects on shell shape and structure in Neogastropods. Ecological Indicators. 2011;11:248–254. doi: 10.1016/j.ecolind.2010.05.001
194. Adams DC, Collyer ML. Phylogenetic ANOVA: Group-clade aggregation, biological challenges, and a refined permutation procedure. Evolution. 2018;72(6):1204–1215. doi: 10.1111/evo.13492
195. Adams DC, Collyer ML. Analysis of character divergence along environmental gradients and other covariates. Evolution. 2007;61(3):510–515. doi: 10.1111/j.1558-5646.2007.00063.x
196. Collyer ML, Adams DC. Analysis of two-state multivariate phenotypic change in ecological studies. Ecology. 2007;88(3):683–692. doi: 10.1890/06-0727
197. Adams DC, Collyer ML. A general framework for the analysis of phenotypic trajectories in evolutionary studies. Evolution: International Journal of Organic Evolution. 2009;63(5):1143–1154. doi: 10.1111/j.1558-5646.2009.00649.x
198. Collyer ML, Adams DC. Phenotypic trajectory analysis: Comparison of shape change patterns in evolution and ecology. Hystrix. 2013;24(1):75–83. doi: 10.4404/hystrix-24.1-6298
199. Adams DC, Collyer ML. Multivariate Phylogenetic Comparative Methods: Evaluations, Comparisons, and Recommendations. Systematic Biology. 2018;67:14–31. doi: 10.1093/sysbio/syx055
200. Guillerme T, Cooper N, Brusatte SL, Davis KE, Jackson AL, Gerber S, et al. Disparities in the analysis of morphological disparity: Analysis of morphological disparity. Biol Lett. 2020;16(7). doi: 10.1098/rsbl.2020.0199
201. Hopkins MJ, Gerber S. Morphological Disparity. In: Nuno de la Rosa L, Müller G, editors. Evolutionary Developmental Biology. Cham: Springer International Publishing; 2017. pp. 1–12. doi: 10.1007/978-3-319-33038-9_132-1
202. Ricardo F, Pimentel T, Moreira A, Rey F, Coimbra MA, Domingues MR, et al. Potential use of fatty acid profiles of the adductor muscle of cockles (*Cerastoderma edule*) for traceability of collection site. Scientific Reports. 2015;5:11125. doi: 10.1038/srep11125
203. Ricardo F, Mamede R, Bruzos AL, Díaz S, Thébault J, da Silva EF, et al. Assessing the elemental fingerprints of cockle shells (*Cerastoderma edule*) to confirm their geographic origin from regional to international spatial scales. Science of The Total Environment. 2022;814:152304. doi: 10.1016/j.scitotenv.2021.152304
204. Christodoulou MD, Clark JY, Culham A. The Cinderella discipline: morphometrics and their use in botanical classification. Botanical Journal of the Linnean Society. 2020;194(4):385–396. doi: 10.1093/botlinnean/boaa055
205. Burdon D, Callaway R, Elliott M, Smith T, Wither A. Mass mortalities in bivalve populations: A review of the edible cockle *Cerastoderma edule* (L.). Estuarine, Coastal and Shelf Science. 2014;150:271–280. doi: 10.1016/j.ecss.2014.04.011

206. Vaz N, Mateus M, Pinto L, Neves R, Dias JM. The Tagus Estuary as a Numerical Modeling Test Bed: A Review. Geosciences. 2020;10(1):4. doi: 10.3390/geosciences10010004
207. Atchley WR, Anderson D. Ratios and the Statistical Analysis of Biological Data. Systematic Zoology. 1978;27(1):71. doi: 10.2307/2412816
208. Marinho TA, Arruda EP. Shell-specific differentiation: how geometric morphometrics can add to knowledge of Macominae species (Tellinidae, Bivalvia). Marine Biodiversity. 2021;51(2):40. doi: 10.1007/s12526-021-01176-x
209. MacLeod N. Generalizing and extending the eigenshape method of shape space visualization and analysis. Paleobiology. 1999;25(1):107–138. doi: 10.1666/0094-8373(1999)0252.3.CO;2
210. Huxley J, Teissier G. Terminology of Relative Growth. Nature. 1936;137:780–781. doi: 10.1038/137780b0
211. Weyl H. Symmetry (Vol. 47). Princeton University Press; 2015
212. Rosen J. Symmetry in Science. In: Symmetry in Science. Praktische Zahnmedizin Odonto-Stomatologie Pratique Practical Dental Medicine. Springer, New York, NY; 1995. pp. 169–183. doi: 10.1007/978-1-4612-2506-5_9
213. Crampton JS. Elliptic Fourier shape analysis of fossil bivalves: some practical considerations. Lethaia. 1995;28:179–86. doi: 10.1111/j.1502-3931.1995.tb01611.x
214. Ryan BD, Bungartz F, Hagedorn G, Rambold G. LIAS glossary – A Wiki-based Online Dictionary for Ascomycete Terminology used by LIAS, the Global Information System for Lichenized and Non-Lichenized Ascomycetes; 2005. [cited 2022 Oct 28] Database: LIAS glossary [Internet]. Available from: http://glossary.lias.net/
215. Brock V, Christiansen G. Evolution of *Cardium* (*Cerastoderma*) *edule*, *C. lamarcki* and *C. glaucum*: studies of DNA-variation. Marine Biology. 1989;102(4):505–511. doi: 10.1007/bf00438352
216. Hummel H, Wolowicz M, Bogaards RH. Genetic variability and relationships for populations of *Cerastoderma edule* and of the *C. Glaucum* complex. Netherlands Journal of Sea Research. 1994;33(1):81–89. doi: 10.1016/0077-7579(94)90053-1
217. MolluscaBase. MolluscaBase. *Cerastoderma lamarcki* (Reeve, 1845); 2022. [cited 2022 Oct 30] Database: WoRMS [Internet]. Available from: https://www.marinespecies.org/aphia.php?p=taxdetails&id=146788
218. Carroll ML, Johnson BJ, Henkes GA, McMahon KW, Voronkov A, Ambrose WG, et al. Bivalves as indicators of environmental variation and potential anthropogenic impacts in the southern Barents Sea. Marine Pollution Bulletin. 2009;59(4-7):193-206. doi: 10.1016/j.marpolbul.2009.02.022
219. Hauser I, Oschmann W, Gischler E. Taphonomic Signatures On Modern Caribbean Bivalve Shells As Indicators Of Environmental Conditions (Belize, Central America). PALAIOS. 2008;23(9):586–600. doi: 10.2110/palo.2007.p07-075r
220. Yan H, Chen J, Xiao J. A review on bivalve shell, a tool for reconstruction of paleo-climate and paleo-environment. Chinese Journal of Geochemistry. 2014;33(3):310–315. doi: 10.1007/s11631-014-0692-0
221. Schöne BR. *Arctica islandica* (Bivalvia): A unique paleoenvironmental archive of the northern North Atlantic Ocean. Global and Planetary Change. 2013;111:199–225. doi: 10.1016/j.gloplacha.2013.09.013
222. Schöne BR, Gillikin DP. Unraveling environmental histories from skeletal diaries — Advances in sclerochronology. Palaeogeography, Palaeoclimatology, Palaeoecology. 2013;373:1–5. doi: 10.1016/j.palaeo.2012.11.026

223. Trofimova T, Milano S, Andersson C, Bonitz FGW, Schöne BR. Oxygen Isotope Composition of *Arctica islandica* Aragonite in the Context of Shell Architectural Organization: Implications for Paleoclimate Reconstructions. Geochemistry, Geophysics, Geosystems. 2018;19(2):453–470. doi: 10.1002/2017gc007239
224. Freitas P, Monteiro C, Butler P, Reynolds D, Richardson C, Gaspar M, et al. An annually-resolved palaeoenvironmental archive for the Eastern Boundary North Atlantic upwelling system: Sclerochronology of *Glycymeris glycymeris* (Bivalvia) shells from the Iberian shelf. 2015 [cited 2022 Nov 07]. Available from: https://meetingorganizer.copernicus.org/EGU2015/EGU2015-10215-1.pdf
225. Freitas P, Richardson C, Chenery S, Monteiro C, Butler P, Reynolds D, et al. Sr/Ca and Mg/Ca in *Glycymeris glycymeris* (Bivalvia) shells from the Iberian upwelling system: Ontogeny and environmental control". 2017 [cited 2022 Nov 07]. Available from: https://meetingorganizer.copernicus.org/EGU2017/EGU2017-15918.pdf
226. Matos FL, Vaz N, Picado A, Dias JM, Maia F, Gaspar MB, et al. Assessment of Habitat Suitability for Common Cockles in the Ria the Aveiro Lagoon Under Average and Projected Environmental Conditions. Estuaries and Coasts. 2022. doi: 10.1007/s12237-022-01136-z
227. Dias JM, Pereira F, Picado A, Lopes CL, Pinheiro JP, Lopes SM, et al. A Comprehensive Estuarine Hydrodynamics-Salinity Study: Impact of Morphologic Changes on Ria de Aveiro (Atlantic Coast of Portugal). Journal of Marine Science and Engineering. 2021;9:234. doi: 10.3390/jmse9020234
228. Dias J, Lopes J, Dekeyser I. Tidal propagation in Ria de Aveiro Lagoon, Portugal. Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere. 2000;25(4):369–374. doi: 10.1016/S1464-1909(00)00028-9
229. Fortunato AB, Nahon A, Dodet G, Pires AR, Freitas MC, Bruneau N, et al. Morphological evolution of an ephemeral tidal inlet from opening to closure: The Albufeira inlet, Portugal. Continental Shelf Research. 2014;73:49–63. doi: 10.1016/j.csr.2013.11.005
230. Andrade C, Marques FMSF, Freitas MC (coordinators). Relatório Final, Projeto Criação e implementação de um sistema de monitorização no litoral abrangido pela área de jurisdição da Administração da Região Hidrográfica do Tejo. Lisbon: FFCUL/APA, I.P.; 2013 Jun. pp. 47.
231. Martins R, Azevedo M, Mamede R, Sousa B, Freitas R, Rocha F, et al. Sedimentary and geochemical characterization and provenance of the Portuguese continental shelf soft-bottom sediments. Journal of Marine Systems. 2012;91(1):41-52. doi: 10.1016/j.jmarsys.2011.09.011
232. Brito P, Amorim A, Andrade C, Freitas MC, Monteiro H. Assessment of morphological change in estuarine systems—A tool to evaluate responses to future sea level rise. The case of the Sado Estuary (Portugal). In II Congresso sobre Planejamento e Gestão da Zona Costeira dos Países de Expressão Portuguesa; 2003 Apr; Recife, Brasil; 2003
233. Guerreiro M, Fortunato AB, Freire P, Rilo A, Taborda R, Freitas MC, et al. Evolution of the hydrodynamics of the Tagus estuary (Portugal) in the 21st century. Revista de Gestão Costeira Integrada - Journal of Integrated Coastal Zone Management. 2015;15(1):65-80.
234. Biguino B, Sousa F, Brito AC. Variability of Currents and Water Column Structure in a Temperate Estuarine System (Sado Estuary, Portugal). Water. 2021;13:187. doi: 10.3390/w13020187
235. Nascimento Â, Biguino B, Borges C, Cereja R, Cruz JPC, Sousa F, et al. Tidal variability of water quality parameters in a mesotidal estuary (Sado Estuary, Portugal). Scientific Reports. 2021;11:23112. doi: 10.1038/s41598-021-02603-6
236. Quintino V, Rodrigues AM. Environment gradients and distribution of macrozoobenthos in three Portuguese coastal systems: Obidos, Albufeira and Alvor. In Ryland JS, Tyler PA, editors. Reproduction, Genetics and Distributions of Marine Organisms: 23rd European Marine Biology Symposium. Olsen & Olsen, Fredensborg; 1989. pp. 469.

237. Oliveira A, Fortunato AB, Rego JRL. Effect of Morphological Changes on the Hydrodynamics and Flushing Properties of the Óbidos Lagoon (Portugal). Continental Shelf Research. 2006;26:917–942. doi: 10.1016/j.csr.2006.02.011
238. Pierce SE, Angielczyk KD, Rayfield EJ. Patterns of morphospace occupation and mechanical performance in extant crocodilian skulls: A combined geometric morphometric and finite element modeling approach. Journal of Morphology. 2008;269(7):840–864. doi: 10.1002/jmor.10627
239. Carvalho FB. Estado atual da população de amêijoa-japonesa (*Ruditapes philippinarum*) do estuário do Tejo e impactes da sua introdução. M.Sc. Thesis. Faculdade de Ciências, Universidade de Lisboa, Portugal. 2017. Available from: http://hdl.handle.net/10451/27642
240. Göldner D, Karakostis FA, Falcucci A. Practical and technical aspects for the 3D scanning of lithic artefacts using micro-computed tomography techniques and laser light scanners for subsequent geometric morphometric analysis. Introducing the StyroStone protocol. PLoS ONE. 2022;17(4):e0267163. doi: 10.1371/journal.pone.0267163
241. Monna F, Messaoud RB, Navarro N, Baillieul S, Sanchez L, Loiodice C, et al. Machine learning and geometric morphometrics to predict obstructive sleep apnea from 3D craniofacial scans. Sleep Medicine. 2022;95:76–83. doi: 10.1016/j.sleep.2022.04.019

7. Annexes

Annex I. The Procrustes Analysis of variance (ANOVA) performed using the "geomorph" package [180-183] is represented as a table in RStudio [173] and is recognized as a table of statistics which includes sums-of-squares, mean squares, F-statistics, and P-values. In this table, such statistics can be applied for different model terms and calculated for different types of sums-of-squares [180, 181]. This ANOVA is based on the package "RRPP" [182, 183] and generalises the statistics to multivariate data so there is no fundamental difference in how univariate and multivariate statistics are calculated [180, 181]. It has been concluded through multiple studies that, for linear models, randomization of residuals (vectors or distances) in a permutation procedure is a reliable method for generating empirical distributions of linear model statistics (*e.g.*, [183, 194, 199]). The nonparametric procedure in the package [182, 183] used for this analysis randomized residuals from null linear models and, when applied in a systematic way, can calculate P-values for various linear model designs, since it holds constant the effects of null models rather than conflate them with the effects that are tested [183]. It has been proven effective for ordinary least-squares estimation of linear model coefficients [183, 194], which is the case of this study. Moreover, the statistical properties of the "RRPP" package [182, 183] have largely been validated (*e.g.*, [94, 183, 199]), making it an ideal tool to use for this analysis.

Annex II. Each specimen's thin-plate spline deformation from the consensus configuration can be decomposed into several components known as principal warps [83, 126, 128, 130]. Scores of each individual on the principal warps, called partial warps, can then be used to calculate relative warps (principal component vectors of the partial warps) [72, 86, 97, 106, 108, 116, 128]. Since the data set undergoes a decomposition that generates a new set of axes along which shape is partitioned as independent components of the shape variation in the original data set, relative warps analysis is an eigen-vector technique [97, 98]. It is used to express and analyse the within-species major trends variation of the structure's form as a deformation over the consensus (*e.g.*, [68, 98, 102, 126, 130, 131, 191]). If, when calculating partial warps, spatial focus (α) equals to zero, then the relative warps analysis resultant from it is considered analogous to Principal Component Analysis (PCA), where the partial warps are shape components of Procrustes analysis (*e.g.*, [68, 83, 86, 97, 116, 126, 130, 191]).

Annex III. Mean height, length and width (in millimetres) of *Cerastoderma edule* and *Cerastoderma glaucum* at the studied systems (Ria de Aveiro, the Óbidos lagoon, the Tagus estuary, the Albufeira lagoon, the Sado estuary) (mean and respective standard error. 95%).

	C. edule			*C. glaucum*		
	Height (mm) ± SE95%	Length(mm) ± SE95%	Width (mm) ± SE95%	Height (mm) ± SE95%	Length(mm) ± SE95%	Width (mm) ± SE95%
Ria de Aveiro	28.604 ± 1.766	31.798 ± 2.105	22.682 ± 1.425	29.519 ± 1.660	33.324 ± 1.929	23.811 ± 1.621
Óbidos lagoon	25.119 ± 1.631	28.712 ± 1.820	19.343 ± 1.631	25.602 ± 1.243	29.713 ± 1.880	19.990 ± 1.640
Tagus estuary	20.350 ± 2.172	22.877 ± 2.324	16.140 ± 1.988	21.500 ± 1.845	24.206 ± 2.051	17.456 ± 1.648
Albufeira lagoon	24.092 ± 1.837	27.678 ± 2.220	18.544 ± 1.627	23.553 ± 1.383	27.133 ± 1.831	18.541 ± 1.290
Sado estuary	24.572 ± 1.443	27.530 ± 1.839	18.988 ± 1.262	25.474 ± 1.633	28.928 ± 2.327	20.246 ± 1.267