Cancer cells accumulate mutations that allow them to grow uncontrollably and eventually acquire the ability to metastasize, that is, spread to other parts of the body. Transmissible or contagious cancers are large-scale metastases in which the cancer cells spread to other individuals beyond the body that originated them. This doctoral thesis provides further insights into the evolution of transmissible cancers in bivalves through the inspection of 7,290 cockles and clams and genomic and transcriptomic analyses of 643 bivalves. The findings reported include multiple mitochondrial horizontal transfers, co-infections of two contagious cancer lineages affecting a single individual, histogenesis for two independent cancer lineages and the description of a novel interspecific contagious cancer. Enjoy the reading!

Evolution of Bivalve Transmissible Cancers

THESIS

Alicia L. Bruzos

2022

INTERNATIONAL DOCTORAL SCHOOL OF THE USC

Alicia L.
Bruzos

PhD Thesis

Evolution of Bivalve Transmissible Cancers

Santiago de Compostela, 2022

Doctoral Programme in Molecular Medicine

DOCTORAL THESIS

EVOLUTION OF BIVALVE TRANSMISSIBLE CANCERS

Alicia L. Bruzos

INTERNATIONAL PHD SCHOOL OF THE UNIVERSITY OF SANTIAGO DE COMPOSTELA
PHD PROGRAMME IN MOLECULAR MEDICINE

SANTIAGO DE COMPOSTELA
2022

DECLARACIÓN DE LA AUTORA DE LA TESIS

Doña Alicia López Bruzos

Título de la tesis: *Evolution of bivalve transmissible cancers*

Presento mi tesis, siguiendo el procedimiento adecuado al Reglamento y declaro que:

1) La tesis abarca los resultados de la elaboración de mi trabajo.
2) De ser el caso, en la tesis se hace referencia a las colaboraciones que tuvo este trabajo.
3) Confirmo que la tesis no incurre en ningún tipo de plagio de otros autores ni de trabajos presentados por mí para la obtención de otros títulos.
4) La tesis es la versión definitiva presentada para su defensa y coincide la versión impresa con la presentada en formato electrónico.

Y me comprometo a presentar el Compromiso Documental de Supervisión en el caso que el original no esté depositado en la Escuela.

En Londres, a 10 de junio de 2022.

Fdo. **Alicia L. Bruzos**

Firma electrónica

AUTORIZACIÓN DEL DIRECTOR / TUTOR DE LA TESIS

Evolution of bivalve transmissible cancers.

Dr. Xulio Maside Rodríguez en calidad de tutor

Dra. Seila María Díaz Costas en calidad de directora

INFORMAN:

Que la presente tesis, se corresponde con el trabajo realizado por Doña Alicia López Bruzos, bajo nuestra dirección/tutorización, y autorizamos su presentación, considerando que reúne los requisitos exigidos en el Reglamento de Estudios de Doctorado de la USC, y que como director de esta no incurre en las causas de abstención establecidas en la Ley 40/2015.

De acuerdo con lo indicado en el Reglamento de Estudios de Doctorado, declaramos también que la presente tesis doctoral es idónea para ser defendida en base a la modalidad de **Monográfica con reproducción de publicaciones**, en los que la participación de la doctoranda fue decisiva para su elaboración y las publicaciones se ajustan al Plan de Investigación.

En Santiago de Compostela, 10 de junio de 2022.

Fdo. Seila María Díaz Costas
Firma electrónica

Fdo. Xulio Maside Rodríguez
Firma electrónica

To all the women who fight for a better world
and, in particular, to my mother.

A todas las mujeres que luchan por un mundo mejor
y, en particular, a mi madre.

Contents

Abstract

Short abstract[1]

English version

Cancer cells accumulate mutations that allow them to grow uncontrollably and eventually acquire the ability to metastasize, that is, spread to other parts of the body. Transmissible or contagious cancers, which are particularly frequent among bivalves, are large-scale metastases in which the cancer cells spread to other individuals beyond the body from which they originated them (*Chapter 1*). In common cockles, two phenotypically different contagious cancer lineages have been described by means of nuclear and mitochondrial DNA variation in a single Spanish location. In this thesis, we report the prevalence rates of 36 populations and 6,719 cockles alongside the distribution range of the species, and we unravel and characterize multiple mitochondrial horizontal transfers by studying the evolutionary history of healthy and cancer individuals, further describing various co-infections of two contagious cancer lineages affecting a single individual (*Chapter 2*). RNA revealed the same histogenesis for two independent cancer lineages pointing to the potential cancer susceptibility of haemolymph (*Chapter 3*). Finally, to investigate the limits of marine contagious cancers, we collected 345 warty venus clams for which we described a contagious cancer present in two distant locations that originated in a different species, the striped venus clam (*Chapter 4*). In summary, this doctoral thesis advances in the understanding of bivalve transmissible cancers providing novel insights and a robust evolutionary framework of mitochondrial horizontal transfer.

Keywords: marine contagious cancer; bivalve transmissible neoplasia; mitochondrial captures; horizontal transfer; histogenesis; interspecies cancer transmission.

[1] As this thesis has relied on the collaboration of people from many countries to whom I am deeply grateful, I have tried to translate the short abstract to their languages; my apologies to the collaborators for whom I did not include the abstract in their native language. The translation of short abstracts of my doctoral thesis has been supported by the following people: English version by Satymaanasa Polubothu; Galician version by Sergio Couso Núñez, Beatriz López Bruzos and Alicia Bruzos Pérez; French version by Aimie Sauvadet; Portuguese version by Sara Rocha and Nicole Knoepfel; Italian version by Adriana Anido and Davide Zecchin; German version by Sarah Bott, Jürgen Bott and Nicole Knoepfel; Russian version by Maria Skazina; Korean version by Yunah Lee and Hansol Park.

Versión en *Galego*
(versión extendida nos apéndices)

As células cancerosas acumulan mutacións que lles permiten medrar sen control e, eventualmente, adquiren a capacidade de metastizar, é dicir, espállanse por outras partes do corpo. Os cancros transmisibles ou contaxiosos, que son especialmente frecuentes entre os bivalvos, son metástases a gran escala nas que as células cancerosas se propagan a outros individuos alén do corpo que as orixinou (*Capítulo 1*). En berberechos dunha localidade galega, describíronse dúas liñaxes de cancros contaxiosos mediante ADN nuclear e mitocondrial que son fenotipicamente diferentes. Nesta tese doutoral, descríbese a prevalencia dos cancros contaxiosos en 6.719 berberechos de 36 poboacións ao longo do rango de distribución da especie, desentrañamos e caracterizamos múltiples transferencias horizontais de mitocondrias estudando a historia evolutiva de individuos sans e con cancro e describimos diversas co-infeccións de dous cancros contaxiosos que afectan a un só individuo (*Capítulo 2*). Co ARN revelouse a mesma histoxénese para as dúas liñaxes independentes de cancro que apuntan á potencial susceptibilidade do cancro na hemolinfa destes animais (*Capítulo 3*). Finalmente, para investigar os límites dos cancros contaxiosos mariños, recolleitamos 345 carneiros, nos cales describimos un cancro contaxioso que se orixinou nunha especie diferente, a ameixa chirla, e agora está presente en dous lugares distantes (*Capítulo 4*). En suma, esta tese doutoral avanza na comprensión do cancro transmisible de bivalvos proporcionando un marco evolutivo robusto para a transferencia horizontal de mitocondrias e informando sobre novos achados non coñecidos até agora.

Palabras chave: cancro contaxioso mariño; neoplasia transmisible de bivalvos; capturas mitocondriais; transferencia horizontal; histoxénese; transmisión de cancro entre especies.

Versión en *Español*

Las células cancerosas acumulan mutaciones que les permiten crecer sin control y eventualmente adquieren la capacidad de metastizar, es decir, se diseminan a otras partes del cuerpo. Los cánceres transmisibles o contagiosos, particularmente frecuentes en los bivalvos, son metástasis a gran escala en las que las células cancerosas se propagan a otros individuos más allá del organismo que las originó (*Capítulo 1*). En berberechos de una localidad española se han descrito dos linajes de cáncer contagioso mediante ADN nuclear y mitocondrial que son fenotípicamente diferentes. En esta tesis doctoral, reportamos la prevalencia de estos cánceres en 6.719 berberechos de 36 poblaciones a lo largo del rango de distribución de la especie, desentrañamos y caracterizamos múltiples transferencias horizontales mitocondriales mediante el estudio de la historia evolutiva de individuos sanos y con cáncer y describimos varias co-infecciones de dos cánceres contagiosos que afectan a un solo individuo (*Capítulo 2*). El ARN reveló la misma histogénesis para los dos linajes de cáncer independientes que apuntan a la susceptibilidad potencial al cáncer de la hemolinfa (*Capítulo 3*). Finalmente, para investigar los límites de los cánceres contagiosos marinos, recolectamos 345 escupiñas gravadas para las cuales describimos un cáncer contagioso que se originó en una especie diferente, la almeja chirla y que está presente en dos lugares distantes (*Capítulo 4*). En pocas palabras, esta tesis doctoral avanza en la comprensión de los cánceres transmisible de bivalvos proporcionando un marco evolutivo sólido para la transferencia horizontal mitocondrial e informando sobre nuevos hallazgos no conocidos previamente.

Palabras clave: cáncer contagioso marino; neoplasia transmisible de bivalvos; capturas mitocondriales; transferencia horizontal; histogénesis; transmisión de cáncer entre especies.

Version en *Français*

Les cellules cancéreuses accumulent des mutations qui leur permettent de se développer de manière incontrôlable et finissent par acquérir la capacité de métastaser, c'est-à-dire de se propager vers d'autres parties du corps. Les cancers transmissibles ou contagieux, particulièrement fréquents chez les bivalves, sont des métastases à grande échelle dans lesquelles les cellules cancéreuses se propagent à d'autres individus au-delà de l'organisme qui les a engendrées (*Chapitre 1*). Chez les coques communes, deux lignées cancéreuses contagieuses phénotypiquement différentes ont été décrites par l'étude d'ADN nucléaire et mitochondrial dans un seul endroit espagnol. Dans cette thèse, nous rapportons les taux de prévalence de 36 populations et de 6719 coques au long de l'aire de distribution de l'espèce, nous démêlons et caractérisons de multiples transferts horizontaux mitochondriaux en étudiant l'histoire évolutive d'individus sains et cancéreux et nous décrivons diverses co-infections de deux lignées cancéreuses affectant un même individu (*Chapitre 2*). L'ARN a révélé la même histogenèse pour les deux lignées cancéreuses indépendantes indiquant la sensibilité potentielle de l'hémolymphe au cancer (*Chapitre 3*). Pour étudier les limites des cancers marins contagieux, nous avons collecté 345 praires communes pour lesquelles nous avons décrit un cancer contagieux présent à deux endroits éloignés et provenant d'une espèce différente, la petite praire ou gallinette (*Chapitre 4*). En une phrase, cette thèse progresse dans la compréhension de la néoplasie transmissible des bivalves en fournissant un cadre évolutif robuste du transfert horizontal mitochondrial et en informant sur de nouvelles découvertes non rapportées auparavant.

Mots clés: cancer marin contagieux; néoplasie transmissible bivalve; captures mitochondriales; transfert horizontal; histogenèse; transmission interspécifique du cancer.

Versão em *Português*

As células cancerosas acumulam mutações que lhes permitem crescer descontroladamente e, eventualmente, adquirir a capacidade de metástizar, espalhalhando-se a outras partes do corpo. Os cancros transmissíveis ou contagiosos, particularmente frequentes entre os bivalves, são metástases em grande escala nas quais as células cancerosas se propagam para outros indivíduos além do corpo que as originou (*Capítulo 1*). Em berbigões comuns, duas linhagens de cancro contagioso fenotipicamente diferentes foram descritas por meio de DNA nuclear e mitocondrial em numa única localidade espanhola. Nesta tese, relatamos as taxas de prevalência em 36 populações e 6.719 berbigões ao longo da área de distribuição da espécie, desvendamos e caracterizamos múltiplas transferências horizontais mitocondriais estudando a história evolutiva de indivíduos saudáveis e de indivíduos afectados e descrevemos vários casos de co-infecção de duas linhagens de cancro num único indivíduo (*Capítulo 2*). O RNA revelou a mesma histogênese para duas linhagens de cancro independentes, apontando para a potencial suscetibilidade da hemolinfa ao processo cancerígeno (*Capítulo 3*). Para investigar os limites dos cancros contagiosos marinhos, coletamos 345 moluscos de ameijoa Pé-de-burro nos quais descrevemos um cancro contagioso presente em duas localidades afastadas, e que se originou numa espécie diferente, conhecida em Portugal como Pé-de-burrinho (*Capítulo 4*). Em poucas palavras, esta tese avança na compreensão da neoplasia transmissível de bivalves fornecendo uma estrutura evolutiva robusta de transferência horizontal mitocondrial e informando sobre novos achados não relatados anteriormente.

Palavras-chave: câncer contagioso marinho; neoplasia transmissível bivalve; capturas mitocondriais; transferência horizontal; histogênese; transmissão interespécies de câncer.

Versione *Italiana*

Le cellule tumorali accumulano mutazioni che consentono loro di crescere in modo incontrollato e di ottenere la capacità di metastatizzare, cioè di diffondersi in altre parti del corpo. I tumori trasmissibili, detti anche contagiosi, particolarmente comuni nei molluschi bivalvi, sono metastasi su larga scala in cui le cellule tumorali si diffondono ad altri soggetti distinti dall'organismo di origine (Capitolo 1). Nelle telline comuni, sono stati descritte due linee evolutive fenotipicamente diverse di cancro contagioso grazie alle variazioni del DNA nucleare e mitocondriale documentate in una singola località in Spagna. In questa tesi di dottorato, riportiamo la prevalenza di questi tumori in 6.719 telline provenienti da 36 popolazioni nell'areale di distribuzione della specie, sveliamo e caratterizziamo diversi trasferimenti mitocondriali orizzontali studiando la storia evolutiva degli individui sani e di quelli malati di cancro, e descriviamo diverse co-infezioni di due tumori contagiosi che colpiscono un singolo individuo (Capitolo 2). L'RNA ha rivelato la stessa istogenesi per le due linee tumorali indipendenti, sottolineando la potenziale predisposizione al tumore dell'emolinfa (Capitolo 3). Infine, per valutare i limiti di diffusione dei cancri contagiosi in mare, abbiamo raccolto 345 vongole *Venus verrucosa*. Per questi molluschi abbiamo descritto un cancro contagioso in due località distanti l'una dall'altra e che si è originato in una specie diversa, la vongola *Chamelea gallina* (Capitolo 4). In breve, questa tesi di dottorato accresce la nostra conoscenza dei tumori trasmissibili dei bivalvi, offrendo un solido quadro evolutivo per il trasferimento mitocondriale orizzontale e descrivendo nuove scoperte.

Parole chiave: cancro contagioso di mare; neoplasia trasmissibile di bivalvi; trasferimento mitocondriale; transferimento orizzontale; istogenesi; diffusione del cancro tra le specie.

***Deutsche* Fassung**

Krebszellen sammeln mehrere Mutationen an die es ihnen ermöglichen unkontrolliert zu wachsen. Eventuell erlangen sie auch die Metastasekapazität, das heißt, sie breiten sich auf andere Teile des Körpers aus. Ansteckende oder übertragbarer Krebsartenq, die bei Muscheln besonders häufig vorkommen, sind in großem Ausmaß Metastasen denen sich die Krebszellen außerhalb des ursprünglichen Individuum verbreiten (Kapitel 1). In Herzmuscheln von einem spanischen Ort wurden zwei phänotypisch unterschiedliche ansteckende Krebslinien anhand von Nuklearer und Mitochondrialer DNA-Variation beschrieben. In dieser Doktorarbeit: berichten wir über des Vorherrschen von diesen Krebsarten in 6.719 Herzmuscheln von 36 Populationen entlang des Verbreitungsgebiets der Art, entschlüsseln und charakterisieren wir mehrere mitochondriale horizontale Übertragungen, indem wir die Evolutionsgeschichte von gesunden und Krebs-Individuen untersuchen; und beschreiben wir verschiedene Co-Infektionen von zwei ansteckenden Krebslinien in ein einzigen Organismus (Kapitel 2). Die RNA zeigte die gleiche Histogenese für beide unabhängige Krebslinien, die auf die potenzielle Krebsanfälligkeit von Hämolymphe hindeutet (Kapitel 3). Um die Grenzen vom ansteckenden Meereskrebs zu erforschen, sammelten wir letztendlich 345 raue Venusmuschel in denen wir einen ansteckenden Krebs beschrieben haben der von einer anderen, in zwei entfernte Orte vorkommende Art stammt, die Gemeine Venusmuschel (Kapitel 4). Kurz gefasst, diese Doktorarbeit fördert das Verständnis von übertragbarem Muschelkrebs, liefert einen robusten evolutionären Rahmen für den mitochondrialen horizontalen Transfer und informiert über neue Erkenntnisse, die bisher nicht bekannt waren.

Schlüsselwörter: ansteckender Meereskrebs; übertragbare Neoplasie von Muscheln; mitochondriale Fänge; horizontale Übertragung; Histogenese; Übertragung von Krebs zwischen Arten.

한국어 버전

암세포는 돌연변이의 축적으로 인해 세포 증식이 조절되지 않고 주변 조직 혹은 다른 장기로 전이하는 능력을 갖는다. 전염성 암은 특히 이매패류(二枚貝類)에서 많이 발생한다고 알려져 있는데, 암세포가 개체 내에서 뿐만 아니라 다른 개체로 이동하는 대규모 전이 양상을 보인다 (1 장). 스페인에서 발견된 새조개 암의 핵 및 미토콘드리아 DNA 를 분석한 결과, 표현형이 다른 두 개의 독립적인 암 계통이 존재하는 것으로 나타났다. 본 논문에서 우리는 36 가지 개체군에 속하는 새조개 6,719 개의 암 유병률을 추정한다. 또한, 암 및 정상 새조개를 비교 분석하여 암세포의 진화 역사를 연구하고 다중적인 미토콘드리아 수평 이동을 밝힘으로써, 하나의 새조개가 독립적인 두개의 암세포에 동시에 감염된 현상을 보고한다 (2 장). 추가적으로, 두개의 독립적인 암 계통에 대한 RNA 분석을 통해, 두 계통이 모두 혈림프 조직에서 유래되었음을 확인했다 (3 장). 또한, 멀리 떨어진 두 지역에서 수집한 345 개의 사마귀금성조개의 암이 다른 종인 줄무늬금성조개에서 기원했음을 확인함으로써, 해양 전염성 암이 종의 한계를 뛰어넘어 종간 암 전파가 가능하다는 사실을 밝혔다 (4 장). 본 박사 학위 논문은 미토콘드리아 수평 이동의 진화 체계 및 추가적인 발견을 제시함으로써 해양 전염성 암에 대한 이해를 넓히는 데 의의가 있다.

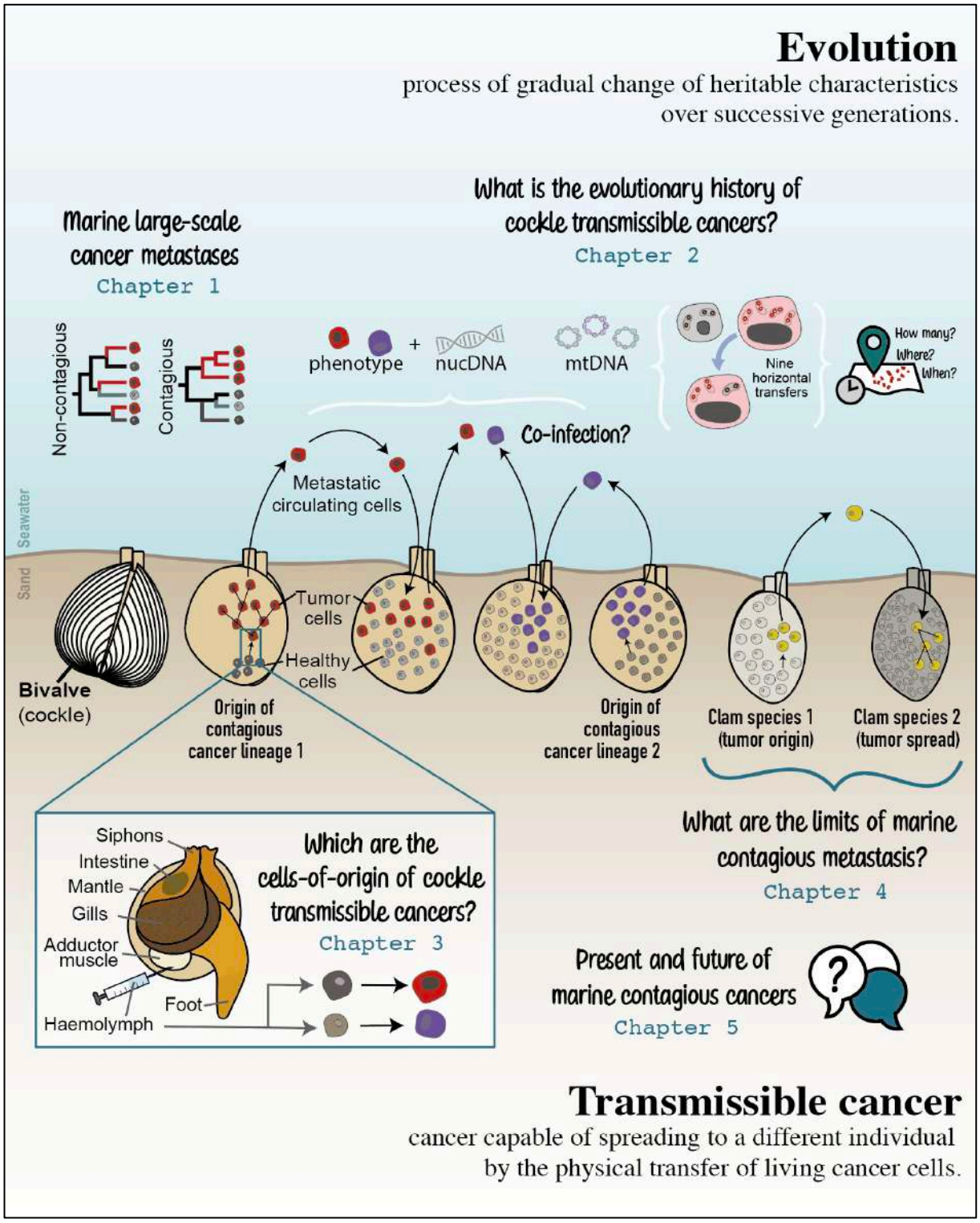
키워드: 해양 전염성 암; 이매패류 전염성 신생물; 미토콘드리아 이입; 수평 이동; 조직 형성; 종간 암 전파.

Русская версия

Раковые клетки способны накапливать мутации, что позволяет им бесконтрольно делиться и в конечном итоге приобретать способность к метастазированию, то есть распространению в другие ткани и органы. Трансмиссивный или контагиозный рак наиболее часто встречается у морских двустворчатых моллюсков. Такой рак представляет собой особую форму метастаз, когда раковые клетки распространяются за пределы организма, в котором они возникли (Глава 1). При анализе ядерной и митохондриальной ДНК у обыкновенной сердцевидки *Cerastoderma edule* из одной испанской популяции были обнаружены две фенотипически различные линии трансмиссивного рака. В данной диссертации приведены результаты следующих исследований. Изучена заболеваемость трансмиссивным раком на примере 6719 сердцевидок из 36 географических популяций в пределах обширного ареала моллюска. В рамках анализа эволюционной истории здоровых и больных особей обнаружены и охарактеризованы множественные горизонтальные переносы митохондриальной ДНК. Описаны различные инфекции сопутствующие заражению моллюска той или иной линией трансмиссивного рака (Глава 2). Анализ РНК показал единое происхождение раковых клеток для двух независимых линий. Так же было показано, что гемолимфа является потенциальной тканью-предшественником раковых клеток у двустворчатых моллюсков (Глава 3). Наконец, для исследования распространения трансмиссивного рака двустворчатых моллюсков было собрано 345 особей *Venus verrucosa*. У этих моллюсков описан трансмиссивный рак в двух географически удаленных популяциях и произошедший от другого вида *Chamelea gallina* (Глава 4).
Результаты исследования в рамках данной диссертации содержат новые данные, которые существенно расширяют представления о трансмиссивном раке двустворчатых моллюсков и о механизмах горизонтального переноса митохондриальной ДНК.

Ключевые слова: межвидовая передача рака; трансмиссивная неоплазия двустворчатых моллюсков; заимствование митохондрий; горизонтальный перенос; гистогенез; морской трансмиссивный рак.

Graphical abstract



Ad astra per aspera

Agradecimientos

Toda aventura comienza con un sueño y rara es la vez que el plan que se traza al principio es el camino que te llevará a tu destino. Si algo he aprendido a lo largo del viaje de la tesis es a ser flexible con el plan y a adaptarme a nuevos escenarios; todo esto no lo habría conseguido sin mis guías, mi directora de tesis Dra. *Seila* y las personas que de alguna forma me guiaron para que no renunciase a seguir caminando. Gracias por todo lo que me habéis enseñado, esta tesis no existiría sin vuestras ideas y conocimientos; no me refiero sólo a las tareas de campo, laboratorio o bioinformática, sino también a las lecciones de vida que no formaban parte de vuestras obligaciones. Mi futuro profesional, en parte, os lo deberé a vosotres.

Un viaje de 250 páginas empieza siempre con un paso y, mi primer paso lo dieron esas mujeres que rompieron las cadenas para que hoy pueda estar yo escribiendo una tesis doctoral. Con los tejados londinenses a través de mi ventana y tarareando en mi cabeza la canción 'Sister Suffragette' de la película favorita de mi hermana, no puedo escribir estos agradecimientos sin una especial mención a todas las *mujeres* que, sin saberlo, la hicieron realidad.

Y hablando de mujeres extraordinarias, esta tesis no existiría sin la mujer a la que más quiero: mi *madre Alicia*. Gracias por alimentar mis sueños -nunca olvidaré cuando nos fuimos en coche a Salamanca para que yo pudiera dar mi primera charla en un congreso-, gracias por acompañarme incluso cuando medio mundo nos separaba (*id est* Corea del Sur y Seattle), gracias por aguantarme en todas mis caídas a lo largo de estos años, gracias por estar a mi lado, gracias por todo lo que aquí no puedo escribir o acabaría siendo una enciclopedia por tomos. A ti te estaré eternamente agradecida. Gracias, *mamá*.

A todos mis compañeres del laboratorio *Genomes and Disease*, gracias por aguantar mis manías de etiquetar todo y gracias por estar siempre dispuestos a ayudar. En especial, gracias *Seila* por uno de los mejores viajes/estancias de mi vida o por todas las veces que saliste a rescatarme sin pensártelo; gracias *Berni* por aquel chiste de berberechos con mala sangre que me encontré en mi teclado una mañana, por todas nuestras partidas a juegos de mesa o por los paseos que podían acabar en 30 km o en noches en el bosque aderezadas con historias terroríficas; gracias *Ana P.* por esas lecciones de biología humana que aprendíamos tomando té; gracias *Martín* por esas cenas otoñales con la mejor tortilla del mundo y por haber sido una gran persona conmigo; gracias *Nita* por todas esas horas que pasamos procesando muestras o, más bien, buscando arroaces por la ventana, gracias *Jr* por tu paciencia con la máquina del tiempo -un secuenciador del siglo pasado-; gracias *Ana O.* por esas visitas a museos tan extravagantes; gracias *Ana C.*, *Artemis*, *Andrea* y *Iago* por ser les mejores estudiantes que pude mentorizar.

Gracias *Jose* por darme libertad para explorar, organizarme y trabajar en un proyecto extraordinario que realmente disfruté. Gracias además por darme la oportunidad de colaborar en otros proyectos con los que descubrí la magia de la investigación.

Gracias a todas esas personas de otros laboratorios que me acompañasteis en este viaje, en especial, gracias *Rubén* por las escapadas a la playa en la hora de comer y por sacarme una sonrisa hasta en los peores momentos; gracias *Paulo* y *Julia* por esas comidas en los troncos del ECIMAT; gracias *Elisa* por ayudarnos desinteresadamente a montar jeringuillas para poder salir antes a disfrutar del tiempo libre; gracias *Javier* por aquellas limas que todavía hoy recuerdo; gracias *Sara*

por tu paciencia y todo lo que me enseñaste, fuiste como una directora para mí; gracias *Laura* por descubrirme lo mucho que me gusta el sushi y cerrar juntas el karaoke; gracias *Michael, Mark, David* y *Young Seok* por abrirme las puertas de vuestros laboratorios; gracias *Marissa, Rachael, Sam* y *Rich* por aquel domingo navegando en vela por Seattle y por el viaje a las Islas San Juan, gracias *Kieran*, *Sean* y *Jean* por presentarme un cachalote; gracias *Yunah*, *Hansol*, *Jisong* y *Hyein* por regalarme un día de princesas orientales en Gongju y por la fiesta de la fresa en Corea del Sur; gracias *Rubén* por extender más allá de la dirección de mi TFG el apoyo que me brindas. Gracias a todes los que me olvido, a quienes pido disculpas de antemano.

Gracias a mis tres confidentes que nunca dejaron de apoyarme en este largo viaje: mi *hermana Beatriz* y mis dos meigas *Sarah* y *Adriana*. Sin vosotras hace tiempo que este barco habría encallado, muchas gracias por estar ahí. Os quiero mucho. Gracias a mi perro *Murphy* que me ha acompañado en toda y cada una de las páginas de esta tesis recordándome que un paseo es la mejor solución para seguir adelante. Gracias a mi gato *Pitágoras* por enseñarme a relativizar los problemas del día a día.

Gracias a todos mis amigues: desde la *Comunidad Bien Avenida* hasta los *Bichólogos* o los *Biorrachos* pasando por todos esos amigues que hice por el camino y, de alguna forma, me han hecho ser quién soy. Gracias a *Isa* por aquel verano recorriendo los Balcanes, gracias a *Sarah* y *Jacko* por acogernos en Extremadura, gracias a *Juanma* por aventurarte conmigo a llegar al Sahara haciendo autostop, gracias a *Adri* y *Campus* por un atardecer en los tejados de una iglesia perdida por el Gêres-Xurés, gracias a *Rubén* por descubrirme la rana escondida en Salamanca, gracias *Sofía* por aquella nota escrita en una servilleta en un bar de Barcelona, gracias a *Miguel* por un fin de semana en el barco de los guerreros del arcoíris … todas estas aventuras, y muchas otras que me dejo en el tintero, forman parte de esta tesis.

Gracias al equipo de *Scientists Meet Artists* y en particular a Sofía por convertir este proyecto en una obra de arte que acercar a les más pequeñes, gracias a la organización del *ForoAcui* que año tras año me han dado la oportunidad de compartir con ellos lo que hacemos en el laboratorio, gracias a *INICE* por descubrirme una parte muy bonita de la investigación, gracias a todes les niñes y adolescentes que se ilusionaron con la ciencia en mis visitas a centros educativos. En definitiva, gracias a todes los que os cruzasteis en mi camino en alguna de las actividades realizadas durante mi tesis.

Gracias a todos esos *berberechos* y *almejas* que involuntariamente dieron su vida por la ciencia (y por mi tesis).

Gracias a les *miembres del tribunal* y a les *expertes internacionales* que le vais a regalar tiempo de vuestras vidas a esta tesis; espero que el proyecto que os presento os apasione tanto como a mi.

Sin extenderme más, como suele acontecerme, *gracias a ti* por leerme; y no te olvides de soñar, porque todas las aventuras empiezan así y estas páginas hacen realidad uno de mis sueños.

Londres, 24 de mayo de 2022.

Alicia L. Bruzos

P.D.: La lengua española no dispone de un género neutro y utiliza el masculino genérico que, en mi opinión, invisibiliza a las mujeres y refuerza estructuras patriarcales. Creo firmemente que les académiques podemos promover la evolución de la lengua para cambiar esta injusta realidad popularizando el uso del "e" como género neutro y, por ello, estos agradecimientos no siguen la normativa gramatical vigente.

Ad astra per aspera

Acknowledgments

An adventure starts with a dream, and rare is the occasion of the initial drawn plan being the path that will take you to your destination. Throughout the journey of the thesis, I have learnt to be flexible with the plan and to adapt to new scenarios which I would not have achieved without my guides, my thesis supervisor Dr. *Seila* and the people have mentored me to not give up on continuing walking. Thank you for everything you have taught me, this thesis would not exist without your ideas and knowledge; I am not referring only to the field, laboratory, or bioinformatics work, but also to the life lessons that were not part of your duties. In part, I will always owe you my professional future.

A 250-page journey always begins with one step, and my first step was taken by those women who broke the chains so that today I can be writing a doctoral thesis. With London rooftops through my window and the song 'Sister Suffragette' from my sister's favourite film humming in my head, I can't write these acknowledgments without a special mention to all the *women* who unknowingly made it happen.

And speaking of extraordinary women, this thesis would not exist without the woman I love the most: my *mother Alicia*. Thank you for feeding my dreams -I will never forget when we drove to Salamanca so that I could give my first talk at a congress-, thank you for accompanying me even when half the world separated us (*i.e*., South Korea, Seattle), thank you for helping me stand up after all my falls throughout these years, thank you for being by my side, thank you for everything that I cannot write here or it would end up being an encyclopaedia. I will be eternally grateful to you. Thanks *Mum*.

To all my colleagues in the *Genomes and Disease lab*, thank you for acclimating to my obsession of labelling everything and thank you for always willing to help. Especially, thanks *Seila* for one of the best sampling trip of my life or for all the times you came to rescue me without thinking about it; thanks *Berni* for that joke of cockles and bad blood that I found on my keyboard one morning, for all the board games or for the walks that could end in 30 km or nights in the forest seasoned with terrifying stories; thanks *Ana P*. for those human biology lessons that we learned drinking tea; thanks *Martín* for those autumn dinners with the best tortilla in the world and for being a great person with me; thanks *Nita* for all those hours we spent processing samples or, rather whale watching through the window, thanks *Jr* for your patience with the time machine -a sequencer from the last century-; thanks *Ana O.* for those visits to such extravagant museums; thanks *Ana C., Artemis, Andrea* and *Iago* for being the best students I could ever supervise.

Thank you, *Jose*, for giving me the freedom to explore, organize and work on an extraordinary project that I really enjoyed. Thank you also for giving me the opportunity to collaborate on other projects with which I discovered the magic of research.

Thanks to all those people from other laboratories who accompanied me on this trip, especially thanks *Rubén* for the trips to the beach at lunchtime and for making me smile even in the worst moments; thanks *Paulo* and *Julia* for those meals in the ECIMAT logs; thank you *Elisa* for helping us so many times to assemble syringes and leave earlier; thanks *Javier* for those limes that I still remember today; thank you *Sara* for your patience and everything you taught me, you have been a supervisor for me; thank you *Laura* for discovering my love for sushi and for ending parties closing the karaoke together; thank you *Michael*, *Mark*, *David* and *Young Seok* for opening the doors of

your laboratories to me; thank you *Marissa*, *Rachael*, *Sam* and *Rich* for that Sunday sailing around Seattle and the trip to the San Juan Islands, thank you *Kieran*, *Sean* and *Jean* for introducing me to a sperm whale; thank you *Yunah*, *Hansol*, *Jisong* and *Hyein* for an amazing oriental princess day in Gongju and for the strawberry party in South Korea; thank you *Rubén* for extending the support you give me beyond the direction of my degree thesis. Thanks to everyone I forget, to whom I apologize in advance.

Thanks to my three besties who never stopped supporting me on this long journey: my *sister Beatriz* and my two witches' friends *Sarah* and *Adriana*. Without you, this ship would have sunk long ago, thank you very much for being there. I love you. Thanks to my dog *Murphy* who has accompanied me in every page of this thesis reminding me that a walk is the best solution to move forward. Thanks to my cat *Pythagoras* for teaching me to relativize everyday problems.

Thanks to all my friends: from *Comunidad Bien Avenida* to *Bichólogos* or *Biorrachos*, passing through all those friends I made along the way and, in some way, they have made me who I am. Thanks to *Isa* for that summer exploring the Balkans, thanks to *Sarah* and *Jacko* for welcoming us in Extremadura, thanks to *Juanma* for venturing with me to reach the Sahara hitchhiking, thanks to *Adri* and *Campus* for a sunset on the roofs of a lost church in the Natural Park of Gêres-Xurés, thanks to *Rubén* for showing me the hidden frog in Salamanca, thanks *Sofía* for that note written on a napkin in a bar of Barcelona, thanks to *Miguel* for a weekend on the ship of the rainbow warriors… all these adventures, and many others that I left in the inkwell, are part of this thesis.

Thanks to the team of *Scientists Meet Artists* and, in particular, to *Sofía* for converting this project into an artwork for the little ones, thanks to the organization of *ForoAcui* that year after year have given me the opportunity to share with them what we do in the laboratory, thanks to *INICE* for unravelling me a beautiful part of the research, thanks to all the *children and adolescents* who enthusiastically asked during my scientific visits to schools. In short, thanks to all of you who crossed my path in any of the activities carried out during my thesis.

Thanks to all those cockles and clams who unwittingly gave their lives for science (and for my thesis).

Thanks to the *viva examiners* and the *international experts* who are going to give time of your lives to this book; I hope that the project that I present to you within these pages will inspire you as much as it does me.

With no further delay, as it usually happens to me when writing, *thanks to you* for reading me; and don't forget to dream, because all adventures start just like that, and this book make one of my dreams come true.

London, 24th May 2022.

Alicia L. Bruzos

List of abbreviations

A	Adenine
AF	Allele frequency
ANOVA	Analysis of variance
AS	Alignment scrore
AU	Approximately unbiased
BAM	Binary Alignment Map
BI	Bayesian inference
bp	Base pair
BWA	Burrows-Wheeler Aligner
C	Cytosine
Ce	*Cerastoderma edule* (common cockle)
Cg	*Cerastoderma glaucum* (lagoon cockle)
CN	Copy-number
COSMIC	Catalogue of Somatic Mutations in Cancer
CTVT	Canine Transmissible Venereal Tumour
DAPI	4′,6-DiAmidino-2- PhenylIndole
DFTD	Devil Facial Transmissible Disease
DFT1	Devil Facial Transmissible Cancer Lineage 1
DFT2	Devil Facial Transmissible Cancer Lineage 2
dN/dS	Normalised nonsynonymous-to-synonymous substitution ratio
DNA	Deoxyribonucleic acid
DUI	Double uniparental inheritance
eDNA	Environmental DNA
FISH	Fluorescence in situ hybridization
freq	Frequency
G	Guanine
Gb	Giga base (one billion bases)
HDP	Highest posterior density
HN	Hemic neoplasia
HT	Horizontal transfer
H&E	Hematoxylin and eosin
indel	Short insertion/deletion
ITS	Internal transcriber spacer
JAK/STAT	Janus kinase signal transducer and activators of transcription
kb	Kilobase (one thousand bases)
MAPK	Mitogen-activated protein kinase pathway

MAPQ	Mapping quality
Mb	Megabase (one million bases)
MCMC	Markov chain Monte Carlo
MHC	Major histocompatibility complex
miRNA	Small single-stranded non-coding RNA (microRNA)
MNP	Multi-nucleotide polymorphism
mRNA	Messenger RNA
mtDNA	Mitochondrial DNA
ML	Maximum-likelihood
MRCA	Most recent common ancestor
MY	Million years
NGS	Next-generation sequencing
NJ	Neighbour-joining
NUMT	Nuclear mitochondrial DNA
ORF	Open reading frame
PBS	Phosphate Buffer Saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PI	Propidium Iodide
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
RQ	Research question
rRNA	Ribosomal RNA
SDS	Sodium dodecyl (lauryl) sulfate
SH	Shimodaira-Hasegawa
SNP	Single-nucleotide polymorphism
SNV	Single-nucleotide variant
T	Thymine
tRNA	Transfer RNA
unk	Unknown
VAF	Variant allele frequency
VCF	Variant Call Format
WGA	Whole genome amplification
WGS	Whole genome sequencing

Gene abbreviations

16S	Small subunit ribosomal RNA molecules of ribosomes
CL17	Satellites of *Venus verrucosa*
CL4	Satellites of *Venus verrucosa*
CDKN2A/B	Cyclin Dependent Kinase Inhibitor 2A/B
c-MYC	c-myelocytomatosis gene
DEAH12	Gene encoding for an ATP-dependent RNA helicase
EF1α	Eukaryotic Translation Elongation Factor 1 Alpha
ERG	Erythroblast transformation-specific related gene
mt-COI	Mitochondrially Encoded Cytochrome C Oxidase I
mt-CO2	Mitochondrially Encoded Cytochrome C Oxidase II
mt-ND4L	Mitochondrially Encoded NADH:Ubiquinone Oxidoreductase Core Subunit 4L
mt-ND6	Mitochondrially Encoded NADH:Ubiquinone Oxidoreductase Core Subunit 6
TP53	Tumour Protein P53
POLR3C	RNA Polymerase III subunit C
PTEN	Phosphatase and Tensin Homolog
RAS	Family of genes that encode small GTPase proteins discovered originally in Rat sarcoma virus
RASL11A	RAS Like Family 11 Member A
RB1	Retinoblastoma gene 1
SCAMP1	Secretory Carrier Membrane Protein 1
SETD2	SET domain-containing 2
ST8SIA2	ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 2
TFIIH	Transcription Factor II Human-like gene

Chapter cover shows the illustration created by Sofía Venzel for the initiative *Scientists Meet Artists* of Campus do Mar from Universidade de Vigo (Spain). Campus do Mar and the artist have granted written permission to reproduce the drawing in this thesis.

CHAPTER 1.
An introduction to the evolution of bivalve transmissible cancers

"Cancer is not merely a lump in the body; it is a disease that migrates, evolves, invade organs, destroys tissues, and resists drugs. […] *It is the Emperor of All Maladies."* Siddhartha Mukherjee

"Not only can cancer be a contagious disease, but it can also threaten an entire species with extinction." Elizabeth Murchison

Cancer disease was named after the shape of a crab, an animal that we can find in the sand or even inside a cockle (Mukherjee, 2010; Longshaw and Malham, 2013). The Greek word '*karkinos'* meant "crab" and it was first used by Hippocrates around 400 B.C. to refer to a malignant mass proliferation in a human body, then the metaphor was extended even further when Galen described the veins surrounding a breast tumour as crab's legs (Ades, Tryfonidis and Zardavas, 2017).

The Greeks did not have microscopes and therefore, they did not imagine that entities called cells when they lose control of their growth were the leading cause of cancer. However, they did understand some key concepts that influenced medicine over the centuries. Galen suggested that you could cut cancer out but one of the four liquids of the body -the bile- would flow right back, like sap seeping through the limbs of a tree (Mukherjee, 2010). In a vague way, he was describing metastasis, that is the migration of cancer from one site to another.

The science of four millennia separate us from the first description of cancer found in an Egyptian papyrus (Mukherjee, 2010). Though science and genetics have illuminated much about the origin and development of this disease, much remains to be understood, especially regarding the biological mechanisms of the metastatic process (Fares *et al*., 2020).

This chapter introduces core concepts and theories with some bites of history to equip you with a state of the art to dive into the research chapters included in this doctoral thesis.

1.1. THE BASIS OF CANCER

1.1.1. THE CANCER GENOME

Cancer is currently defined as a collection of more than 100 distinct diseases all caused by the uncontrolled growth of cells unleashed by mutation (M. R. Stratton, Campbell and Futreal, 2009). And yet it has been a known disease for millennia, a papyrus with the teachings of Imhotep already described breast cancer in 2625 B.C., but we still lack understanding of the genetic mechanisms that rule this disease of the genome (Mukherjee, 2010).

In the twentieth century, the pathogenesis of cancer was proposed based on pioneering studies of chromosomes from cancer cells (Hansemann, 1890; Boveri, 1914; Mitelman, Obe and Natarajan, 1990), and soon found support in early experimental studies of carcinogenesis (Loeb and Harris, 2008). The origin of cancer happens in a single cell that acquires alterations, or mutations, in its hereditary material. Mutations include a wide range of changes in the DNA sequence, such as single-nucleotide variants (substitutions of a single base), small or large insertions or deletions of DNA sequence, increases or reductions in the number of copies of a DNA segment, rearrangements of the sequence, integration of mobile elements or exogenous DNA (notably viral sequences) or even epigenetic changes (gene activity and expression). While most of this damage is repaired by cellular systems, a small fraction escapes repair and is passed to daughter cells as mutations (Báez, 2019).

Cancer genomics is the study of structure, function and inheritance of the genetic material of cancer cells and how alterations can lead to cancer; which started by looking at the chromosomes under the microscope in the 60's (Lobo, 2008). Cancer cells quickly brought the attention of scientists because the chromosomes of cancers cells are characterized by bizarre chromosomal aberrations compared to healthy [2] cells (Lobo, 2008). In fact, just by comparing the chromosomes of a healthy human cell (Figure 1A) with a cancer cell (Figure 1B), we can rapidly understand that genomes of cancer cells can be incredibly messed up. Thousands of chromosomal aberrations have been discovered in different types of cancer and provide diagnostic tools for cancers such as chronic myeloid leukaemia (Lobo, 2008).

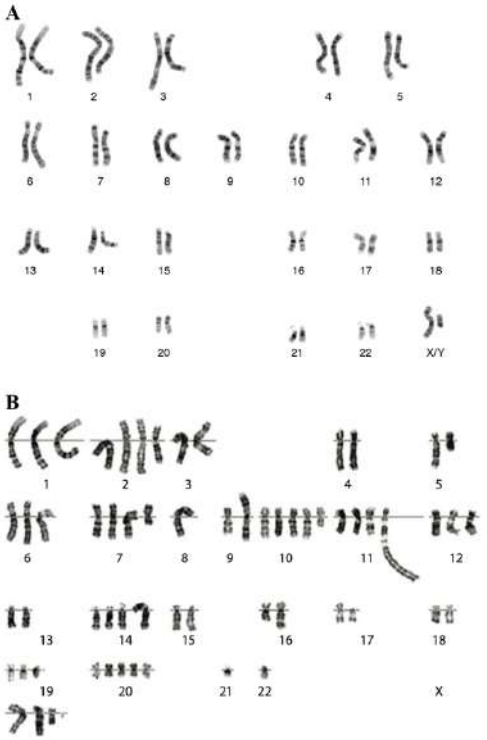


Figure 1. Karyotypes showing in bright field G-banding the complete set of metaphase chromosomes sorted by length and centromere location in **(A)** a human healthy cell (Courtesy: National Human Genome Research Institute, genome.gov) and **(B)** squamous cell carcinoma (adapted from French, 2012; reprinted with permission from Annu. Rev. Pathol. Mech. Dis., see Appendix H).

The genome of a normal cell regulates cell division and death while the genome of a cancer cell turns it into a cell that cannot stop dividing and growing; how exactly do they do this remains unclear. But cancer is not just a genetic disease in its origin, it is genetic in its entirety. Mutant genes do not only allow the uncoordinated proliferation of cells into a mass but promote its survival, accelerate its growth, enable its mobility, recruit blood vessels,

[2] Throughout this thesis, the term 'healthy' is used to refer to non-cancer cells, tissues, or individuals; it could happen that they were affected by other pathologies.

enhance nourishment or draws oxygen. In other words, mutations in the genome of a cancer cell sustain cancer's life (Mukherjee, 2010).

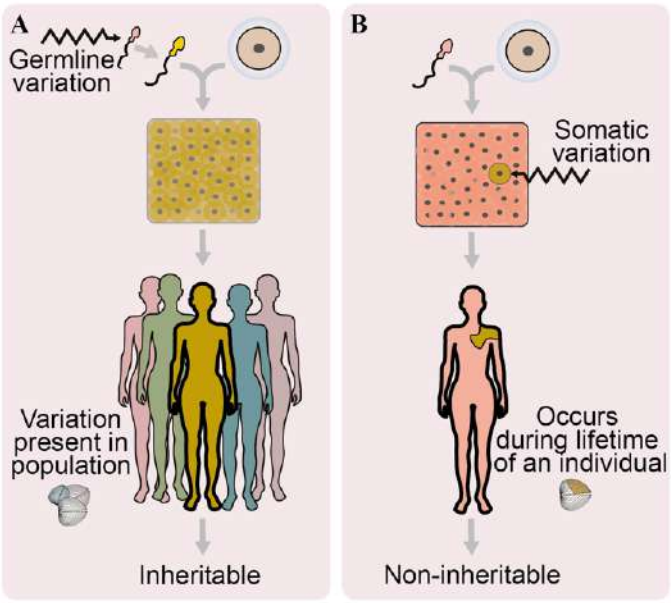


Figure 2. Germline and somatic variation in a population. **(A)** Germline variation is present in the population, and it is heritable. **(B)** Somatic variation is the genetic variation arising in somatic cells during the lifetime of an individual.

Mutations that are inherited from the progenitors are called **germline variants** while the mutations acquired by all type of cells along the organism life are called **somatic variants** (M. Stratton, Campbell and Futreal, 2009; Figure 2). Cancer cells will have germline variants as well as somatic mutations and some of them must be affecting important genes for cancer development (Stratton, 2013). Which genes need to be affected to develop cancer? How many mutated genes are necessary? What are the genetic steps to convert a normal cell into a cancer cell?

In the decade of the 80's and 90's, more than one hundred genes involved in cancer development were identified and categorized as **proto-oncogenes** if their function was involved in normal cell growth or as **tumour suppressor genes** if they normally inhibit growth (Weinberg, 1994). If the genome is so densely composed by these genes waiting to push a cell toward cancer, then why is the human body not exploding with cancer every minute? A single mutation on these genes only produces a step towards cancer, mutations are rare events that need to activate proto-oncogenes and inactivate tumour suppressor genes thus two independent mutations must inactivate each copy of the latter gene category which is even rarer (Haber and Fearon, 1998).

As not every mutation is valid for cancer, in just six rules, Weinberg and Hanahan summarized in 2000 the hallmarks of cancer: (1) acquisition of pathological mitosis to proliferate by the activation of oncogenes, (2) inactivation of tumour suppressor genes that normally inhibit growth, (3) evasion of apoptosis (*ie.* the programmed cell death), (4) limitless replicative potential by activating specific gene pathways to render immortality, (5) capacity to draw out their own supply of blood (*ie.* angiogenesis) and (6) ability to migrate to other organs, invade their tissues and colonize them resulting in the spread of cancer throughout the body (*ie.* metastasis) (Hanahan and Weinberg, 2000).

1.1.2. METASTASIS: THE CANCER JOURNEY

Metastasis is an extremely complex process in which tumour cells escape from the primary site of origin, disseminate to a secondary location, survive, adapt to the new site, and finally colonize and proliferate forming a new tumour while evading immune surveillance (Hunter *et al.*, 2018). The word metastasis is a curious mix of *meta-* ("beyond") and *-stasis* ("stillness") that captures the peculiar instability of this process. Cancer is an expansionist disease that invades through tissues from one organ to another filling the bodies with too many cells: the pathology of excess (Mukherjee, 2010).

When then President Nixon declared war against cancer in 1971, he was probably not expecting such a long war. Certainly, there have been so many major triumphs during the past 50 years, we now have treatments for acute lymphocytic leukaemia and other cancers, as well as the development of methods for early diagnosis that are among the great achievements of modern medicine (Sporn, 1996). However, cancer continues to be a leading cause of death worldwide, accounting for nearly 10 million deaths in 2020 and metastases are the primary cause of cancer deaths (Ferlay *et al.*, 2020). In fact, it is estimated that metastasis is responsible for about 90% of cancer deaths and this estimate has changed little in more than 50 years (Seyfried and Huysentruyt, 2013).

Despite extensive research, the fact that metastatic establishment of cancers at distant organs continues to be largely uncurable attests to the failure in managing the disease once it disseminates through the body (Seyfried and Huysentruyt, 2013; Bergers and Fendt, 2021). This is mainly because most research does not involve metastasis in the *in vivo* state (Seyfried and Huysentruyt, 2013) and the predominant cancer treatments focuses on inhibiting cancer growth, with little emphasis on metastasis (Guan, 2015). Therefore, a better understanding of metastases from various points of view would provide us with new opportunities to develop cancer therapies to try to reduce this high death rates.

Nonetheless, metastasis formation itself is a rare event in tumours because cancer cells need to overcome multiple hurdles before they can successfully proliferate in other organs of the body (Bergers and Fendt, 2021). As for cancer, four hallmarks of metastasis have also been defined to provide conceptual framework and advance in the knowledge of this process (Figure 3). As a first step, cancer cells need to become (1) motile and invasive to enter the stream or route that will get them to other location, either as single cells or collective migration. Once they have migrated, a striking characteristic of metastases is their (2) ability to modulate the new environment, that is negating antitumor actions of the immune system or even manipulating the behaviour of other cancer cells. Likewise, the capacity to grow in more than one location requires the capacity to adapt, that is (3) plasticity. And finally, the dissemination of cancer cells needs to be successful, each cancer cell that disseminates has the potential to metastasize, but it is not yet realized unless they (4) colonize (Welch and Hurst, 2019; Bergers and Fendt, 2021). To cut a long story short, the majority of cancer cells will succumb during their journey, with only a few able to travel and successfully colonize other organs.

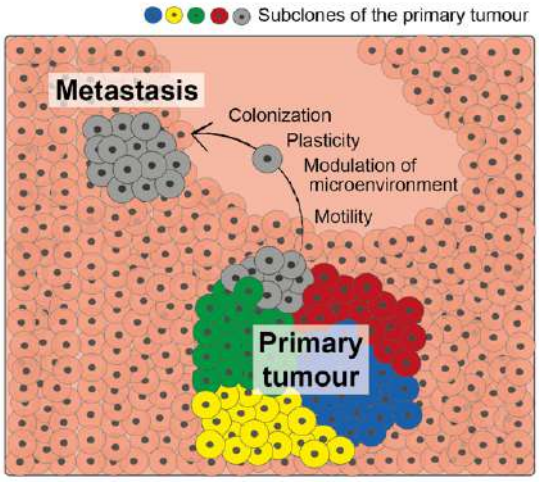


Figure 3. The metastatic establishment of cancer at distant organs in an organism requires four distinguishing features: motility and invasion, ability to modulate the secondary site, plasticity and ability to colonize.

1.1.3. CANCER EVOLUTION

Cancer is a *clonal disease* defining clone as cells that share a common genetic ancestor (Figure 4A). Every known cancer originates from one ancestral cell that, having acquired the mutations of cancer, gives rise to numbers of descendants that will continue to divide (Mukherjee, 2010). But growth without evolution would not make cancer able to invade,

survive and metastasize. Therefore, cancer is not simply a clonal disease, it is a *clonally evolving disease*.

Every generation of cancer cells creates some cells that are genetically different from its parents (Mukherjee, 2010). This result of multilineage somatic evolution of genetically unstable cancer cells it is known as genetic heterogeneity (Figure 4B; Rübben and Araujo, 2017).

Upon activating the cancer hallmarks, cancer cells continue to evolve into a life-threatening metastatic cancer cell. Metastasis is thought to be the ultimate manifestation of a cancer cell's evolution toward becoming autonomous from the host (Welch and Hurst, 2019).

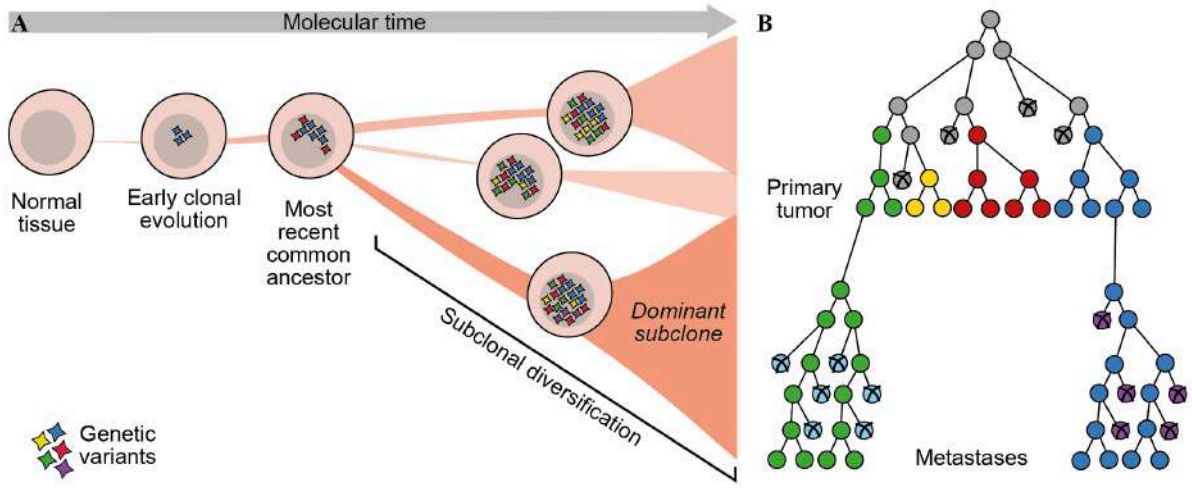


Figure 4. Cancer is a clonal evolving disease. **(A)** Model of clonal evolution of cancer, subclonal diversification is followed by sequential selection of dominant genetic populations (adapted from Nik-Zainal *et al.*, 2012; reprinted with permission from Elsevier CC-BY 3.0, see Appendix H). **(B)** Schematic representation of somatic cancer evolution as a phylogenetic tree where different colours represent subclones and show the heterogeneity in the primary tumour and its metastases (adapted from Rübben and Araujo, 2017; reprinted with permission from Springer Nature CC-BY 4.0, see Appendix H).

Analogously to the Darwinian evolution of species, cancers evolve through the interaction of two processes: acquisition of genetic variation, and the action of natural selection (Báez, 2019).

1. The **acquisition of genetic variation or mutations** occur throughout the lifetime of the individual starting in the fertilised egg (Figure 5). Mutations can occur because of random events (errors of DNA replication), life cell or environmental factors (UV exposure, tobacco…). But most of them will not make any difference at all: they are innocent **passenger mutations** to the cells in which they occur. Passenger mutations provide insights into the underlying mutational processes operative in each case. However, occasionally a mutation does make a difference, it has an effect, it gives the cell an advantage and leads the cells to clonal expansion. These are the ones called **driver mutations** and they are the important ones for cancer development. It is typically thought that you need a handful of these driver mutations to develop a cancer (Hanahan and Weinberg, 2000; M. R. Stratton, Campbell and Futreal, 2009; Stratton, 2013).

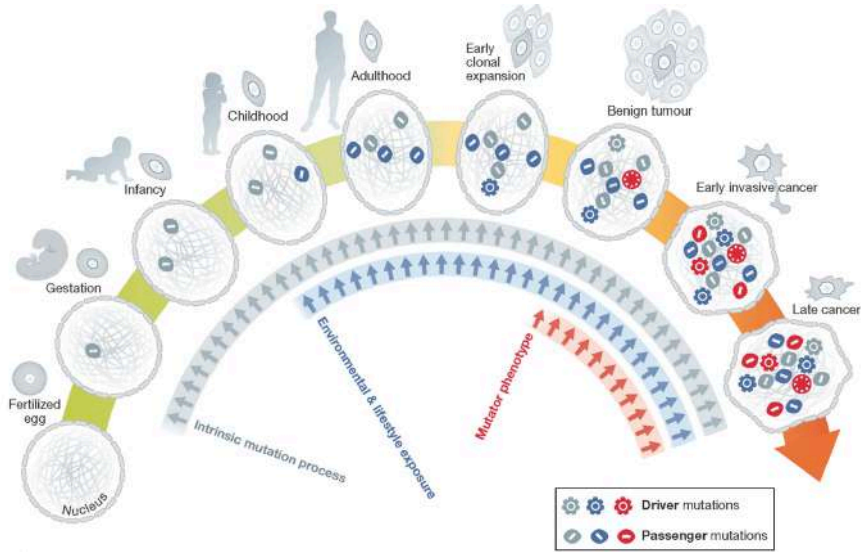


Figure 5. Accumulation of driver and passenger somatic mutations throughout a cellular lineage that connects the fertilized egg and a fully malignant cancer cell. Coloured arrows represent exposure to various types of mutational processes (adapted from Stratton, 2013; reprinted with permission from John Wiley and Sons, Ltd on behalf of EMBO, see Appendix H).

2. The **action of natural selection** on the resultant cells occurs via cell competition and selective pressures from the cellular microenvironment (Stratton, 2013). At a certain time in the tumour's evolutionary history, one cell acquires an additional driver mutation that gives it an advantage over the rest of cancer cells and embarks in a new clonal expansion and leads to a new sub-clone in the tumour that might be more resistant to treatments or have the ability to metastasize. The evolution of a cancer is probably dependent not only on the acquired driver mutations, but also on the genomic landscape, cell type, cellular environment, and tissue architecture wherein such mutations arise. In fact, certain genes are strongly associated with particular cancer types suggesting the existence of cell-type specific evolutionary trajectories leading to malignant phenotypes across tissues (Bailey *et al.*, 2018). The genomic intra-tumour heterogeneity has been largely observed within individual tumours suggesting that cancer development may proceed along a variety of possible evolutionary trajectories (McGranahan and Swanton, 2017; Yates, 2017; Kuo and Curtis, 2018; Ju, 2021).

Understanding cancer from an evolutionary perspective might propose alternative approaches and achieve better therapies (Heng *et al.*, 2011; Gillies, Verduzco and Gatenby, 2012; Enriquez-Navas, Wojtkowiak and Gatenby, 2015).

1.1.4. CANCER ACROSS THE TREE OF LIFE

Cancer has been recognized and defined anthropocentrically by how it appears in humans. Considering cancer-like phenomena broadly including neoplastic growths characterized by abnormal proliferation and differentiation (Aktipis *et al.*, 2015), we can have a look at the tree of life and see that cancer is not only an ancient human's disease (Figure 6).

Cancers have been observed across most vertebrates, regardless of body size and lifespan, although mammals tend to have higher cancer rates than birds or reptiles (Effron, Griner and Benirschke, 1977; Caulin and Maley, 2011) and larger and longer-lived animals, such as whales and elephants, have lower cancer rates than what would be expected given the number of cells and divisions. Only two species of vertebrates stand out having little if any cancer: naked mole rats and blind mole rats (Aktipis *et al.*, 2015).

But cancer is not restrained to vertebrates, it has also been reported in plants, algae, fungi, and invertebrates such as sea urchins, starfish, corals, insects (flies, spiders…), or bivalves

(cockles, clams, mussels, oysters…). In fact, *Drosophila* flies have been widely exploited to study the genetic causes of cancer and have assisted both in deciphering the genetic basis of human cancers and in identifying novel cancer genes. Paradoxically, cancer in crabs, which etymologically name the cancer disease, have rarely been observed (Vogt, 2008).

Animals appear to be more susceptible to cancer than the other branches of the tree, although we cannot completely eliminate the possibility of this conclusion being influenced by sampling bias as there have been vastly more studies of cancer in animals than in other organisms. On the contrary, animals do show higher number of proliferative cells, cell types and higher metabolic rates which increases the cancer risk directly.

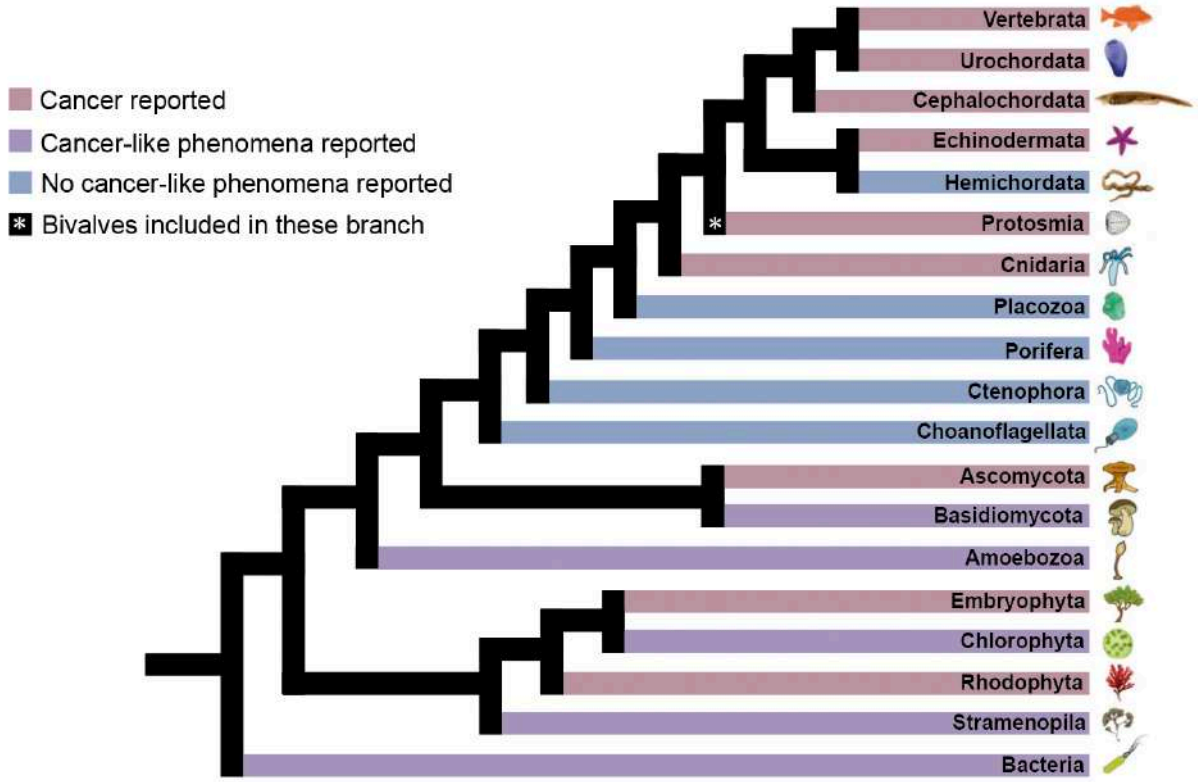


Figure 6. Wildlife cancers reported across the tree of life (adapted from Aktipis et al., 2015; reprinted with permission from Royal Society, see Appendix H). Phylogenetic tree showing the occurrence of cancer. It is not representative of all ancestral states, it is a general representative of major clades. Colour branches represent whether a cancer phenotype -invasion or metastasis- was reported for that lineage (red branches), a cancer-like observation -abnormal proliferation or differentiation- (purple branches) or no cancer-like phenotype has been described (blue branches). White star denotes the branch in which the bivalve molluscs studied in this doctoral thesis are included.

Aside from a few rare observations in plants, metastasis appears to be restricted to animals (Aktipis *et al.*, 2015). At its core, metastasis requires the dissemination of cells away from the originating tumour. Animals also have circulatory systems that transport cells and resources which probably make them more susceptible to metastasis than organisms that only transport resources. Researchers default to think on bloodstream as the route of metastatic spread however, it has been observed that metastatic cells enter not only the cardiovascular system, but some migrate along nerves (Marchesi *et al.*, 2010; Sleeman, Nazarenko and Thiele, 2011), along the basal side of endothelial cells (Lugassy *et al.*, 2004) or through the lymphatic vessels (Welch and Hurst, 2019). Therefore, once again we cannot rule out the possibility that plants might have metastasis ever reported.

1.2. CONTAGIOUS METASTASES

Clonally transmissible cancers, also called contagious cancers, are somatic cell lineages that are transmitted between individuals via the transfer of living cancer cells, meaning that they can survive beyond the host that spawned them (Murchison *et al.*, 2014).

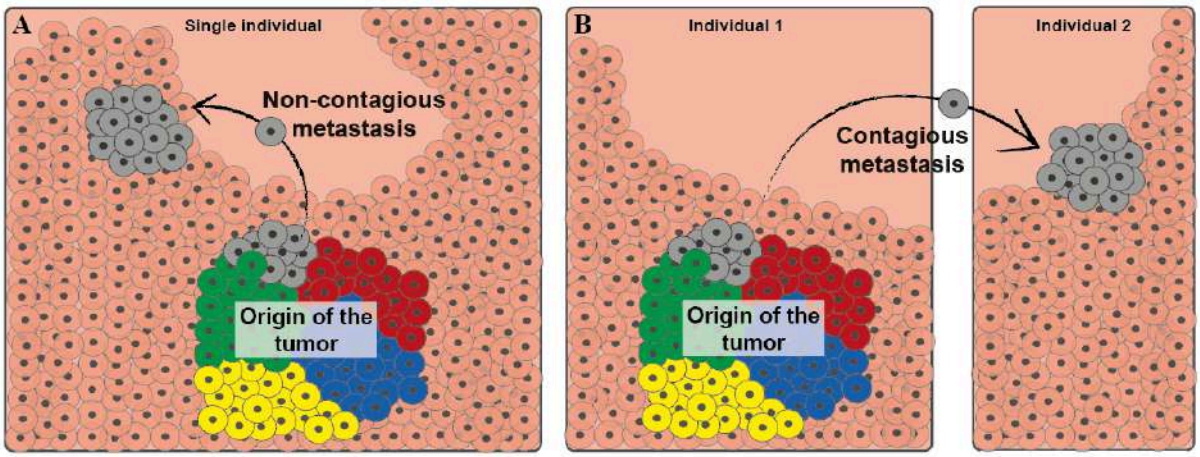


Figure 7. Cancer types regarding the scope of their metastasis. **(A)** Canonical non-contagious cancer that originates in a tissue of the body and metastasize in other organs or tissues of the same individual. **(B)** Contagious cancers originate in an individual, eventually one or several cells are transmitted from one individual to another that will develop the cancer of that cell lineage that was contagious.

While in a 'canonical' metastasis (Figure 7A), tumour cells escape from the tumour of origin and disseminate to a secondary location forming a new tumour (Hunter *et al.*, 2018), in a contagious cancer are tumour cells metastasize between different hosts (Figure 7B). For this reason, they represent an interesting and unique model to illuminate novel insights into the general mechanisms of cancer development and metastasis.

No virus, bacteria or parasite infects the new host, it is the cancer cell itself that will be established in the new individual and then starts to divide to form a new tumour. In other words, these cancer cells acquire the ability of contagion or transmissibility (Pearse and Swift, 2006; Metzger and Goff, 2016).

Therefore, these diseases are distinct from those in which a tumour is initiated by an infectious agent, such as human papillomavirus or human hepatitis B virus. In such cases, although the causative agent is transmissible, the tumour itself remains confined within an individual host (Báez, 2019; Álvarez *et al.*, 2021).

This transmission ability is equivalent to the creation of a new infectious "parasite": the "parasitic" cancer cell will infect an individual different from the one that originated it, it will divide, and its "daughter" cells will continue to infect other individuals (Figure 8).

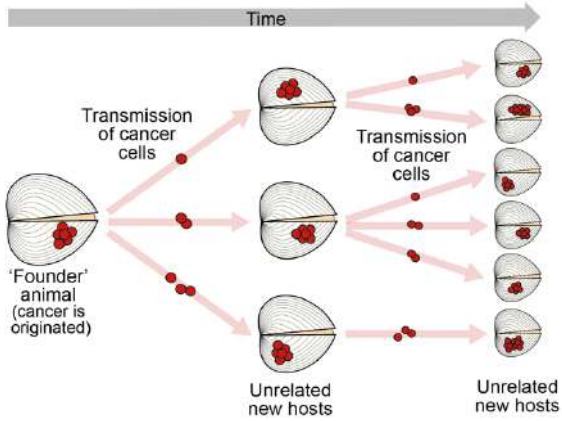


Figure 8. Horizontal spread of a clonally transmissible cancer affecting cockles (*Cerastoderma edule*), that is the transfer of cancer cells between unrelated hosts (adapted from Strakova *et al.*, 2015; reprinted with permission from Elsevier Ltd., see Appendix H).

As for cancer and metastasis, four hallmarks of a contagious cancers have been described as follows: (1) shedding of tumour cells from the affected host, (2) survival of tumour cells

during the host–host transit, (3) a permissive environment facilitating invasion and (4) adaptation to novel habitats and evasion of immune attacks in the foreign host. While this rare confluence of traits explains the rarity of tumour cell transmission, it also suggests that when it happens, multiple emergences can theoretically happen if the favourable window persists (Ujvari, Gatenby and Thomas, 2016).

1.2.1. TRANSMISSIBLE CANCERS THROUGH THE LENS OF SEQUENCING

The first inkling of a transmissible cancer came from a study of canine transmissible venereal tumour, CTVT, dating back to 1876 (Novinski, 1876). The theory of transmission as an allograph –transplant of cells from one individual to another that is not an identical twin– came from artificial transmission experiments and the discovery of genetic markers in the twentieth century. However, it was not till this current century that next-generation sequencing studies have proven that, in these cases, the cancer genomes of different individuals were very similar and different from those of its hosts. Nowadays, contagious cancers are usually studied from a genetic point of view to clarify their transmissible nature. So far, ten contagious cancers have been described, but thanks to sequencing advances, it is possible that many more cases will be identified in the next decade in other species (Murgia *et al.*, 2006; Pearse and Swift, 2006; Metzger *et al.*, 2015, 2016; Pye *et al.*, 2016; Yonemitsu *et al.*, 2019; Garcia-Souto *et al.*, 2022; Michnowska *et al.*, 2022).

To unravel the contagious nature of a cancer, whole genomes of tumour and non-tumour cells coexisting in the same individual are compared with those of other infected individuals (Murgia *et al*., 2006; Pearse and Swift, 2006; Rebbeck, Leroi and Burt, 2011; Pye *et al*., 2016; Strakova, 2017; Báez, 2019). In the case of a non-contagious cancer, phylogenetic analysis reveal that tumour cells are more similar to the non-tumour cells of the same individual than to the tumour cells of other individuals because cancer was independently originated in each individual (Figure 9A). On the other hand, when we face a transmissible cancer, tumour cells of an individual are more similar to the tumour cells of other individuals than to the non-tumour cells of themselves so they will cluster together in a phylogeny (Figure 9B).

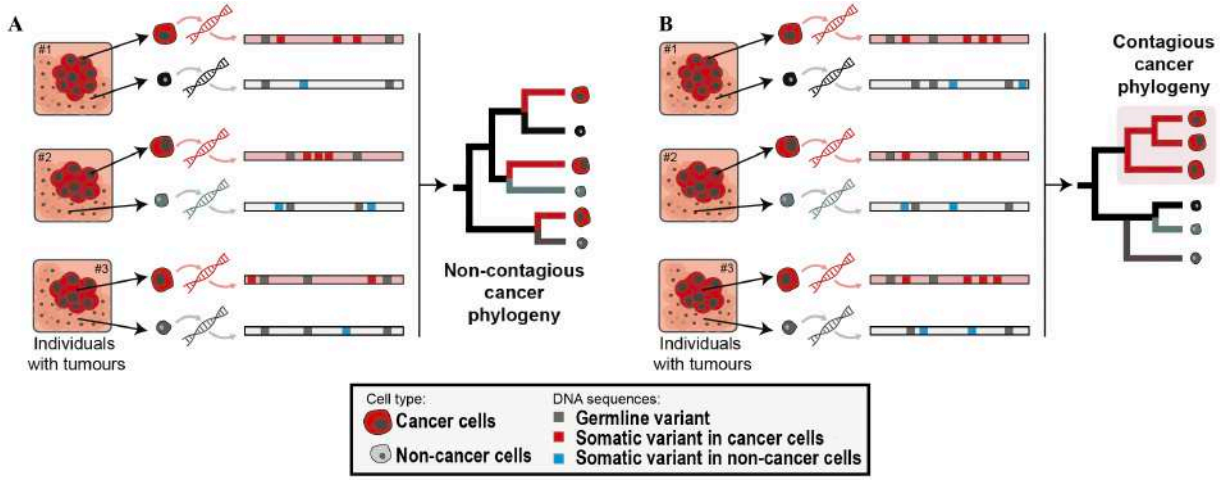


Figure 9. Sequencing data analysis of contagious cancers. **(A)** Phylogenetic model of cancer and non-cancer cells extracted from three individuals affected by a canonical non-contagious cancer. **(B)** Phylogenetic model of cancer and non-cancer cells extracted from three individuals affected by a contagious cancer.

1.2.2. CLONALLY TRANSMISSIBLE CANCERS IN NATURE

Transmissible cancers are a rare phenomenon in nature. Most cancers remain within the body that originated them (Strakova, 2017), however there are three known types of naturally occurring clonally transmissible cancers: one of which is a leukaemia-like cancer found in nine marine bivalves, called disseminated neoplasia (Metzger and Goff, 2016; Yonemitsu *et al.*, 2019; Garcia-Souto *et al.*, 2022; Michnowska *et al.*, 2022). The other two affect mammals –a venereal tumour in dogs (Murgia *et al.*, 2006) and a facial sarcoma in Tasmanian devils (Pearse and Swift, 2006; Pye *et al.*, 2016)– requiring physical contact between animals by coitus or biting for the contagion.

Furthermore, a transmissible cancer was suggested to spread by cannibalism and mosquitoes bite in a laboratory population of golden Syrian hamsters (*Mesocricetus auratus*) but it is no longer maintained (Brindley and Banfield, 1961; Cooper, Mackay and Banfield, 1964; Banfield *et al.*, 1965; Ostrander, Davis and Ostrander, 2016).

1.2.2.1. Canine Transmissible Venereal Tumour (CTVT)

The first description of this disease was written more than 200 years ago (Blaine, 1810). In nature, the transfer of tumour cells from one dog to another generally occurs during mating resulting in the development of tumours on the genitals of females and males (Figure 10A; Das and Das, 2000).

The notion that the tumour is naturally transmissible as an allograft came from three lines of observation (Murgia *et al.*, 2006): (1) CTVT can only be experimentally induced by transplanting living tumour cells, and not by killed cells or cell filtrates (Cohen, 1985), (2) CTVT karyotype is aneuploid but has characteristic marker chromosomes in tumours collected in different geographic regions (Weber, Nowell and Hare, 1965; Oshimura, Sasaki and Makino, 1973) and (3) a long interspersed nuclear element (LINE-1) insertion near *c-myc* has been found in all tumours examined (Katzir *et al.*, 1987).

It was not till 2006 that the transmissible nature of the tumour was genetically confirmed (Murgia *et al.*, 2006) analysing genetic markers, microsatellites, and mitochondrial DNA in naturally occurring tumours and matched blood samples. In each case, the tumour was genetically distinct from its host and all tumours were derived from a single neoplastic clone.

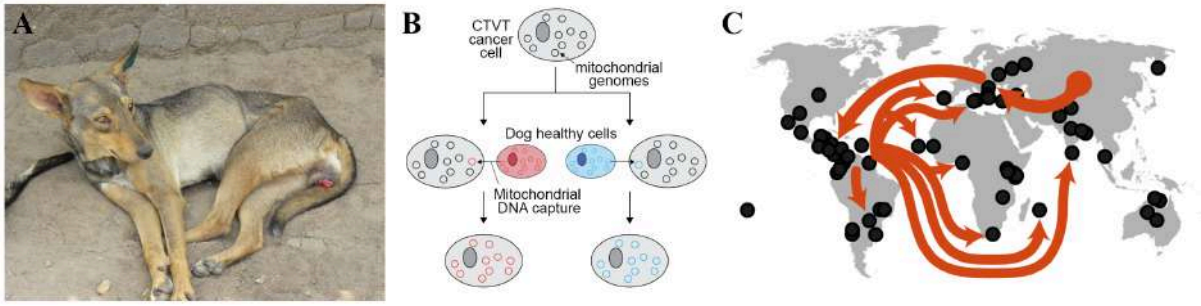


Figure 10. Canine Transmissible Venereal Tumour. **(A)** Dog affected by CTVT (source Strakova and Murchison, 2014; reprinted with permission from BioMed Central Ltd., see Appendix H) **(B)** diagram of the horizontal transfer of mitochondrial genomes (adaptation Strakova and Murchison, 2015; copyright 2015, Elsevier Ltd.) and **(C)** the cancer's phylogeographic history (source Báez, 2021; reprinted with permission from AAAS, see Appendix H).

Five years later, mitochondrial sequences from dogs, wolves and CTVT tumours were used to build a phylogeny and the authors discovered that CTVT mitochondrial genomes do not share a clonal origin. In other words, mitochondrial genomes did not support the phylogeny build

with nuclear sequences. This led to the proposal that CTVT cells periodically capture mitochondria from their hosts (Figure 10B; Rebbeck, Leroi and Burt, 2011).

The histogenesis of CTVT remains unclear although immunophenotypic suggested a histiocytic origin of CTVT (Murgia, 2006; Hendrick, 2017). Moreover, tumour cells in culture undergo a morphological transformation from round cells to fibroblast-like cells (Murgia, 2006).

In humans we can trace back breast cancer to the Persian queen Atossa in 500 BC (Mukherjee, 2010) but today we know of an even older cancer that is still alive: CTVT. Recent phylogenetic studies (Báez, 2019) of the mutations in these tumours suggested that CTVT originated 4000–8500 years ago in Central or Northern Asia, and probably travelled to Europe along the Silk Road. Sixteenth-century Europeans subsequently introduced CTVT into the Americas, from where it spread to dogs worldwide in an unfettered sweep enabled by the transoceanic trade routes of the eighteenth and nineteenth centuries (Figure 10C; Báez, 2021). Indeed, this is the oldest known cancer to date.

Potential driver mutations have been identified in *SETD2*, *ERG*, *CDKN2A*/B, *PTEN* and *RB1*, the only putative driver with experimental support is a mutation of a retrotransposon near c-*MYC* (Murchison, 2014; Decker *et al.*, 2015; Strakova and Murchison, 2015; Báez, 2019). Remarkably, mutations on these genes are also known to be driver in human non-contagious cancers.

The mutations also provided insights of the mutagenic processes that have acted on cancer cells. In fact, ultraviolet-light-mediated DNA damage and the tumour collection latitude have been associated, providing evidence that sunlight-induced DNA damage is exacerbated in low-latitude regions (Báez, 2019, 2021; Báez *et al.*, 2019).

Regarding CTVT natural selection, no evidence of ongoing selection has been found what contrasts to observations of positive selection in human cancers. This suggests that tumour evolution proceeds differently over short and long timescales, being long-term cancer evolution dominated by neutral processes such as genetic drift, rather than natural selection (Báez, 2019, 2021; Báez *et al.*, 2019).

1.2.2.2. Devil Facial Tumour Disease (DFTD)

Tasmanian devils (*Sarcophilus harrisii*) are marsupial carnivores endemic to the Australian island of Tasmania. This species was considered endangered by the International Union for Conservation of Nature (Hawkins *et al.*, 2008) due to the emergence of a clonally transmissible cancer known as devil facial tumour 1 (DFT1; Pearse and Swift, 2006) because since 1996, when it was observed for the first time, DFT1 had spread widely throughout Tasmania, causing significant declines in devil populations (Hawkins *et al.*, 2006; Lazenby *et al.*, 2018). Regardless the extinction predictions of epidemiological models, natural population recovery has been observed in the latest years because Tasmanian devils are responding to the strong selection pressure imposed by DFTD, particularly given the low genetic diversity in Tasmanian devil populations (Jones *et al.*, 2019).

In 2014, routine diagnostic screening revealed a second transmissible cancer in Tasmanian devils that was called DFT2 (Pye *et al.*, 2016). Remarkably, DFT1 and DFT2 present a similar appearance of solid tumours located in the face, neck or mouth (Figure 11A), same transmission route –biting which is a common behaviour of the species– and life cycle, but are histologically and genetically distinct (Pearse and Swift, 2006; Pye *et al.*, 2016; Stammnitz *et al.*, 2018). However, DFT2 is geographically restricted to a peninsula in south central Tasmania.

Both DFT1 and DFT2 have been confirmed as clonally transmissible by means of karyotypic (Figure 11B) and genetic analyses (Murchison *et al.*, 2010, 2012; Pye *et al.*, 2016; Stammnitz *et al.*, 2018), and both have been proposed to be neural-crest-derived tumours (Murchison *et al.*, 2010; Stammnitz *et al.*, 2018). DFT1 is characterised by having an X chromosome while DFT2 has been found to have pieces of chromosome Y suggesting the independent origin of these two cancer lineages in devils of different sexes.

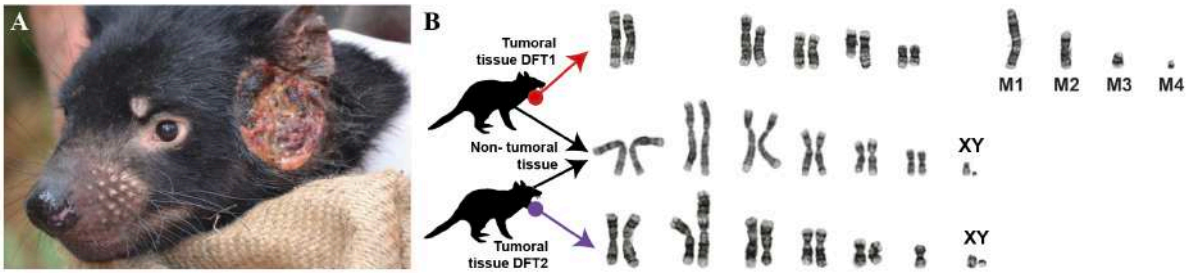


Figure 11. Devil Facial Tumour Disease. **(A)** Gross appearance of DFTD (source Stammnitz *et al.*, 2018; reprinted with permission from Elsevier Ltd., CC BY 4.0, see Appendix H). **(B)** Karyotypes of DFT1, DFT2 and non-tumoral tissue of Tasmanian devils showing that DFT1 and DFT2 bear no detectable similarities among them and against normal cells (adaptation from Pye *et al.*, 2016; reprinted with permission from PNAS, see Appendix H).

Surprisingly, female devils are more tolerant to infection, with males suffering bigger declines of their body condition and having smaller tumours (Ruiz-Aravena *et al.*, 2018). This increased female survival was associated with genes such as *ST8SIA2* involved in chronic inflammation, *SCAMP1* that has immune functions and *POLR3C* implicated on non-self-recognition (Margres *et al.*, 2018).

1.2.3. CANCER TRANSMISSION IN HUMANS

Despite the recent discovery of contagious cancers, they have already been found in several species and some of them are known to have arisen independently more than once in a particular species what highlights the possibility that they arise in nature with greater frequency than expected and scares the idea that they could affect us.

Numerous cancer patient sequencing studies have been conducted in recent decades, and to my knowledge, there is no evidence that a contagious cancer similar to that of dogs, Tasmanian devils or bivalves is spreading in humans.

While a viral infection can lead to the development of cancer (Álvarez *et al.*, 2021), it is the virus what is transmitted from one individual to another, therefore, they do not account as cancer transmission. However, there are some cases where cancer cells have been transmitted from one human to another resulting on cancer development in the second one.

The only natural route available for transfer of cancer cells between individuals is via the placenta (Greaves and Hughes, 2018) and, as a matter of a fact, maternal-foetal transmission (Figure 12A-B) and foetal-foetal transfer of cancer cells (Figure 12C) during pregnancy have been largely described (Greaves *et al.*, 2003; Tolar and Neglia, 2003; Seckl, Sebire and Berkowitz, 2010; Nancy *et al.*, 2012; Greaves and Hughes, 2018; Arakawa *et al.*, 2021). Leukaemia, choriocarcinomas and uterine cervical tumours are the common cancers transmitted between humans via this natural route and their genetic analysis suggests that cancer cells acquire mutations that allow them to evade the child's defences (Arakawa et al., 2021; Seckl, Sebire and Berkowitz, 2010; Greaves and Hughes, 2018; Greaves et al., 2003). Genetic studies of monochorionic twins with a shared clonal origin of childhood leukemia have provided

unique insights into the disease propagating cells in these diseases (Hong *et al.*, 2008). A recent case of a cancer –myelofibrosis– originated in the utero origin was found in two adult monozygotic twins supporting that the latency between acquisition of an initiating driver mutation and presentation with overt cancer can be prolonged (Sousos *et al.*, 2022), therefore, these cancers are probably more common.

In adults, cancer only seems to spread when a person's defences are not working properly. Interestingly, cancer contagion can occur between humans by organ transplant, patients who receive organ transplants may be vulnerable to getting cancer from their donor (Tolar and Neglia, 2003; Gandhi and Strong, 2007; Matser *et al.*, 2018). In 2018, four patients developed breast cancer after receiving kidneys, lungs and liver from a 53-year-old donor who had died in an accident (Matser *et al.*, 2018, Figure 12D). The cancer cells did not match those of the patients, but rather those of the donor who did not have the disease at the time of transplantation.

Melanoma transplantation into the rectus of a woman was performed in an attempt to produce tumour antibodies that might be helpful for the treatment of her daughter that had melanoma; unfortunately, the recipient died with disseminated melanoma months later presumed to be originated from the injected cells (Scanlon *et al.*, 1965).

Three examples of human cancer contagion with no recurrence after removal have been reported in medical or scientific workers who accidentally cut themselves with scalpels or needles that carried cancer cells developed a tumour in the wound area (Figure 12E): (1) cancer transfer from patient to surgeon during an operation (Gartner *et al.*, 1996), (2) accidental inoculation in the hand during biopsy (Greaves and Hughes, 2018) and transfer to a laboratory worker from a cell line (Gugel and Sanders, 1986).

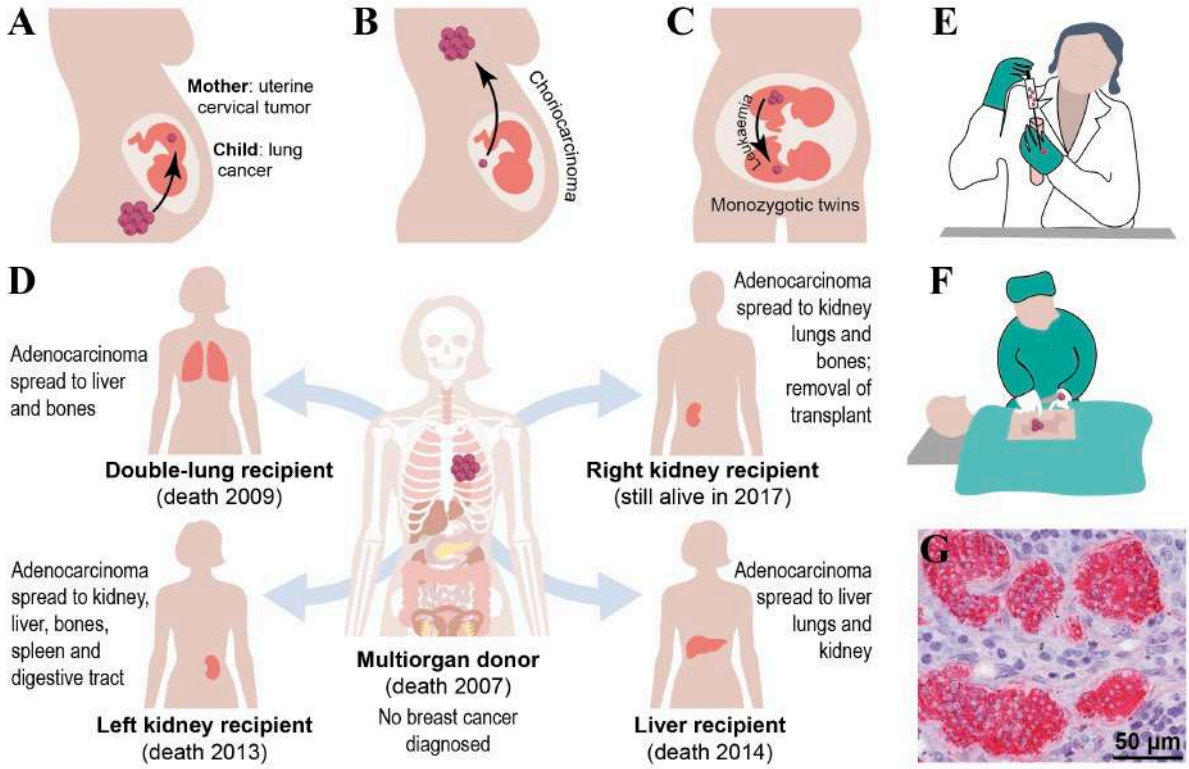


Figure 12. Human cancer contagions. **(A)** Two cases reported vaginal transmission of cancer from mothers with cervical cancer to their infants that develop lung cancer (Arakawa *et al.*, 2021). **(B)** Intraplacental choriocarcinoma can result in disseminated disease in the mother, infant or both (Seckl, Sebire and Berkowitz, 2010; Greaves and Hughes, 2018). **(C)** Twin to twin dissemination of leukaemia in utero (Greaves *et al.*, 2003). **(D)** Transmission of breast cancer by a single multiorgan donor to 4 transplant recipients (Matser *et al.*, 2018).

(E) Accidental transfer from an adenocarcinoma cell line to a healthy laboratory worker during an experiment (Gugel and Sanders, 1986). **(F)** Accidental transfer from patient to surgeon during an operation (Gartner *et al.*, 1996). **(G)** Parasite-derived cancer cells in a human host shown by in situ hybridization with the use of a cestode DNA probe (reproduced with permission from Muehlenbachs *et al.*, 2015; copyright Massachusetts Medical Society, see Appendix H).

Lastly, all previous examples were cancer contagion from human to human, nevertheless, a case of parasite-derived cancer cells metastasizing in a human host has been reported: a tapeworm infection of a cestode ended up in nests of undifferentiated cells with invasive behaviour in the lymph-node and lung of its human host that was immunosuppressed because of other pathological conditions (Muehlenbachs *et al.*, 2015; Figure 12F).

1.3. BIVALVE TRANSMISSIBLE NEOPLASIAS

Leukaemia-like diseases known as disseminated or haemic neoplasia (HN) were reported in many bivalve species in the twentieth century albeit the clonal transmission of a HN case was not established till 2015 (Metzger *et al.*, 2015).

HN (Elston *et al.*, 1988) has also been called sarcomatoid proliferative disease (Farley, 1969b), proliferative atypical haemocytic condition (Lowe and Moore, 1978), epizootic sarcoma (Farley, Otto and Reinisch, 1986), sarcomatous neoplasia (Brousseau, 1987), transmissible sarcoma (Farley, Plutschak and Scott, 1991), systemic neoplasia (Moore *et al.*, 1991) and disseminated neoplasia (Galimany and Sunila, 2008; Carballal *et al.*, 2015). In this doctoral thesis, the nomenclature HN will be used because the cell-of-origin of the disease has already been studied in cockles (see *Chapter 3*).

1.3.1. HISTORY, PREVALENCE AND DISTRIBUTION

Benign and malignant neoplasia cases, including HN and gonadal neoplasia, have been described in many marine bivalve species from four continents and all oceans (Carballal *et al.*, 2015; Skazina *et al.*, 2022), but only HN has been recently reported to be transmissible in nine species (Table 1).

In the late 60's, HN was firstly described in the oysters *Crassostrea virginica* and *Crassostrea gigas* (Farley, 1969a) but it was not till the 80's that HN would be reported in cockles *Cerastoderma edule.* Cockle's HN (Table 2) was discovered in Brittany, France (Poder and Auffret, 1986) and Cork Harbour (Ireland) (Twomey and Mulcahy, 1984) and fifteen years later in several locations of the northwest coast of Spain (Carballal *et al.*, 2001; Villalba, Carballal and López, 2001; Ordás and Figueras, 2005; Díaz *et al.*, 2016). In 2021, a cockle's parasites study reported cockle's HN in two additional countries: Portugal and The Netherlands (Montaudouin *et al.*, 2021).

HN has been found in at least 19 bivalve species (House and Elston, 2006) that includes oysters (*Magallana gigas*[3]*, Crassostrea rhizophorae, Crassostrea virginica, Ostrea chilensis, Ostrea lurida, Ostrea edulis, Saccostrea commercialis*), mussels (*Mytilus edulis, Mytilus galloprovincialis*), cockles (*Cerastoderma edule, Cerastoderma glaucum*) and clams (*Arctica islandica*, *Macoma balthica, Macoma calcarean, Macoma nasuta, Macoma irus, Mya arenaria, Mya truncata, Tagelus plebeius*).

Prevalence of the disease varies depending on the species studied and the spatio-temporal variation of the disease in the screened populations. Cockles *C. edule* from the western Atlantic coast of Europe, softshell clams *Mya arenaria* from the north-eastern Atlantic coast of America, mussels *Mytilus trossulus* from the north-western Pacific coast of America, and clams *Macoma balthica*[4] from the Chesapeake Bay and the Baltic Sea reported high prevalence of HN associated with significant mortalities (Christensen, Farley and Kern, 1974; Elston and Moore, 1992; Pekkarinen and Lei, 1994; Thiriot-Quiévreux and Wolowicz, 1996; Díaz *et al.*, 2016; Dairain *et al.*, 2020; Geoghegan *et al.*, 2021; Hammel *et al.*, 2021). Among these species, the majority have been analysed for cancer transmission and it was found that multiple contagious cancer lineages are spreading through them (Table 1).

[3] Citation referred as *Crassostrea gigas,* current taxonomic name *Magallana gigas* (Salvi and Mariottini, 2017).
[4] Sometimes referred as *Limecola balthica,* current taxonomic name *Macoma balthica* (WoRMS database).

Table 1. Marine bivalve species that have been corroborated to have clonally transmissible neoplasia (missing data is indicated *NA*).

Cancer diagnosed species (Common name)	Species of cancer origin (Common name)	Locations of animal collection	Year of collection	Cancer* / collected samples	Cancer samples analysed	Contagious cancer lineages	Genetic analysis performed suggesting cancer transmission	Reference/s
Clams								
Mya arenaria (Soft shell clam)	Same	P.E.I., Canada Maine, USA New York, USA *Total: 2 countries*	2009-2010 2013 2014	4 / *NA* 3/92 + 1/*NA* 1 / 12	4 4 1 *Total: 9*	**MarBTN**	Nuclear: 10 microsatellite loci, *Steamer* integration sites Mitochondrial: *mtCOI* locus, *CYTB* locus	Metzger *et al.*, 2015
Polititapes aereus (Carpet shell clam)	*Venerupis corrugata* (Pullet carpet shell)	O Bohído, Spain	2014	31 / 74	6	**VcoBTN**	Nuclear: *EF1a* quantitative PCR Mitochondrial: mtCOI locus, rDNA *ITS*	Metzger *et al.*, 2016
Venus verrucosa (Warty venus clam)	*Chamelea gallina* (Striped venus clam)	Ribeira, Spain Mahón, Spain	2017 2018	3 / 30 5 / 67	8	**CgaBTN**	Nuclear: DEAH12 locus Mitochondrial: coverage analysis of mtDNA with WGS data, *mtCOI* locus	Garcia-Souto *et al.*, 2022 ****
*Macoma balthica*** (Baltic clam)	Same	Gulf of Gdansk, Poland	2019	4/100	4	**MbaBTN**	Nuclear: partial *EF1a* locus Mitochondrial: *mtCOI* locus	Michnowska *et al.*, 2022
Mussels								
Mytilus trossulus (Foolish mussel)	Same	Copper Beach, Canada Esquimalt, Canada	2015 2015	2 / 28 9 / 250	2 7	**MtrBTN1**	Nuclear: partial *EF1a* locus Mitochondrial: *mtCOI* locus	Metzger *et al.*, 2016
		Gaydamak Bay, Russia	2019	4 / 226	4	**MtrBTN2**	Nuclear: EF1*a* locus, microsatellite *Mgm3* loci Mitochondrial: *mtCR* locus (including *16S* rRNA), *mtCOI* locus	Skazina *et al.*, 2021
		Nagaev Bay, Russia *Total: 2 countries*	2020	11 / 214	3 *Total: 13*	**MtrBTN1** **MtrBTN2**	Nuclear: EF1*a* locus Mitochondrial: *mtCR* locus (including *16S* rRNA), *mtCOI* locus	Skazina *et al.*, 2022
Mytilus edulis (Blue mussel)	*Mytilus trossulus* (Foolish mussel)	Arcachon, France Normandy, France Wadden Sea, Netherlands Chausey Island, France Brittany, France	2016 2015 2009 2009 2017	2 / *NA* 1 / *NA* 5 / 938 2 / >4000 2 / ~100	2 1 4 2 2	**MtrBTN2**	Nuclear: partial *EF1a* locus, *H4* locus Mitochondrial: *mtCR* locus (including lrRNA), *mtCOI* locus	Yonemitsu *et al.*, 2019
		10 Atlantic and English Channel locations, France Wadden Sea, Netherlands *Total: 2 countries*	2009-2017 2009	22 / 430 1 / 23	22 1 *Total: 35*	**MtrBTN2**	Nuclear: 74 SNPs, partial *EF1a* locus Mitochondrial: *mtCOI* locus	Hammel *et al.*, 2022
Mytilus chilensis (Chilean mussel)	*Mytilus trossulus* (Foolish mussel)	Beagle Channel, Argentina Castro, Chile *Total: 2 countries*	2012 2018	6 / 60 6 / 200	6 3 *Total: 9*	**MtrBTN2**	Nuclear: partial *EF1a* locus, *H4* locus Mitochondrial: *mtCR* locus (including lrRNA), *mtCOI* locus	Yonemitsu *et al.*, 2019
Mytilus galloprovincialis (Mediterranean m.)	*Mytilus trossulus* (Foolish mussel)	1 location, Croatia Brittany, France *Total: 2 countries*	2009-2017 2009-2017	1 / 12 1 / *NA*	1 1 *Total: 2*	**MtrBTN2**	Nuclear: 74 SNPs, partial *EF1a* locus Mitochondrial: *mtCOI* locus	Hammel *et al.*, 2021
Cockles								
Cerastoderma edule (Common cockle)	Same	O Sarrido, Spain	2014	9 / 150	6	**Ced-a-BTN1** **Ced-b-BTN2** ***	Nuclear: 9 microsatellite loci, partial *EF1a* locus Mitochondrial: *mtCOI* locus	Metzger *et al.*, 2016

* Cancer diagnosis varies among the studies from histology, cytology, flow cytometry, genetic tests, or a combination of methods. ** Sometimes referred as *Limecola balthica*, current taxonomic name *Macoma balthica* (WoRMS database). *** Cancer lineages diagnosed in independent samples and corresponding to different histological features (Metzger *et al.*, 2016). **** This research is part of this doctoral thesis.

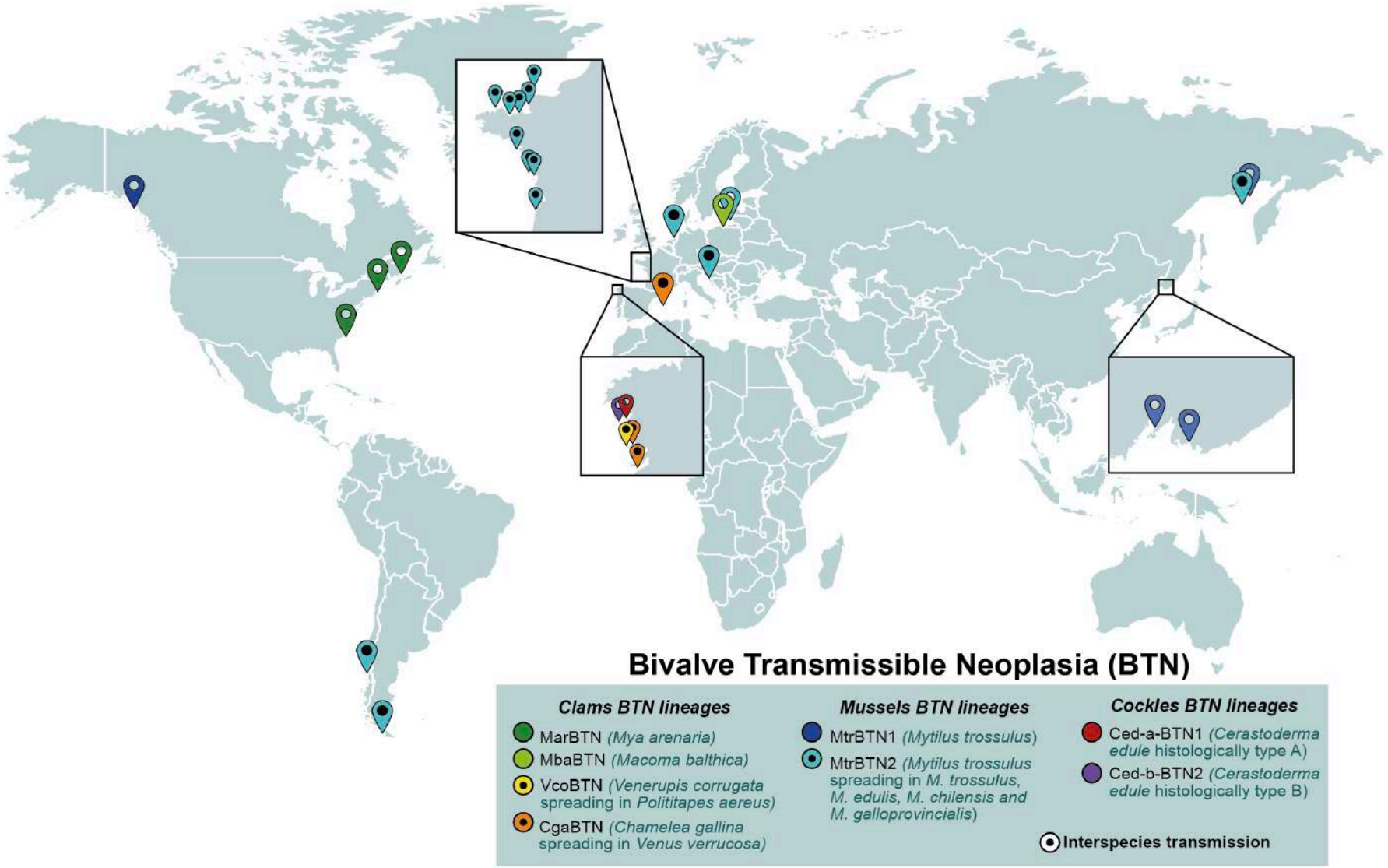


Figure 13. Worldwide distribution of bivalve contagious cancer lineages. Locations of reported cases of HNs that have been proven to be transmissible, colours represent the cancer lineage identified. The cancer lineage (green) affecting clams *M. arenaria* in the east coast of USA and Canada was the first cancer lineage ever identified to be transmissible in bivalves. A cross-species cancer lineage transmission (middle black point) was identified to have originated in the calms *V. corrugata* and be currently spreading among *P. aereus* individuals, later two additional cases were described. The mussels BTN1 lineage (dark blue) was only identified in *M. trossulus* populations in Canada, while the BTN2 lineage (light blue) was probably originated *M. trossulus* and currently identified to be spreading among *M. trossulus, M. edulis, M. galloprovincialis* and *M. chilensis* populations in both the Pacific and Atlantic Oceans. Two independent cancer lineages have been described in cockles *C. edule* corresponding to different morphologies of cells: type A (red) and type B (purple).

Contagious HN metastases have been reported worldwide (Figure 13) although some cancer lineages are restricted to a local area while others have spread thousands of miles (Metzger *et al.*, 2016; Yonemitsu *et al.*, 2019; Garcia-Souto *et al.*, 2021). In both mussels *Mytilus trossulus* and cockles *Cerastoderma edule*, more than one independent cancer lineage has been identified to have arisen and spreading in the host species (Metzger *et al.*, 2016; Yonemitsu *et al.*, 2019).

Table 2. Location and prevalence reports of HN affecting cockles *Cerastoderma edule*.

Country	Location	Prevalence	References
Cockles *C. edule*			
Ireland	Cork	22%-94%	Twomey and Mulcahy, 1984; Collins and Mulcahy, 2003; Barber, 2004
	Dundalk	<1%	Montaudouin *et al.*, 2021
France	Britanny	2.2-46%	Poder and Auffret, 1986; Le Grand *et al.*, 2010
	Arcachon	28%	Montaudouin *et al.*, 2021
	Somme	2%	Montaudouin *et al.*, 2021
Spain	Galicia	0-84%	Carballal *et al.*, 2001, 2015; Villalba, Carballal and López, 2001; Da Silva *et al.*, 2005; Diaz, 2005; Romalde *et al.*, 2007; Diaz *et al.*, 2013; Ruiz *et al.*, 2013; Montaudouin *et al.*, 2021
Portugal	Aveiro	2%	Montaudouin *et al.*, 2021
	Formosa	7%	Montaudouin *et al.*, 2021
Netherlands	Texel, Wadden Sea	2%	Montaudouin *et al.*, 2021
Cockles *C. glaucum*			
Spain	Galicia	1 case	Rodriguez *et al.*, 1997
		~2%	Carballal *et al.*, 2016
Poland	Gdansk	unk	Ogrodowczyk, 2017

Since most of the bivalve HN analysed for transmission are attributable to contagious cancers, it is reasonable to predict that HN in other bivalves will be found to be contagious as well (Metzger and Goff, 2016).

1.3.2. AETIOLOGY AND HISTOGENESIS

The aetiology of HN, that is the cause of the condition, has been debated since the discovery of the disease.

Sublethal levels of biotoxins, the presence of stressors and various pollutants *(i.e.,* fuel or chlordane) have been proposed to induce the development of HN (Yevich and Barszcz, 1976; Balouet *et al.*, 1986; Twomey and Mulcahy, 1988; Farley, Plutschak and Scott, 1991; Landsberg, 1996), although none of these hypotheses were supported by experimental data (Romalde *et al.*, 2007).

Transmission of HN was suggested (Oprandy *et al.*, 1981; Oprandy and Chang, 1983; Appeldoorn, Brown and Chang, 1984) before its recent genetic study demonstration (Metzger *et al.*, 2015); in fact, successful transplantation of cockles' HN (Twomey and Mulcahy, 1988; Díaz *et al.*, 2017) was already achieved in the eighties although whether the transplanted cells proliferated in the new host or released an infectious agent remained an open question (Collins and Mulcahy, 2003) indicating that an infective agent could be involved. In addition, in cockles high reverse retrotranscriptase activity was observed supporting the idea of a viral transmission on these cancers (Romalde *et al.*, 2007). Thus, ultrastructural examination of neoplastic cells from cockles did not reveal a clear pathogenic agent (Poder and Auffret, 1986; Elston and Moore, 1992), once virus-like particles were observed in a neoplastic cockle but it was not confirmed in other samples (Romalde *et al.*, 2007).

Recent investigations suggested that HN spreads through shedding of cancer cells from infected animals into seawater, from which they are subsequently filtered by susceptible animals (Metzger and Goff, 2016; Metzger *et al.*, 2016). HN tumour cells behave like metastatic cells, leaving their hosts to dive in the marine environment until they reach a new host and propagate inside it (Figure 14). Supporting this hypothesis, the ability of cancer cells to survive in artificial seawater has been tested and detection of cancer cells from natural seawater has been found (Giersch *et al*., 2022).

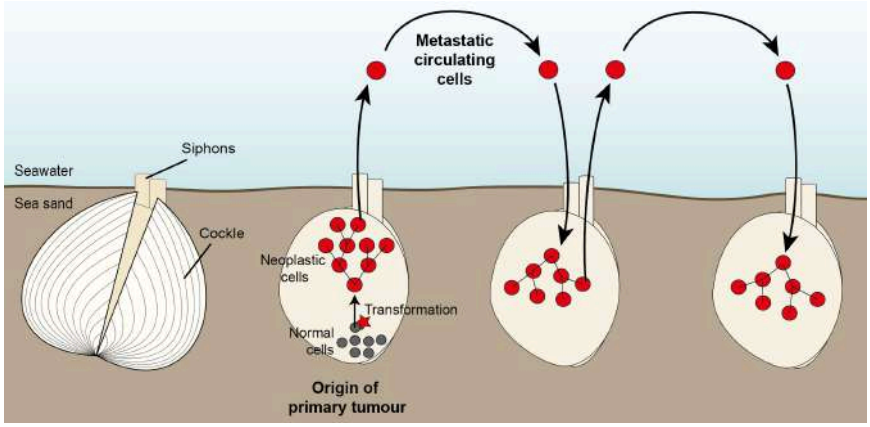


Figure 14. Bivalve transmissible cancers are thought to spread through shedding of cancer cells from infected animals into seawater, from which they are subsequently filtered by susceptible animals (this scheme has been adapted from the Scuba Cancers ERC proposal, courtesy of Jose Tubío).

Despite the discoveries of HN aetiology in the understanding of cancer causation in multiple bivalve species, the cell-of-origin of these cancer cells remains unknown in all of them. It is generally considered to be a sarcoma (neoplasia of mesoderm-derived tissues) although a haematopoietic and a gonadal origin have also been proposed (Alderman, Green and Balouet, 2017). In 1969, Farley et al. described the first HN as a probable neoplastic disease of the hematopoietic system.

Because neoplastic cells are first observed in the haemolymph, with increasing prevalence over normal haemocytes as the disease progresses, and because normal and neoplastic haemocytes share receptors for the same monoclonal antibodies (Reinisch, Charles and Troutner, 1983; Smolowitz, Miosky and Reinisch, 1989; Muttray and Vassilenko, 2018), it is believed that normal and neoplastic haemocytes are ontogenetically related and that neoplastic cells are of haemocytic origin. However, similar neoplastic diseases in other bivalves, *M. balthica* (Christensen, Farley and Kern, 1974) and *P. aureus* (Carballal et al. 2013), appeared to have the gill epithelium as the origin of neoplastic cells, which subsequently spread to other organs.

We cannot rule out the possibility of a non-haemocytic cell line being the ancestry of HN cancer cells. Interestingly, histopathology and gene-expression profiles of tumours often remain relatively stable during progression from primary tumour to metastasis and even end-stage disease (Visvader, 2011) providing a good scenario to investigate the origin of cancer cells.

1.3.3. MORPHOLOGICAL CHARACTERISTICS

HN is characterised by the proliferation of large, anaplastic circulating cells in the haemolymph – i.e., the fluid analogous to the vertebrate's blood that circulates in the interior of molluscs (Barber, 2004; Carballal *et al.*, 2015). It is currently unknown whether all the

identified cases of HN are clonally transmissible, but it is very likely that more cases will be described soon (Metzger and Goff, 2016). Non-transmissible cases have also been reported in mussels coexisting in populations with transmissible cancers (Hammel *et al.*, 2021).

In cockles *C. edule*, HN has been reported in individuals ranging from 10 to 40 mm in length, with the highest prevalence and severity in cockles of intermediate size/age and the sex seemed not to influence susceptibility to HN (Díaz *et al*, 2016).

HN cannot be diagnosed by the external examination of individuals (Farley, 1969a), for that reason, cyto-histological and genetic methods have been developed for its diagnosis: *histology* (Farley, 1969a; Yevich and Barszcz, 1976) and *haemocytology* (Peters, 1988) consisting on the observation of neoplastic cells, *immunoassays* (Smolowitz and Reinisch, 1986) to detect antibodies raised against specific antigens of cancer cells, *flow cytometry* (Elston, Kent and Drum, 1988, (Vassilenko and Baldwin, 2014) to detect cancer cells measuring DNA content and *genetic testing* (Metzger *et al.*, 2016) using as molecular markers certain cancer-related genes and, in some cases, the method developed as a quantitative PCR.

1.3.3.1. Histological features

Histology was the first method used to diagnose this disease (Farley, 1969a; Yevich and Barszcz, 1976) because, as cancer cells can are morphologically different, they can be observed in the tissues (Figure 16). Several studies established scales of neoplasia progression based on the number of neoplastic cells observed and the tissues and organs affected (Carballal *et al.*, 2015).

Two morphologically different types of neoplastic cells were distinguished in mussels *Mytilus spp*. (Lowe and Moore, 1978; Mix, Hawkes and Sparks, 1979; Moore *et al.*, 1991), in the softshell clam *M. arenaria* (Brown *et al.*, 1985) and in cockles *C. edule* (Carballal *et al.*, 2001). Most of the affected cockles had neoplastic cells like those previously described (Twomey and Mulcahy, 1984; Poder and Auffret, 1986) but another type of neoplastic cell was seen in some cockles and called ‘neoplasia B’. The latter were smaller and had round to oval nuclei with a single nucleolus; they were more tightly packed in the connective tissue than the former neoplastic cells (Figure 15).

Abundant and swollen mitochondria and altered Golgi complexes are ultrastructural features often observed in these cancer cells (Poder and Auffret, 1986, Díaz *et al.* 2011).

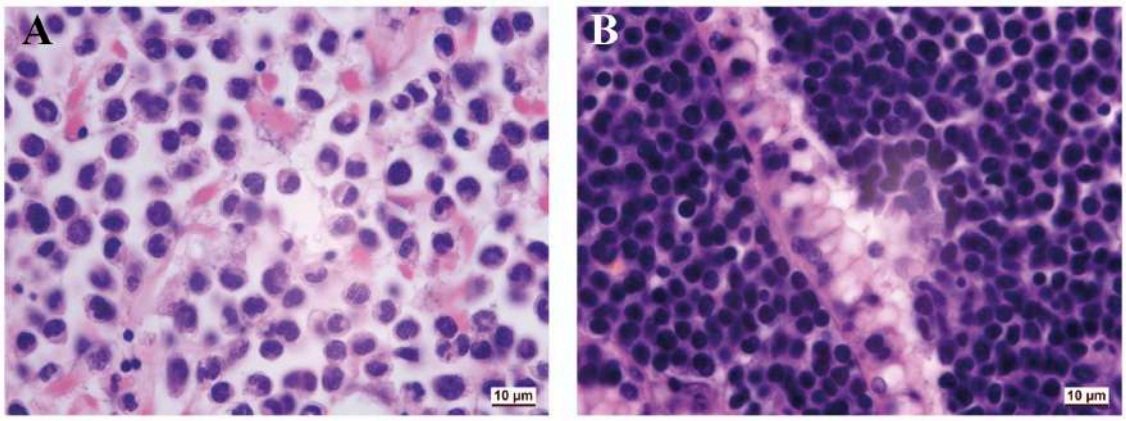


Figure 15. Histology (H&E stain) of two cockles HN affected (Courtesy of Seila Díaz). **(A)** Neoplasia type A and **(B)** type B.

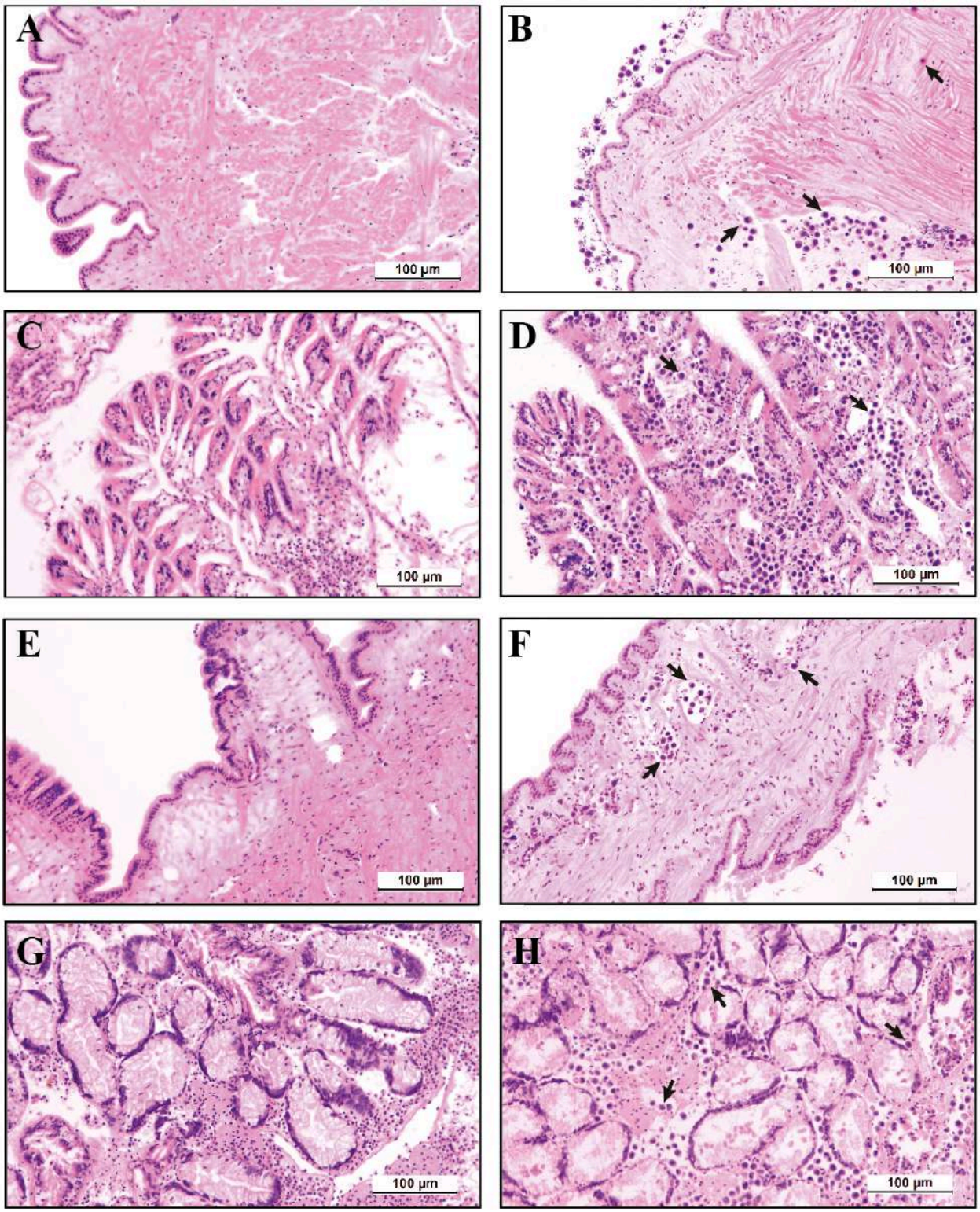


Figure 16. Histological (H&E stain) comparison of HN and non-cancer tissues of cockle *C. edule* highlighting neoplastic cells (arrows). Foot section of **(A)** non-cancer cockle and **(B)** a HN cockle. Gills section of **(C)** non-cancer cockle and **(D)** a HN cockle. Mantle section of **(E)** non-cancer cockle and **(F)** a HN cockle. Digestive gland section of **(G)** non-cancer cockle and **(H)** a HN cockle. These micrography pictures have been taken by the doctoral candidate for this thesis within the framework of Scuba Cancers ERC project.

1.3.3.2. Cytological features

HN involves the occurrence of neoplastic circulating cells and for this reason the disease can also be detected in haemolymph samples.

By extracting of haemolymph from either the pericardial region or the posterior adductor muscle and cyto-centrifugating it onto on slides, the haemolymph can be directly stained and examined with bright-field microscopy. In 1988, Peters et al. referred to this technique as "haemocytology" and this method allows repeated sampling of living individuals as it does not require to kill the animal (Carballal *et al.*, 2015).

Neoplastic cells are larger and rounder, with a nucleus cytoplasm ratio much higher than that of hemocytes, less or no pseudopodia emission and they present frequent mitotic figures (Díaz, 2015).

Cooper, Brown and Chang (1982) found a positive correlation between the number of circulating neoplastic cells and the histopathological lesions, which provided support for using the number of neoplastic circulating cells as an indicator of the degree the disease.

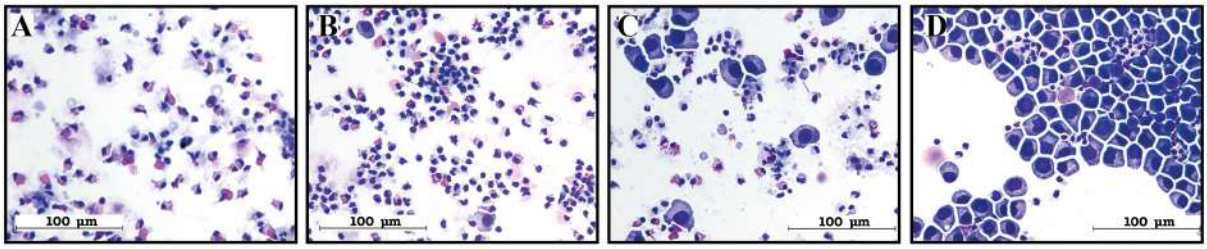


Figure 17. Cytological severity scale for the diagnosis of HN in cockles *Cerastoderma edule* (Diaz et al., 2010). **(A)** Non-affected -N0- when not a single cancer cell was seen under the microscope; **(B)** early-stage cancer -N1- when individuals showed proportion of cancer cells lower than 15% in the haemolymph cell monolayers; **(C)** medium-stage cancer -N2- when the proportion ranged from 15% to 75%; and **(D)** severe-stage -N3- when the proportion was higher than 75%. These micrography pictures have been taken by the doctoral candidate for this thesis within the framework of Scuba Cancers ERC project.

Several studies established scales of haemocytology to quantify the severity and progression of HN (Carballal *et al.*, 2015); Figure 17 shows a haemocytological HN scale for the species *C. edule*.

1.3.4. GENETIC INSIGHTS

The genetic alterations that characterize HN in bivalves remain largely unexplored. The investigation of the molecular basis of HN has been mainly focused on the genetic characterization of *p53*-family proteins (Muttray and Vassilenko, 2018). Analysis of HN in soft-shell clams (*Mya arenaria)* revealed that p53 proteins are sequestered in the cytoplasm of neoplastic haemocytes by the action of mortalin-like proteins, resulting in loss of wild-type p53 function (Walker and Böttger, 2008). In mussels, researchers suggested an oncogenic role of a truncated p53-family isoform, and up-regulation of a Mdm2-like protein as a potential negative regulator of p53-family (Muttray, Schulte and Baldwin, 2008; Muttray *et al.*, 2010). In cockles, HN shows high expression of mutant p53 protein in neoplastic samples, which is not expressed in the disease-free samples (Díaz *et al.*, 2010), and higher transcriptional expression of *ras* in only some stages of the development of the disease (Ruiz *et al.*, 2013).

In terms of genetic instability, it has been suggested that the induction of retrotransposons could accelerate the progression of cancer (Arriagada *et al.*, 2014). In addition to the reverse transcriptase activity found in neoplastic samples (Oprandy *et al.*, 1981; Oprandy and Chang,

1983; House, Kim and Reno, 1998; Romalde *et al.*, 2007; AboElkhair, Siah, *et al.*, 2009; AboElkhair, Synard, *et al.*, 2009; Manso *et al.*, 2012), RNA sequencing of haemolymph from cancer and non-cancer soft-shell clams (*M. arenaria*) allowed the identification of the Steamer retroelement that was highly active neoplastic cells. of neoplastic. Its DNA characterization revealed an element with long terminal repeats encoding a single large protein with similarity to mammalian retroviral Gag-Pol proteins. DNA copy number of Steamer per genome was present at high levels in cancer cells indicating extensive reverse transcription and retrotransposition (Arriagada *et al.*, 2014).

1.3.4.1.Chromosomal abnormalities

As described for human cancers in *section 1.1*, chromosomal abnormalities and polyploidy have also been described in several bivalve HNs (Carballal *et al.*, 2015). In cockles HN, the number of chromosomes (Figure 18A-B) do not correspond with those of a healthy diploid cockle cell (2n=38) and the range of chromosomes found (41-145) are wider than the range detected in other bivalve HNs (Diaz *et al.*, 2013). In other species' HNs, it has been described a similar feauture, ranging from 44-80 chromosomes in *Mya arenaria* (Muttray and Vassilenko, 2018) to 59-105 chromosomes in *Macoma balthica* (Smolarz *et al.*, 2005).

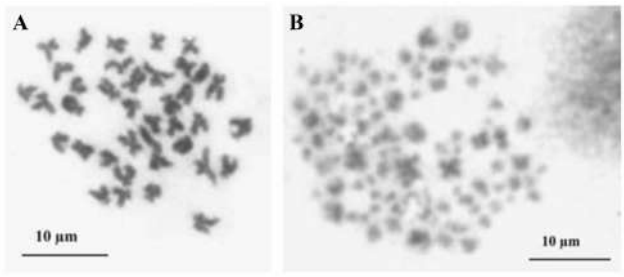


Figure 18. Micrographs of metaphases (Giemsa staining) from gill cells of **(A)** a healthy cockle *C. edule* showing the standard diploid set of 2n=38 chromosomes and **(B)** a severe HN cockle showing a higher number of chromosomes. Adapted from Díaz et al., 2013; reprinted with permission from Elsevier Ltd., see Appendix H.

Flow cytometry analysis showed a cell population of larger and more complex cells in concordance with histological and ultrastructural characteristics. Moreover, neoplastic cells showed a variable ploidy value ranging between 3.1n and 15.2n while healthy cockle cells showed the two expected peaks of DNA content – 2n and 4n (Diaz *et al.*, 2013).

1.3.4.2.Clonal transmission analysis

Metzger et al. (2015) discovered the clonal transmissible nature of HN in the soft-shell clam *Mya arenaria*, when they observed that neoplastic cells from different individuals shared common retrotransposon integration sites that were not present in the normal tissues from the same diseased animals. The analysis of microsatellite variation (Figure19A) and mitochondrial SNPs provided further evidence confirming the monoclonal origin of HN in clams (Figure19B), that is HN from different clams are descendants of the same clone, sharing common alleles that are different from their hosts.

HNs of other species were investigated the following years using nuclear and mitochondrial markers in all cases, more details and references can be found in Table 1. In common cockles, analyses of mitochondrial DNA and microsatellites on neoplastic haemocytes isolated from six diseased cockles revealed the existence of at least two unrelated cancer clones in cockle HN (Metzger *et al.*, 2016). This finding is important because it demonstrates the polyphyletic origin of cockle HN, which strongly suggests that many other unrelated clonal lineages are possible and that cockles are genetically or behaviourally predisposed to develop transmissible cancers (Yonemitsu *et al.*, 2019). These two cancer lineages genetically identified in cockles correspond with the HN subtypes previously described with light microscopy (Figure 15).

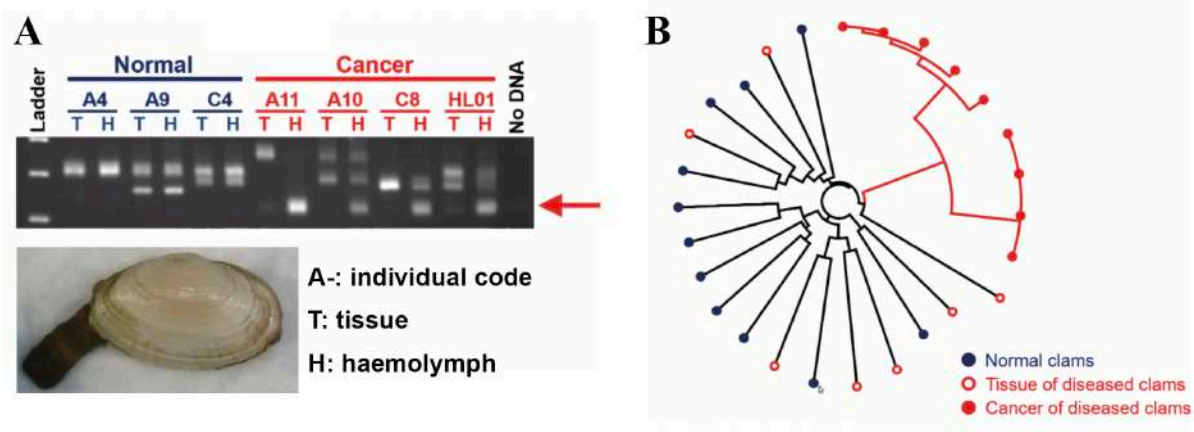


Figure 19. Analysis of transmissible cancer in the soft-shell clam. **(A)** Microsatellite loci amplified in tissue -T- and haemolymph -H- from normal/healthy and diseased -neoplastic- clams; picture of a soft-shell clam is displayed below the electrophoresis gel. **(B)** Neighbour-joining phylogenetic tree based on nine microsatellite loci showing a monophyletic origin of hemic neoplasia. Adapted from Metzger et al., 2015; reprinted with permission from Elsevier Ltd., see Appendix H.

1.3.4.3.Bivalve references genomes

Next-generation sequencing (NGS) technologies are revolutionizing life sciences (Ellegren, 2014). Until recently, genome sequencing projects were limited to biomedical model organisms and required the efforts of large consortia (Ekblom and Wolf, 2014). However, substantial advances in NGS sequencing technologies, combined with lower costs, have allowed the rapid growth of new fields such as marine genomics (Kelley *et al.*, 2016; Van Nimwegen *et al.*, 2016). Despite the fact that the number of reference genomes of marine organisms is considerably lower than that of terrestrial species, the recent increase of these generates an important potential to answer the questions of marine biology from the genomic point of view (Kelley *et al.*, 2016).

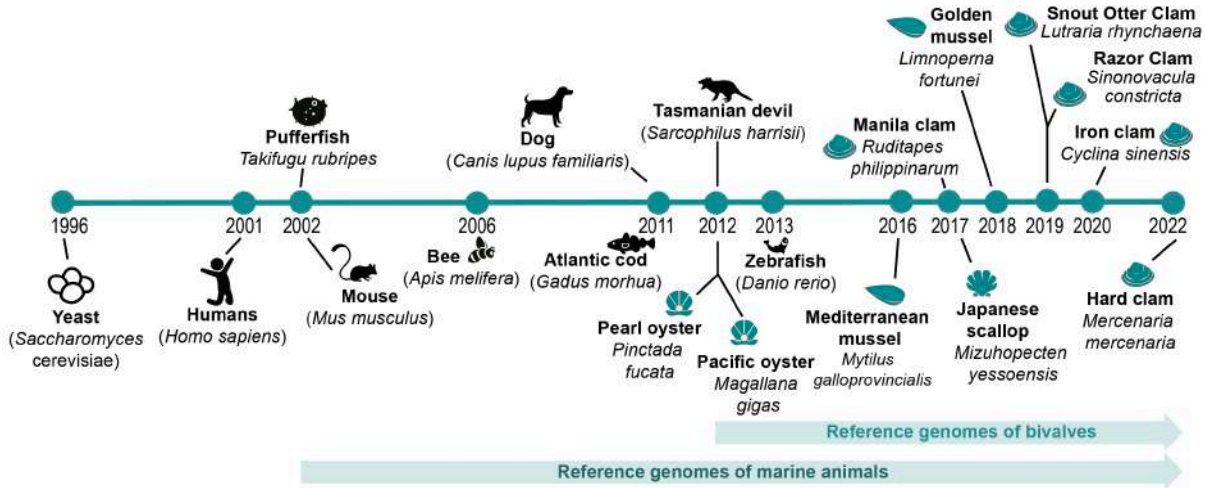


Figure 20. Dates of some releases of references genomes, not all reference genomes have been compiled, just a few are shown to give a general idea. Yeast strain S288c, was sequenced and released in 1996 being the first complete, high quality genome sequence of an eukaryal organism (Goffeau *et al.*, 1996). In 2001, the Human Genome Project and the company Celera Genomics published human genomes (Craig Venter *et al.*, 2001; Lander *et al.*, 2001). Surprisingly, the first publicly available draft vertebrate genome to be published after the human genome was from a marine organism, the pufferfish (Aparicio *et al.*, 2002). In the following years, many reference genomes were published such as the bee (Weinstock *et al.*, 2006), the domestic dog (Boyko, 2011), the Tasmanian devil (Murchison *et al.*, 2012), the Atlantic cod (Star *et al.*, 2011) or the Zebrafish (Howe *et al.*, 2013). The first reference genome of a bivalve was the one of pearl oyster (Takeuchi *et al.*, 2012) and some months later the Pacific oyster genome was also released (Zhang *et al.*, 2012). Recently, other bivalve genomes have been published such as the Mediterranean mussel (Murgarella *et al.*, 2016), the Japanese scallop (Wang *et al.*, 2017),

the manila clam (Mun *et al.*, 2017), the golden mussel (Uliano-Silva *et al.*, 2018), the snout otter clam (Thai *et al.*, 2019), the razor clam (Ran *et al.*, 2019), the iron clam (Wei *et al.*, 2020) or the hard clam (Farhat *et al.*, 2022).

Although the first eukaryotic reference genome was from a model organism, today most species sequenced are non-model organisms. However, the list is biased in favour of certain taxonomic groups, for example: more than 0.1% of all vertebrate genomes have already been sequenced, with mammals being the most characterized group (Ellegren, 2014). Few genomic resources currently exist for the invertebrate organisms although they represent 95% of animal biodiversity (Lopez *et al.*, 2019).

The phylum Mollusca is one of the most diverse groups of animals since it comprises eight lineages, Bivalvia being one of the largest phyla. Bivalvia class includes ~ 20,000 living species but the number of genomic resources available in public databases for these organisms is quite limited, and generally limited to their transcriptomes (Takeuchi *et al.*, 2012; Murgarella *et al.*, 2016). To date there is no available reference genome for common cockles *Cerastoderma edule*, lagoon cockles *C. glaucum*, warty venus clam *Venus verrucosa* or striped venus clam *Chamelea gallina*.

Genome size estimation has been shown to be more accurate using the k-mer method than flow cytometry (Guo *et al.*, 2015; He *et al.*, 2016) because the latest method quantifies total cellular DNA without discriminating nuclear genetic material. In fact, several bivalve reference genome projects (Table 3) have revealed discrepancies between both methods (Elliott and Gregory, 2015; Murgarella *et al.*, 2016). In most cases, the differences between the sizes of the genomes of closely related species are due to the variation in the number of repetitive sequences (He *et al.*, 2016).

Table 3. Comparative of size estimation with two methods (K-mers and flow cytometry) in several bivalve species.

Species	Sex	Tissue	Genomic library	Sequencing technology	Size estimation		Karyotype
					K-mers	Flow cytometry	
Cerastoderma edule	Male	Haemolymph	350 pb	Illumina paired end	0,8 Gb	1,34 Gb*	2n = 38
Pinctada fucata martensii	Male	Gonad	4 kb, 10 kb	Roche 454 GS-FLX paired-end	*NA*	1,15 Gb	2n = 28
			3 kb, 10 kb	Illumina mate-pair			
			(Takeuchi *et al.*, 2012)				
Magallana gigas[5]	Female	Adductor muscle, gills, mantle, gonad	170 pb, 500 pb, 800 pb	Illumina paired end	545 Mb (17-mer)	637 Mb	2n = 28
			5 kb, 10 kb, 20 kb	Illumina mate-pair			
			(Zhang *et al.*, 2012)				
Mytilus galloprovincialis	NA	Adductor muscle	180 pb, 500 pb, 800 pb	Illumina paired end	1,6 Gb (17-mer)	1,4 – 1,9 Gb	2n = 28
			(Murgarella *et al.*, 2016)				
*Mizuhopecten yessoensis***	Male	Adductor muscle	180 pb, 300 pb, 500 pb	Illumina paired end	1,43 Gb (19-mer)	1,44 Gb	2n = 38
			2 kb, 5 kb	Illumina mate-pair			
			(Wang *et al.*, 2017)				

* published data in Rodriguez-Juiz, Torrado and Mendez, 1996
**scientific name used in the article: *Patinopecten yessoensis* (Jay, 1857).

[5] Citation referred as *Crassostrea gigas,* current taxonomic name *Magallana gigas* (Salvi and Mariottini, 2017).

Quality assemblies for bivalve genomes are usually challenging due to several factors such as the composition of repetitive elements and high levels of heterozygosity (Gomes-dos-Santos *et al.*, 2020) nevertheless, sequencing a reference genome offers valuable information on the genes involved in disease resistance and allows to understand the genetic alterations that lead the infection.

1.3.5. COMPARISON WITH MAMMAL CONTAGIOUS CANCERS

Notably, bivalve transmissible neoplasia (BTN) differs in several characteristics from known mammal contagious cancers previously described in CTVT and DFTD (*Section 1.2.2*). Table 4 summarizes common and different characteristics of BTNs against mammal contagious cancers that have already been reported in other sections of this Chapter.

Table 4. Comparison of the known naturally occurring contagious cancers.

	Bivalve Transmissible Neoplasia	**Canine Transmissible Venereal Tumour**	**Devil Facial Tumour Disease**
Host species	**Clams, mussels, and cockles**	Dogs	Tasmanian devil
Species class	**Bivalvia**	Mammalia	Mammalia
Spread locations	**Oceans and seas of America, Asia and Europe**	All continents except Antarctica	Tasmanian island
Oldest description of tumour	**1969**	1810	1996
Transmission	**Probably through water filtration**	Sexual intercourse	Biting
Tumour type	**Leukaemia-like (neoplastic cells found in haemolymph)**	Sarcoma	Sarcoma
Most common tumour location	**Haemolymph**	Genitals	Face
Cell-of-origin	**Not investigated yet. Most likely a haemocyte.**	Myeloid cell	Schwann cell
Tumour age	**At least probably 40 years**	~ 8,000 years	At least 25 years
Lineages	**At least 8**	1	2
Mitochondria acquisition	**Not known**	Yes	Not known
Genetic diversity	**High**	Low	Low
Regression	**Yes**	Spontaneously or after treatment with vincristine	Yes
Interspecies transmission	**Yes**	No	No

1.4. BEYOND THE LIMITS OF METASTASES

Bivalve transmissible neoplasia are the only natural-occurring contagious cancers that have been able to infect animals of a different species of the one that originated the cancer. In this section, we will review interspecific contagion cases found in bivalves and the mechanisms that animals have to fight against contagious cancers.

1.4.1. INTERSPECIES TRANSMISSION OF CANCER

Cancer contagion is rare because cancer cells need to overcome the shedding from the origin host, survive in the sea water, invade and adapt to a new host and its immunological responses (Ujvari, Gatenby and Thomas, 2016). Therefore, a cancer capable of transmission from one species to another seems even rarer. However, three cases have been described (Table 1, Figure 13), one of which was reported in the article reproduced in the *Chapter 3* of this thesis.

The cancer observed in golden carpet shell clams (*Polititapes aureus*) was found to have originated in a different, but related, species, the pullet carpet shell clam (*Venerupis corrugata*), first ever known case of interspecies transmission of cancer (Figure 21A). Surprisingly, only sporadic cases of HN are found in the pullet shell clams that co-habitat with golden carpet shell clams pointing to a potential adaptation of the pullet shell clam to resist infection by the transmissible cancer that first arose in a member of its own species; despite this, the cancer has survived by engrafting to a new host species (Metzger *et al.*, 2016; Murchison, 2016).

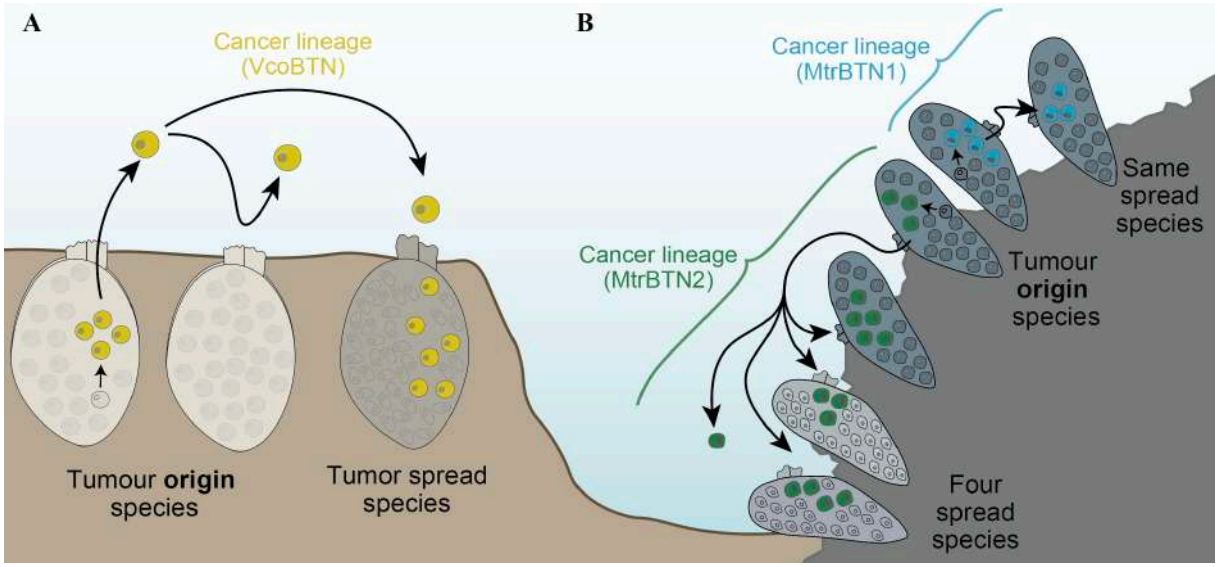


Figure 21. Interspecies transmission scenarios. **(A)** The pullet carpet shell clam originated a contagious cancer that is no longer spreading among its species, but it engrafted into the golden carpet shell clams. **(B)** Two cancer lineages originated in the foolish mussel; both are spreading among this species but one of them has also engrafted into three additional related species.

Two cancer lineages arose in the foolish mussel (*Mytilus trossulus*) and are currently spreading among the population (Table 1). However, one of those cancer lineages, in addition to infecting foolish mussels (Yonemitsu *et al.*, 2019; Skazina *et al.*, 2021), it has been able to engraft into three additional mussel species – the blue mussel (*M. edulis*; Yonemitsu *et al.*, 2019), the Chilean mussel (*M. chilensis*; Yonemitsu *et al.*, 2019) and the Mediterranean mussel (*M. galloprovincialis*; Hammel *et al.*, 2021) – and even into hybrid mussels (Figure 21B), thus extending the known spreading complexity of BTN (Hammel *et al.*, 2021).

1.4.2. MECHANISMS AGAINST TRANSMISSION

Metastasize in a new host does not only mean overcoming physical barriers but also the immunological response of the host, in some cases natural regression of cancer is accomplished by the host. HN usually results in death of the individual, though remission has been known to occur (Elston, Kent and Drum, 1988; Collins and Mulcahy, 2003). Therefore, regression has been reported in all contagious cancers (Table 4) although CTVT and DFTD have been more studied leading to the fact that both evade the recognition of major histocompatibility complex (MHC).

1.4.2.1. Bivalves' defences and immunity

The immune system in mammals is traditionally classified into two categories, the innate system and the highly specific system called 'adaptive' or 'acquired' associated with the existence of immune memory, which allows it to develop a better defence during a second infection by the same pathogen strain. On the other hand, invertebrates have long been considered to rely exclusively on nonspecific innate immune mechanisms (Escoubas *et al.*, 2016; Gerdol *et al.*, 2018). However, recent studies have provided arguments for the existence of a form specific recognition and immune memory (Pradeu and Du Pasquier, 2018; Miccoli *et al.*, 2021) but the molecular mechanisms remain poorly understood (Odintsova, 2020).

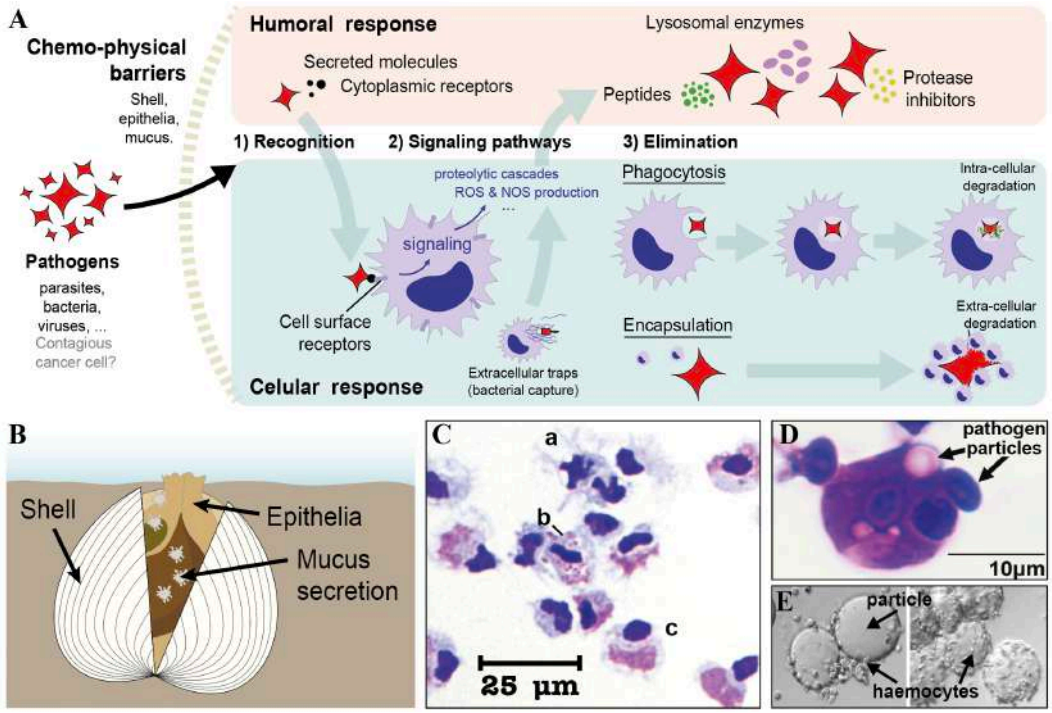


Figure 22. Overview of defence mechanisms and immune responses in bivalves. **(A)** Schematic representation of defences and immunity; once a pathogen has overcome the chemo-physical barriers, humoral and cellular responses take place. Most invaders are (1) recognized by pattern-recognition receptors that bind to conserved molecules expressed on microbial surfaces and trigger the (2) activation of intracellular signalling cascades. Small invaders are (3) eliminated by the phagocytic haemocytes, while large invaders are eliminated by encapsulation. **(B)** First bivalve defences are chemo-physical barriers such as the shell, the epithelia, and the mucosal layer. **(C)** Haemolymph cell monolayer from a healthy cockle where three types of haemocytes can be observed: (a) hyalinocytes, (b) granulocytes and (c) type III; this micrography picture has been taken by the doctoral candidate for this thesis within the framework of Scuba Cancers ERC project. **(D)** Light microscopic image of the phagocytic haemocytes of the mussel *Lamellidens marginalis* engulfing multiple yeast particles, reproduced from Chakraborty, Ray and Ray, 2021, reprinted with permission from Elsevier Ltd., see Appendix H. **(E)** Encapsulation of particles (positively charged beads) by haemocytes of *Cerastoderma edule*, reproduced from Wootton, Dyrynda and Ratcliffe, 2006; reprinted with permission from Company of Biologists Ltd., see Appendix H.

Shells of molluscs act as **physical barriers** (Figure 22A-B) that prevent some pathogens from penetrating into the host's body (Al-Khalaifah and Al-Nasser, 2019). The second physical barrier beyond the shell is provided by the skin and epithelial cells produce and secrete a wide range of bioactive molecules that are embedded in mucus. All mucosal epithelia of bivalves are capable of endocytosing biotic and abiotic particles and colloids (Allam and Raftos, 2015).

Bivalves possess an innate immune system composed of humoral factors and cell-mediated mechanisms (Figure 22A). **Humoral factors** include lectins (agglutinins, opsonins), lysosomal enzymes (phosphatase acid, lysozyme and various hydrolytic enzymes), antimicrobial peptides, protease inhibitors and cytokine-like molecules among others (Chu, 1988; Villalba *et al.*, 2008). Circulating cells known as haemocytes (Figure 22C) are the main effectors of the **cellular response** although they are also involved in many other fundamental roles such as nutrient transport, shell calcification, digestion and excretion processes or wound repair (Escoubas *et al.*, 2016). One of the first reactions observed following stress is an increase in the quantity of circulating haemocytes and haemocyte infiltration within the affected tissues (Mayrand, St-Jean and Courtenay, 2005; Hammel, 2022).

If a pathogen overcomes the chemo-physical barriers, the first step is its **recognition** by (i) secreted molecules, (ii) cell surface receptors or (iii) cytoplasmic receptors that bind to molecules expressed on microbial surfaces and trigger the activation of intracellular **signalling cascades**. Haemocytes produce reactive oxygen species (ROS) and reactive nitrogen species (RNS) in response to endogenous and exogenous stimuli. Bivalve genomes also encode components of the mitogen-activated protein kinase pathway (MAPK) and the janus kinase signal transducer and activators of transcription (JAK/STAT) pathway; pathways that are key players in multiple processes including cell growth and differentiation as well as immunity and inflammatory processes in mammals but their role in bivalves is yet unknown (Escoubas *et al.*, 2016). In the end, two mechanisms are used for the **elimination** or destruction of pathogens: phagocytosis and encapsulation. Phagocytosis (Figure 22D) involves the migration toward a chemical stimuli released by non–self, its recognition and attachment, believed to be mediated by lectins, then the internalization to finalize with an intracellular degradation (Soudant, E. Chu and Volety, 2013). When phagocytosis fails or when particles are too large to undergo phagocytosis, haemocytes are recruited in large numbers to surround and encapsulate (Figure 22E) the invader pathogen and to release cytotoxic products for extracellular killing (Allam and Raftos, 2015).

Specific recognition of self/nonself discrimination has been observed and it has been suggested that it is due to the existence of polymorphic and diversified putative immune receptor variants that vary considerably between individuals, yielding an enlarged repertoire of putative recognition molecules. If invertebrates possess diversified immune receptors involved even partly in the specific recognition of pathogens, it can be speculated that they also possess a kind of immune memory. Long-term increase in antimicrobial response after an infection and enhanced resistance to a second infection has been observed in bivalves and this acquired resistance has been named immune priming (Escoubas *et al.*, 2016).

1.4.2.2. Immunity in the context of bivalve contagious cancers

Most bivalve immune responses have been characterized in the context of bacterial or viral infections and eukaryotic parasites. Thus, in the context of a contagious cancer cell –eukaryotic microparasite genetically close to its host– recognition pathways by the immune system remain unknown (Hammel, 2022). Eight cancer lineages are currently spreading among bivalve species, two cancer lineages have arisen in at least two bivalve species, cases of interspecies

transmissions have been reported; by looking to these findings (Table 1), it does not seem that bivalves are usually successful in the recognition and elimination of contagious cancer cells.

Bivalves do not possess any known form of histocompatibility barriers (Báez, 2019) which might explain the susceptibility to contagious cancers, and probably contributes to explaining the striking frequency at which such diseases have been found to affect bivalves, relative to their apparent rarity in vertebrates (Metzger et al., 2016; Metzger and Goff, 2016). Though, most BTNs are restricted to the species where they were originated suggesting that bivalves may possess unidentified mechanisms for preventing the engraftment of cells from another species (Mateo, MacCallum and Davidson, 2016; Báez, 2019).

1.4.2.3. Regression of contagious cancers

Initially, researchers thought that DFTD was able to transmit from one animal to another due to the lack of genetic diversity in this species, but experiments showed that devils are able to recognise as foreign and reject skin drafts from other devils (Kreiss *et al.*, 2011; Miller *et al.*, 2011; Siddle *et al.*, 2013). However, surface molecules for immune system recognition of DFTD cancer cells are not present which allows the cancer to hide and grow uncontrollably (Siddle *et al.*, 2013). Moreover, when comparing devils that had tumour regression with devils with aggressive tumours, putative tumour suppressor genes were associated with tumour regression (Siddle *et al.*, 2013). In addition to those host mechanisms, mutations associated with regression have also been found in tumours. A single point mutation in the 59 untranslated region of the putative tumour suppressor *RASL11A* significantly contributes to tumour regression. *RASL11A* was found to be expressed in regressed tumours but silenced in wild type, non-regressed tumours, consistent with RAS pathway downregulation in human cancers (Margres *et al.*, 2020).

In the case of HN, encapsulation previously described in bivalves for the elimination of pathogens has been occasionally detected in HN cockles (Figure 23) while phagocytic activity of haemolymph cells has been reported to decrease in the late stages of the disease. In fact, remission of the disease does not seem frequent. The progressive and lethal nature of this contagious cancer supported by histological observations would presage a significant mortality in the cockle population consistently with the prevalence recorded (Díaz *et al.*, 2016).

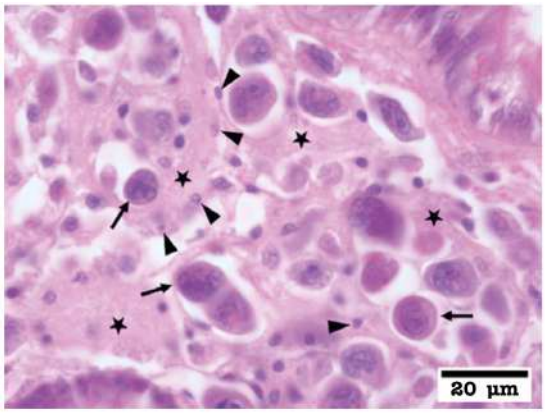


Figure 23. Histological section (H&E stain) of the digestive gland showing the encapsulation of cancer cells (arrows) by layers of connective fibres (stars) and fibroblasts (arrowheads) in cockles. Adapted from Díaz et al., 2016; reprinted with permission from Elsevier Ltd., see Appendix H.

In cockles, the study of HN cells has shown a higher lysosome biovolume, non-specific esterase activities and an increase of ROS production which are parameters related with cancer and host defence (Díaz *et al.*, 2011).

1.5. HYPOTHESIS AND OBJECTIVES OF THIS THESIS

This doctoral thesis, *Evolution of Bivalve Transmissible Cancers*, intends to identify the genomic alterations that shaped the evolution of marine transmissible cancers from the origin to the genetic causes that made them evolve as parasitic clonal lineages in the marine environment, trying to illuminate processes that make these cancers contagious and identify potential targets for advancing their prevention, detection, monitoring and/or treatment.

Following, the goals and hypothesis behind the next chapters of this thesis are described:

Chapter 1 and 5 present an introduction and general discussion to dress the experimental research presented on chapters 2, 3 and 4.

Chapter 2 presents the evolutionary history of cockles' HN throughout Europe. Although the transmissible nature of cockles' HN has already been reported, there are still many unanswered questions that deserve specific attention.

Hypothesis

Contagious cancers in cockles have arisen at least twice, studying the clonal stratification of these cancers, will allow us to understand the evolution and the transmission of the disease, and could help to establish a classification of HN according to phenotypes or transmissible behaviours.

Objectives

- Determine the prevalence of HN in common cockles throughout its distribution range
- Analyse the number of independent cancer lineages by means of mitochondrial DNA and validate them with nuclear markers (*e.*g., Microsatellites)
- Estimate the potential region and date of cancer origin
- Assess the genetic architecture of cockle transmissible cancer cells

Chapter 3 presents the cell-of-origin of cockles' HN. Leukaemia-like cancers have been reported on many bivalve species since the late 60s but the histogenesis of this disease has not been explored yet.

Hypothesis

HN is generally considered to be a sarcoma (neoplasia of mesoderm-derived tissues) although a haematopoietic and a gonadal origin have also been proposed. Hence, elucidating the cell-of-origin of known contagious cancer lineages of cockles' HN might offer new insights to understand the evolutionary changes that underlie a cell to become cancerous and develop a metastatic behaviour that goes beyond the body limits.

Objectives

- Analyse the diversity of gene expression among different healthy cockle tissues and larval stages
- Identify the potential cell-of-origin of cockles' HN lineages.
- Examine the histogenesis differences and similarities of both cancer lineages.

Chapter 4 presents the report of a novel transmissible cancer affecting bivalves which was originated in a different species.

Hypothesis

It has been demonstrated that transmissible cancers can be naturally transmitted between different species. Investigating other bivalve species with no HN clearly reported to date, such as the warty venus clam, could give as new models to study interspecific transmission of cancer in bivalves between close species.

Objectives

- Examine warty venus clams from different locations for leukaemia-like cancers
- Morphological and karyotypic characterization of cancer cells
- Evaluate its contagious nature and identify the species of cancer origin

Chapter cover shows a sampling of cockles in Cork (Ireland) in April 2019. People in the photograph: from right to left, Dr. Seila Díaz, Eoin MacLoughlin, and the doctoral candidate Alicia L. Bruzos. All people in the photograph have granted written permission to reproduce the picture in this thesis.

Acknowledgments. Sara Rocha, Laura Tomás and Tamara Prieto provided essential knowledge and resources for the clonal deconvolution and phylogenetic analysis of this chapter.

Chapter 2.
Evolution of cockle transmissible cancers

"The most dangerous cancer cells are actually the ones that are more like stem cells, which have this ability to produce themselves over and over again." Elizabeth Blackburn

"It is not the strongest of the species that survives, nor the most intelligent; it is the one most adaptable to change." Charles Darwin

2.1. BACKGROUND

2.1.1. BIVALVE TRANSMISSIBLE NEOPLASIA

Bivalve transmissible neoplasia (BTN) are naturally occurring leukaemia-like cancers that are transmitted between bivalve individuals. They behave as clonal cell lineages that spread within the populations by the transfer of living cancer cells most likely using ocean currents (Metzger and Goff, 2016).

Marine bivalves are molluscs enclosed by a shell consisting of two hinged parts and the majority are filter feeders. They include clams, oysters, mussels, scallops, and cockles among others (Gosling, 2015). The existence of contagious cancers infecting these animals has only been confirmed in nine species (Table 1). In this study we have focused on the BTN affecting common cockles.

2.1.2. COMMON COCKLES

The common cockle *Cerastoderma edule* (Linnaeus, 1785) is a bivalve mollusc (Figure 24A) with a wide geographical distribution along the north-eastern Atlantic coastline from the western region of the Barents Sea to the Iberian Peninsula, and south along the coast of West Africa to Senegal (Tebble, 1976; Maia, Barroso and Gaspar, 2021). This species lives buried just under the surface in clean sand, muddy sand, mud or muddy gravel bottoms and it is commonly found in intertidal flats and shallow subtidal areas of estuaries, coastal lagoons and sheltered coastline bays (Kater, Geurts Van Kessel and Baars, 2006; Maia, Barroso and Gaspar, 2021).

The common cockle (hereafter 'cockle' or *C. edule* as appropriate) is one of the main non-cultured bivalve species harvested in western European waters (Carss *et al*., 2020). Cockles are

one of the most abundant mollusc species in European bays and estuaries where population densities of 10,000 per m^2 have been recorded (Tyler-Walters, 2007). Animals mature when reaching ca. 20 mm shell length (study performed in the UK) and live up to 10 years in some habitats but more commonly to 2–6 years (Carss *et al*., 2020).

2.1.2.1. Closest species in the area: lagoon cockles

The other European cockle species is the lagoon cockle, *Cerastoderma glaucum* (Figure 24B) which is morphologically similar to the common cockle *C. edule*. Discrimination between the two cockle species can be achieved with qualitative shell characters. For instance, the shell rib-number differ in the two species in a common environment: lagoon cockle has fewer ribs than common cockle because rib-number is directly related to salinity and lagoon cockle usually colonizes areas with lower salinities (Boyden, 1973). Given the difficulties to differentiate the species morphologically, molecular techniques based on sequence differences found in the internal transcribed spacer region (ITS) of the ribosomal DNA of the two cockles have been designed (Freire *et al.*, 2011).

Figure 24. Shells of two cockle species (Courtesy of Olivier Caro). **(A)** Common cockle *Cerastoderma edule*. **(B)** Lagoon cockle *Cerastoderma glaucum*.

Distribution of lagoon cockles is restricted to brackish water habitats, however, both cockles overlap part of their range, coexisting with common cockles in several localities of Portugal, Spain, France and UK. In addition, lagoon cockles are recorded in the Mediterranean, Black and Baltic Sea where common cockles are not found (Carballal *et al.*, 2016).

2.1.2.2. Genetic diversity and population structure

Genetic diversity is crucial for the adaptation of natural populations to environmental changes thereupon, different studies aiming to unravel it found significant differences between regions with high heterozygosity levels and gene flow in particular regions (Hummel, Wolowicz and Bogaards, 1994; Martínez *et al.*, 2013).

Mitochondrial DNA sequencing of cockles within its distribution range showed two differentiated groups in northern and southern areas (Krakau *et al.*, 2012; Martínez *et al.*, 2015). Microsatellites nuclear DNA markers subdivided these groups into: (i) a southern region (cockle populations of Morocco, Portugal, Spain and France up to the English Channel); (ii) an intermediate region including cockle populations from Ireland, Great Britain and southern North Sea (the Netherlands and Germany); and (iii) a northern group (Scotland, Denmark, Norway and Russia) (Martínez *et al.*, 2015; Vera *et al.*, 2021).

The genetic homogeneity detected of northern and southern populations may be the result of both ocean currents and demographic processes that likely play a leading role in connectivity within this group of populations (Martínez *et al.*, 2015). In fact, models of larval dispersal suggested a barrier for larval dispersal linked to the Ushant front that could explain these northern-southern genetic clusters (Vera *et al.*, 2021).

This genetic diversity distribution has some similarity with that of two mussel species that are able to hybridize. In Europe, the regions west and south of the English Channel are dominated by the Mediterranean mussel (*M. galloprovincialis*) while the east and north by the blue mussel (*M. edulis*); being the boundary between the two groups of populations, the same as for common cockles (Krakau *et al.*, 2012).

2.1.2.3. Mortalities and pathologies

During the last decades cockle stocks have shown a progressive declining trend mostly due to mass mortality episodes and recruitment failures; both provoked by climate-related events (Peteiro *et al.*, 2018). However, in addition to parasite infections, a leukaemia-like cancer described in this species is one of the main pathologies affecting mortalities (Díaz, 2015).

Common cockles in the region of Galicia (Northwest of Spain) have undergone an important population decline associated with the parasitic protozoan *Marteilia cochillia* (Villalba *et al.*, 2014). Nonetheless, the leukaemia-like cancer is found at a non-negligible prevalence in many populations (Table 2). Yet, none of the parasites and diseases reported in common cockles is harmful for human consumers which means that these parasites are not zoonotic as far as we know (Montaudouin *et al.*, 2021).

2.1.3. COCKLE HAEMIC NEOPLASIA

Hemic neoplasia (HN), also known as disseminated neoplasia, manifests itself with the appearance of tumoral cells in the haemolymph (*ie*. circulatory system of these animals) and infiltrating all tissues of the animal in the latest stages of the disease.

The first known reports of HN were made in the late 1960s and it has subsequently been reported in several bivalve species (Carballal *et al.*, 2015). In 2015, the transmissible nature of HN affecting soft-shell clams was demonstrated by studying their DNA (Metzger *et al*., 2015) and the following years several additional HN were corroborated to be contagious (Metzger *et al.*, 2015, 2016; Yonemitsu *et al.*, 2019; Garcia-Souto *et al.*, 2021; Hammel *et al.*, 2021; M. Skazina *et al.*, 2021; Michnowska *et al.*, 2022).

In cockles, HN was reported in the 80's in France (Poder and Auffret, 1986) and Ireland (Twomey and Mulcahy, 1984) and later more populations in southern European countries were reported (Table 2). Cancer cells are morphologically characterized by being bigger and rounder than haemocytes, high nucleus/cytoplasm ratio, frequent observation of mitotic figures, no pseudopods, pleomorphic nuclei and a big nucleolus (Díaz, 2015).

2.1.4. GENOMICS OF COCKLE TRANSMISSIBLE CANCERS

In a similar way to how the haemic neoplasia of American soft-shell clams was described in *Section 1.3.4*, the contagious nature of Galician cockles' HN was demonstrated through a genetic screen of mitochondrial and nuclear DNA (Table 1). Analyses of microsatellites (Figure 25A-B) and mitochondrial DNA (Figure 25B) on neoplastic haemocytes isolated from six diseased cockles revealed the existence of at least two unrelated cancer clones (Ced-a-BTN1 and Ced-b-BNT2) in cockle HN (Metzger *et al.*, 2016).

This finding is important because it strongly suggests the polyphyletic origin of cockle HN, which suggests that many other unrelated clonal lineages are possible and that cockles are genetically or behaviourally predisposed to develop transmissible cancers (Yonemitsu *et al.*,

2019). These two cancer lineages genetically identified in cockles correspond with the HN subtypes previously described with light microscopy (Figure 15).

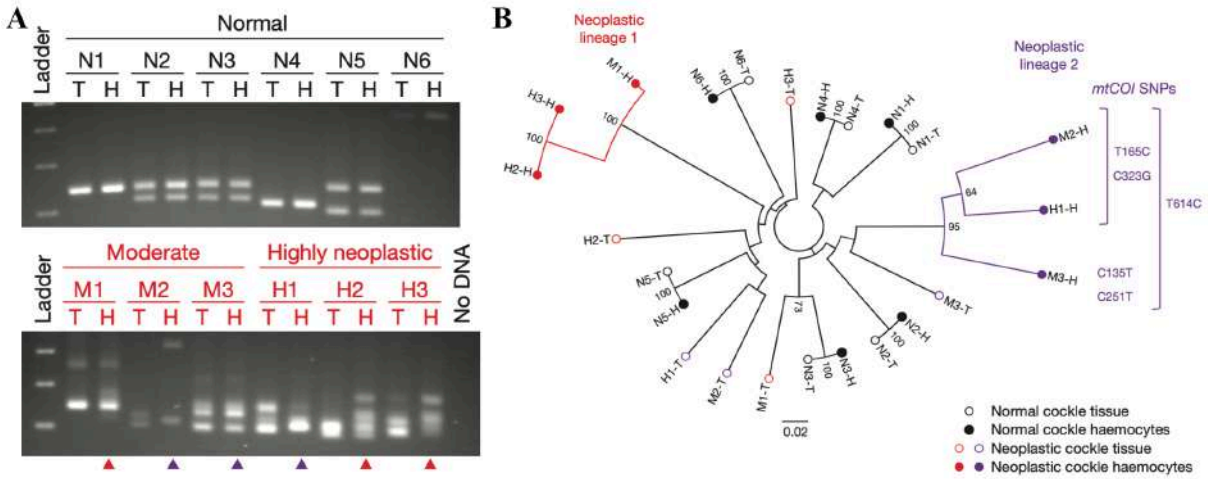


Figure 25. Analysis of cockle transmissible cancers. **(A)** Microsatellite loci amplified in tissue -T- and haemolymph -H- from normal/healthy and diseased -moderate and highly neoplastic- cockles. **(B)** Neighbour-joining phylogenetic tree based on nine microsatellite loci highlighting only bootstrap values over 50 showing a polyphyletic origin of cockle hemic neoplasia; in cancer lineage 2, unique *mtCOI* SNPs are displayed. Adapted from Metzger et al., 2016; reprinted with permission of Springer Nature, see Appendix H.

The genomes of cockles and their transmissible cancer lineages are thus of interest for the insights they may provide into the origins, somatic evolution and population genetics of these recurrent contagious cancers emerging at least twice in this species.

2.2. MATERIALS AND METHODS

2.2.1. SAMPLE COLLECTION

Samples (n=6,719) were always collected from natural beds (Figure 26A-B) throughout 12 countries along the distribution range of *Cerastoderma edule* (FAO, 2019) from the northern Barents Sea and to the south coast of Morocco (Figure 26D, Appendix A: Supplementary material - sampling summary table).

All samples arrived at the laboratory alive and were maintained in a tank with closed-circuit of running seawater for 48 h before the diagnosis and further procedures (Figure 26C), animals from different sampling locations were never mixed in the same tank and bleach cleanings of the tanks were performed between sample arrivals. Animal facility details and ethical approvals are disclosed in Appendix G.

2.2.2. SAMPLE DIAGNOSIS

2.2.2.1. Cytology

HN was firstly diagnosed by examination of haemolymph cell monolayers. Haemolymph was withdrawn from the adductor muscle of every bivalve sample using a 23-gauge needle attached to a 5 ml syringe (Figure 26E). 50 µl of haemolymph were mixed with 150 µl of cold modified Alsever's anti-aggregate solution (Bachère, Chagot and Grizel, 1988) and cyto-centrifuged onto slides (130 g, 7 min, 4 °C). The haemolymph cell monolayers were fixed and stained (Figure 26F) with the kit Hemacolor (Merck) and examined on a Leica CTR6 LED light microscope for HN diagnosis and cell counting.

Cockles were ranked according to a scale of disease severity (Figure 17) by manually counting 500 cells: **non-affected** (N0), when not a single cancer cell was seen under the microscope; **early-stage cancer** (N1), when individuals showed proportion of cancer cells lower than 15% in the haemolymph cell monolayers; **medium-stage cancer** (N2), when the proportion ranged from 15% to 75%; and **severe-stage** (N3), when the proportion was higher than 75% (Diaz *et al.*, 2010)

2.2.2.2. Histology

Previous diagnosis was verified through histological sections and samples with unclear or no cytological diagnosis confirmed or discarded. In addition, HN samples were categorized in type A or B according to its morphological characteristics.

For each specimen, 5 mm section containing almost all organs (visceral mass, gills, mantle, and foot) were dissected (Figure 26G-H), fixed in Davison's solution (10% glycerin, 20% formaldehyde 36–40%, 30% ethanol, 30% filtered seawater, 10% acetic acid) and embedded in paraffin. Then, 5 µm thick sections were micro-dissected, stained with Harri's haematoxylin and eosin and examined using a light microscope for histopathological analysis.

Similarly to the preceding cytological classification, cockles were labelled as follow: **non-affected** (N0), no cancer cells are detected in the tissues; **early-stage cancer** (N1), detection of isolated cancer cells in the tissues; **medium-stage cancer** (N2), presence of small foci in one or more organs and **severe-stage** (N3), involvement of most organs by foci or masses of neoplastic cells (modification of Diaz *et al.*, 2016). Neoplasia types were differentiated by size

and cell interaction where (i) type A were larger and more scattered and (ii) type B smaller, clustered and more compressed (Carballal *et al.*, 2001; Figure 15).

Infiltration of cancer cells through organs was examined being the most infiltrated regions the vessels and sinuses of circulatory system, connective tissue of gonad, digestive gland and gills. However, organs made up mostly of muscle tissue showed less infiltration. Therefore, the apical foot area, edge of the mantle or the adductor muscle were usually selected as matched-normal of HN.

Figure 26. Sample collection and processing. **(A)** Cockle fishing in Noia, Spain. **(B)** Digging in the sand to find cockles in Noia, Spain. **(C)** Cockle maintenance in tanks in the laboratory. **(D)** Map showing the distribution range of common cockles (Cerastoderma edule), the locations in which we have collected samples and the codes used for labelling. **(E)** Haemolymph extraction from the adductor muscle of a cockle. **(F)** Staining drying of cyto-centrifugated preparations of haemolymph. **(G)** Opened common cockle showing the soft tissues inside the shell. **(H)** Dissected cockle showing the shell (left) and the tissues (right). **(I)** Colchicine overnight treatment of cockles to obtain chromosomes.

2.2.3. SAMPLE STORAGE

Hemolymph withdrawn from adductor muscle was centrifuged to eliminate plasma, cell pellet was mixed with 150 µl RNAlater. Dissected tissues were separated (Figure 26H), and all tubes frozen in liquid nitrogen before transferring them to a -80 freezer for long-term storage. In some sampling points where resources were limited, tissues and hemolymph were preserved in ethanol 100%.

Notation of samples followed a code of Country, Place, Specific name, Year, Sample Number and Tissue (*e.g.*, ENCE16/154H would be a sample from Spain, Noia, *Cerastoderma edule*, cockle numer 154, Haemolymph; Appendix A: Supplementary material – schematic workflow of sample processing).

2.2.4. DNA ISOLATION, WGA AND SEQUENCING

DNA was isolated using QIAamp DNA Mini Kit (Qiagen) and an additional precipitation step with 600 µL of 20% SDS/CH_3COOH (70ºC, 10 minutes) was included for the precipitation of histones and other DNA binding proteins right after the RNAse digestion. Along with the proteinase K, 20 µL β-mercaptoethanol reducing agent was used.

DNA purity was evaluated with Nanodrop One (Thermo Fisher Scientific), DNA yield was measured in a Qubit fluorometer (Thermo Fisher Scientific), and DNA integrity was evaluated in a 4200 TapeStation System (Agilent). When the quality controls showed insufficient DNA quantity to perform whole-genome sequencing, an intermediate whole-genome amplification (WGA) step was done using REPLI-g Mini Kit (Qiagen). This step was avoided as much as possible performing another extraction from the same or another tissue to avoid the bias produced by WGA in downstream analysis.

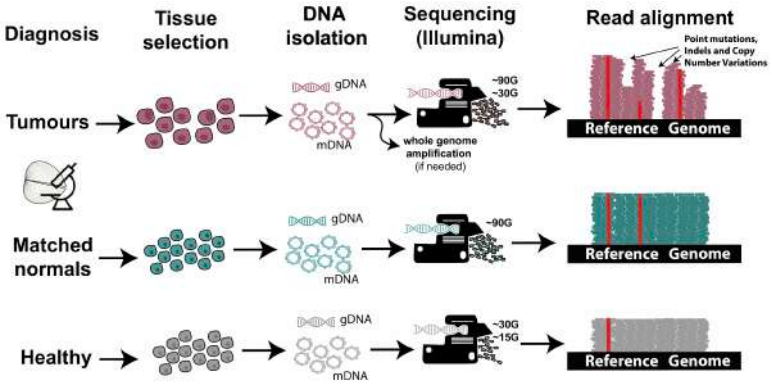


Figure 27. Schematic representation of the steps performed from diagnosis to read alignment for the three sample types sequenced (tumours, matched-normals and healthy).

Whole-genome sequencing (WGS) libraries were prepared and sequenced with 100 bp paired-end reads using the Illumina NovaSeq6000 platform (Macrogen, Seoul, South Korea). Depending on the purpose of the sample (*e.g.*, tumour N3, tumour N2, healthy…), different number of reads (*i.e.*, 15, 30 or 90 Gb) were obtained from each sample (Figure 27).

Selection of samples to be submitted for sequencing was done considering the cancer purity of the haemolymph using the results of cell counting and quality of nucleic acids measured in terms of integrity, purity and concentration. All populations where cancer was diagnosed were included, in cases where no samples met the requirements (*i.e.*, no severe cases of HN or no high purity of DNA), conditions were relaxed.

As not all samples met the requirements (*i.e.*, DNA quantity) for sequencing, whole-genome amplification (WGA) protocol was used in 63% (44/70) of the sequenced tumours to increase the number of samples suitable, however, while 72% (8/11) of Irish samples needed WGA, only 33% of French (1/3) and English (1/3) were done with WGA (Figure 28A, red).

Several matched-normal tissues were sequenced (Figure 28A, blue), most of the matched-normal sequenced were foot samples (55%); although adductor muscles (25%) and mantles (20%) were also sequenced in some cases (Figure 28B), selection of matched-normal tissues was based done based on the general infiltration of cancer cells in tissues of severe-stage cockles (Figure 28C).

Moreover, for cancer samples belonging to populations well represented in our sequencing dataset (*i.e.*, Spain and Portugal) almost half of the samples were sequenced at low coverage (Figure 28A, green).

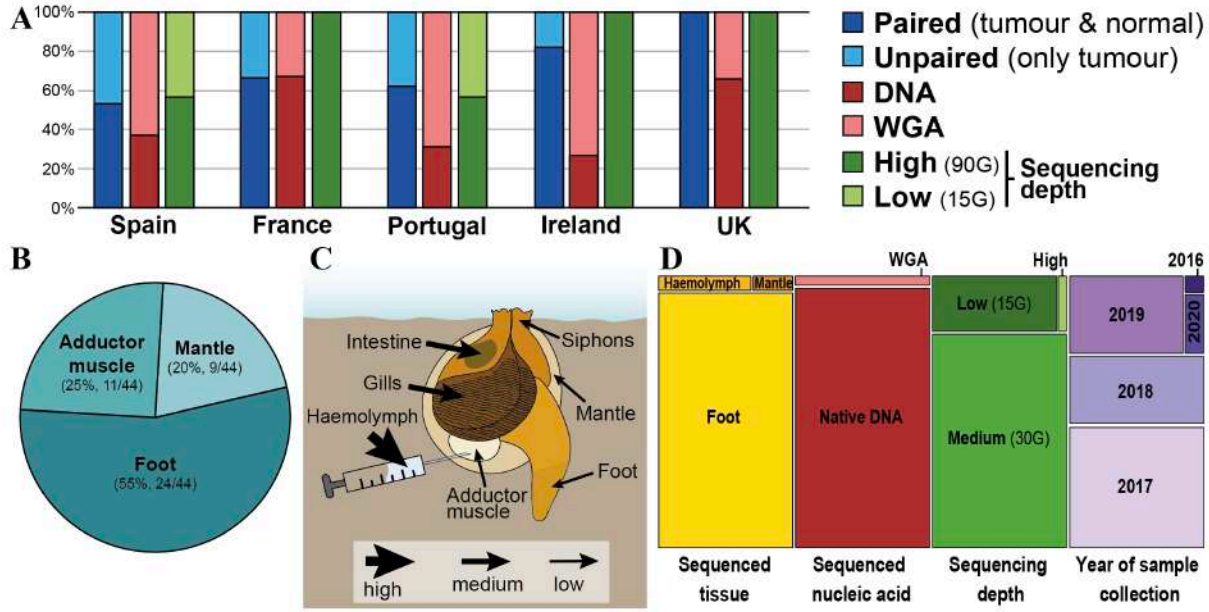


Figure 28. Sequenced samples characteristics. **(A)** Stacked columns describing the cancer samples that were sequenced along with a matched-normal tissue against those that only haemolymph was sequenced (paired vs unpaired), the samples that were sequenced using a whole-genome amplification (WGA) protocol prior to library preparation against those prepared with native isolated DNA (WGA vs DNA) and samples were sequenced at high coverage against those at low coverage. **(B)** Pie chart displaying the proportion of tissues sequenced as matched-normal (foot, adductor muscle and mantle). **(C)** Schematic representation of cockle tissues with arrows indicating the a histological estimation of the number of cancer cells usually found in that tissue in late stages of HN. **(D)** Tree maps outlining the healthy cockle samples sequenced by means of sequenced tissue (yellow), nucleic acid used for library preparation (red), sequencing depth (green) and year of sample collection (purple).

In addition, non-cancer individuals were sequenced to create a panel of normal individuals (PoN). For the most part of the PoN, native DNA isolated from foot tissue was sequenced at a medium sequencing depth of samples collected from 2016 to 2020 (Figure 28D).

2.2.5. SPECIES DETERMINATION

Species determination was performed by species-specific PCR amplification of their ribosomal DNA ITS region (Freire *et al.*, 2011).

Table 5. Species-specific primers to differentiate two cockle species co-habiting in some regions (*Cerastoderma edule* – Ce and *Cerastoderma glaucum* – Cg) in a single PCR amplification (Freire *et al.*, 2011).

Forward primer	Primer sequence (5' → 3')	Reverse primer	Primer sequence (5' → 3')
ITS-forward	GTTTCCGTAGGTGAACCTG	ITSCe-R	AAGCAGCGAGAAGCCGTTC
		ITSCg-R	AATTCGCCATCGTCGG

Amplifications were performed in a final volume of 25 µl; the reaction mixture contained 20 ng/ µl of genomic template DNA, 1 mol/L of each primer, 2.5 µl of dNTPs at 2 µM, 0.5 µl of Taq polymerase (Sigma Aldrich, 5 uds/µl) and 2.5 µl of the polymerase buffer. Denaturation, annealing, extension and number of cycles were used as specified in Freire *et al.*, 2011. PCR products were checked on 2% agarose gels stained with SYBR-Safe and photographed in an LM-20 transilluminator. Species determination was performed by looking at the gel bands of different sizes depending on the species. Two positive controls with samples of *Cerastoderma edule* and *Cerastoderma glaucum* were included.

2.2.6. COCKLE REFERENCE GENOME

To obtain the reference genome of common cockle *Cerastoderma edule,* a first initial sequencing of a male cockle with no pathologies or chromosomal abnormalities was carried out. DNA isolation was made from the foot with a commercial kit of DNA extraction from human blood; quality controls were good enough to proceed with library preparation and posterior sequencing. After this pilot experiment, we proceeded with the sequencing of a large male cockle from Spain that was used to build the reference genome. Assembling and annotation was performed by Jorge Zamora and Daniel García-Souto (unpublished data), then I characterized the initial results (exon distribution, repeats distribution, mtDNA) before starting to use it as our standard reference genome.

2.2.7. WGS DATA ALIGNMENT

The dataset comprising paired-end sequencing samples was aligned to the common cockle reference genome using BWA-mem 0.7.17-r1188 (Li, 2013) with default settings and samtools v.1.9 (Li *et al.*, 2009) was used to sort and index the files. Duplicate reads were marked using the package biobambam/bammarkduplicates (Tischler and Leonard, 2014).

2.2.8. MITOCHONDRIAL ANALYSIS

2.2.8.1. Visual inspection of alignments

By visual inspection using the Integrative Genomics Viewer (Robinson *et al.*, 2011) all tumours and healthy cockles were checked. Unexpected regions with higher coverage in the coordinates MT:9018-10168 were detected and annotated for several tumour samples.

2.2.8.2. Variant calling and filtering

Variant calling was made for the mitochondrial genomes individually using GATK Mutect2 v4.1.6.0 (Poplin *et al.*, 2018; Van der Auwera and O'Connor, 2020) setting the flag "mitochondria-mode", which automatically sets parameters for variant calling in mitogenomes. A maximum number of 100 reads were retained per alignment start position and the filtering of duplicates disabled as advised. Sites with median mapping quality below 50 were skipped and calling of MNPs disabled (each variant was called independently for each alignment position). Because of the specific GC bias of the mitochondria, an orientation bias model was built and used to filter mtDNA calls. A median autosomal coverage of 50, estimated with samtools 1.9 (Li *et al.*, 2009), was assumed for filtering potential polymorphic NUMTs (nuclear integrated mtDNA copies). The minimum alt reads required on both forward and reverse strands for calling a variant was set to 1. Variants were not called in coordinates MT:9018-10168 (portion of ~1Kb) as tandem amplifications were detected in that region (see *Section 2.2.7*); that portion was posteriorly removed from the alignment for downstream analyses.

Variant allele frequency (VAF) plots were built for each sample and visually examined. In general, the identification of different mtDNA haplotypes was straightforward (tumour, matched-normal and their shared variants), which are present at different frequencies within a sample, but this first visualization allowed us also to identify the overall presence of many likely false positives as well (*e.g*., variants not called in both samples of a pair whose frequencies differ from the main variant sets of that sample), as well that samples subject to WGA that had usually many low frequency variants especially many indels, that were clearly false positives. All variants from all samples were plotted in an occupancy matrix for better visualization and decision-making and performed a second round of filtering as follows.

First, we performed variant-type based filtering:

(1) *Biallelic indels* (61) were discarded for simplicity, as they were almost exclusively found in samples subject to WGA, and almost exclusively at very low frequencies, with strong evidence thus of the huge majority being amplification artifacts.

(2) *Multiallelic positions* were individually examined across all samples and labelled as to keep or to exclude based on the concordance between their frequency and the ones of the genomes of the respective samples (155 kept from 197 called). Indels within these (101 called) were discarded for concordance with previous. Most of these positions had also clear evidence of being false positives (low frequency across all samples), mostly (though not only) related to WGA samples.Yet, they were all examined individually, and the 12 that are likely true (existence in more than 1 sample and concordance with frequency of sample genome(s)) were annotated for posterior examination. They were mostly on tRNAs and 12S and 16S rRNAs across different samples.

For the remaining biallelic SNPs called (1666), filtering was sample type based, as follows:

(1) *Healthy individuals* (*i.e.,* without tumour diagnosis, herein N0s), for which all variants were usually at frequency ~1, variants found in samples at freq > 0.5 (but less than 1) had their frequency converted to 1, and the ones with freq < 0.5, it was converted to zero. The rationale here is that the first case (1 > freq > 0.5) may be explained by mapping errors (of other reads, which decrease the variant frequency in that position), coverage issues, or even unidentified copy number variations or high frequency heteroplasmies, but that by considering those we are considering the "majority" genome of that sample; and in the second case (0.5 > freq > 0) we are getting rid of false positives and low frequency heteroplasmic positions.

(2) *Paired tumour and normal samples* (*i.e.*, those for which there were tumour and a matched-normal tissue from the same individual); we compared both corresponding genomes in both samples.

(3) *Only tumour samples* (*i.e.,* individuals with cancer diagnosis but for which there was only a sample from one tissue - generally haemolymph), we examined all VAF plots in detail and when possible, established a frequency threshold below which variants were eliminated.

2.2.8.3. Clonal deconvolution

Clonal deconvolution algorithms commonly applied to cancer bulk data to identify and separate different clonal lineages such as Clomial (Zare *et al.*, 2014), LICHeE (Popic *et al.*, 2015) and CloneFinder (Miura *et al.*, 2018) were tested in our dataset with all samples together to separate cancer and host genomes across samples, but with no satisfactory results. A high number of clones with nonsense frequencies were inferred for each sample in the different methods, as well as many variants were not assigned to any clone, rendering its application impossible to this dataset; furthermore, results differed across the methods. This performance

is probably related to ploidy (mtDNA genomes are haploid and these methods are built to infer clones in diploid genomes), to the high number of samples/genomes and their divergence (the co-existence of distant clones in the same sample due to the nature of cancer transmission), given that these methods expect that clones/genomes within a given sample are related to each other by a (preferably low) number of mutations (Miura *et al.*, 2018, 2020).

Thus, clonal deconvolution to separate cancer and host genomes across samples was performed "manually" by directly inspecting VAF ranges together with the ratio of tumour cells in each sample -that is cell counting previously described in *Section 2.2.*2- to separate and label tumours and host haplotypes within each sample. In cases where "host" and "tumour" genomes IDs could not be confidently attributed (mostly unpaired samples with no correspondence between tumour cell counting and genome frequencies), these samples were discarded. There were a few cases where more than two genomes were present in both sequenced tissues (tumour and matched-normal). All were extracted and included in the analyses. For samples within which more than two genomes seemed to co-exist but 1) the third genome did not appear in both tissues of the individual (for paired samples), and/or 2) the frequency at which it appeared was very low and/or originated a long branch in the phylogeny suggesting possibly false variants and not related to any of other tumoral lineages, these were also discarded in a conservative approach.

In-house R scripts were used to reconstruct these mtDNA genome sequences, from a multisample VCF file and the mtDNA reference sequence, including a list of the filtered/unfiltered variants and filtering thresholds per sample, when appropriate.

2.2.8.4. Phylogenetic inference

MtDNA genomes alignment was visually inspected using GeneiousPrime v.11.03 (www.geneious.com) to check the correctness of reading frames across coding genes and basic alignments statistics. As indels were not called, resulting generated sequences had all the same length. Region MT:9018-10168 was excluded due to the existence of high unexpected coverage in cancer genomes (see *Section 2.2.8.1*).

As average divergence was very low (~1%), "preliminary" NJ trees were constructed and used to examine the placement of some sequences we were not completely confident about (cases of host/tumour unknown cell counting and samples for which we suspected that false positive variants could exist). If they turned out to be long branches and/or not grouping with known host/tumour lineages as expected, they were (conservatively) discarded.

ModelTest-NG (Darriba *et al.*, 2020) was used to select the best-fit nucleotide substitution model for the dataset. Models were estimated for each gene/region separately (30 regions; some overlapping and/or contiguous tRNAs/intergenic regions were merged), as well as a single model for the complete dataset and models for a three-partitioned dataset (coding-regions, rRNAs and tRNAs), and chosen according to Bayesian information criteria.

Phylogenetic relationships were inferred using maximum likelihood (ML) and Bayesian inference (BI). For ML, we used RAxML-NG v0.8.1 (Kozlov *et al.*, 2019) with 10 parsimony starting trees and 1000 bootstrap replicates. Partitioned analyses were implemented, using the 30 partitions described above (exploratory analyses made using a single and three partitions gave identical results). BI analyses were conducted with MrBayes v3.2.7 (Huelsenbeck and Ronquist, 2001), again implementing different models for the 30 a priori established partitions. Branch lengths were linked, and four simultaneous Markov chains were run, for 15 million generations, sampling every 1500. At least two runs were made. Congruence of runs and

convergence of both parameters and topologies was accessed with RWTY (Warren *et al.*, 2017), a 50% majority rule consensus tree was built and 10% of the run was discarded as burn-in.

All trees were inferred without an outgroup and are midpoint rooted for presentation. We attempted several approaches to rooting, but none was successful. Dataset was aligned to *C. glaucum* mitogenome assembled by D. Garcia-Souto (unpublished), its closest known species, but it was roughly ~20% divergent, and different trials of phylogenetic inference using it (nucleotides and coding regions only, 1st and 2nd positions only or aminoacids), all resulted in similar trees with a very long branch leading to the outgroup and completely unresolved relationships within the ingroup. Computational approaches to find a root within the ingroup were also tried (Bettisworth and Stamatakis, 2021), using the ML inferred tree, but support for the inferred root (an Irish healthy cockle) was very low (0.09) and thus the result considered unreliable.

2.2.8.5. Divergence time estimation

To infer the timing of diversification of this species and the ages of origins of the tumour lineages, we estimated a time tree using BEAST2 v2.6.2 (Bouckaert *et al.*, 2019). Estimates were done both using a fixed tree (the ML tree, midpoint rooted) and co-estimating the tree. We used the uncorrelated relaxed clock model (Drummond *et al.*, 2006) with a normally distributed prior on the substitution rate with mean 0.01 and standard deviation of 0.003 substitutions per million year and min/max values of 0-0.5. This was estimated as an average overall rate for invertebrates (Allio *et al.*, 2017), and we believe it to be appropriate. Runs were implemented with a single or 3 partitions (coding regions, rRNAs and tRNAs), and not more, to reduce bias on node ages caused by increased partitioning (Jin and Brown, 2018), with linked clock models and tree topology, and both coalescent and Yule priors used on the tree topology. Several independent MCMC chains were run for 200M generations, sampling every 20,000. Convergence was checked using Tracer v.17.1 (Rambaut *et al.*, 2018) and TreeAnnotator used to calculate a consensus tree (MCC) and summarize the posterior estimates.

2.2.8.6. Tree topology tests

Multiple tests were used to compare alternative hypotheses to the obtained unconstrained phylogeny and therefore discard a common origin of the cancer mtDNA lineages described in this work. Support for alternative topologies was evaluated using Shimodaira-Hasegawa (SH) and the approximately unbiased (AU) tests as implemented in iqtree2 (Minh *et al.*, 2020) as well as through (bayesian) posterior probabilities odds as in Bergsten et al. (2013). The posterior probability for each hypothesis was calculated by filtering the posterior (post-burnin) tree sample in PAUP* v4.0a168 (Swofford, 2002).

2.2.8.7. Estimation of selection

Coding regions were used to test hypotheses of relaxation or intensification of natural selection along (tumour vs normal) branches of the mtDNA tree, using RELAX (Wertheim *et al.*, 2015).

2.2.9. MICROSATELLITES

2.2.9.1. Identification of novel cockle microsatellites

Using the contigs of the initial draft of the assembled reference genome of common cockle, SciRoKo (Kofler, Schlötterer and Lelley, 2007) was run to identify microsatellites. Then, the following criteria were applied to select the most suitable ones: (1) microsatellite repeats were trinucleotides or tetranucleotides; (2) total length of the microsatellite was in a range between 80-140 nucleotides; and (3) among the 100 flanking bases there were no repetitive sequences.

Primers were designed using the web interface of the Primer 3 (Untergasser *et al.*, 2012), and an in-silico PCR was performed to study the possible existence of base complementarity of the primers in other regions of the genome removing them to facilitate the amplification of a single band in the PCR (or two, in the case of heterozygous individuals). Microsatellites were tested and a selection of 10 was used for the subsequent analysis presented on this thesis. More information about the initial tests can be found in Ruiz Arribas (2017).

2.2.9.2. Identification of novel cockle microsatellites

Twelve microsatellite markers developed by Martínez *et al.* (2009) were genotyped.

2.2.9.3. Amplifications and genotyping

PCR amplifications consisted of an initial denaturation step at 95 °C for 7 min, followed by 35 cycles at 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 45 s., a final phase of extension at 72 °C for 7 min and then kept at 4 °C. All PCRs were performed in a final volume of 25 µL containing 2 µL of DNA (10 ng/µL), 1 µL of forward and reverse primers at 10 µM (Sigma), 2.5 µL of dNTPs (Invitrogen) at 2 mM, 0.5 µL of Taq polymerase (Sigma) at 5 units/µL, 2.5 µL of 10X PCR buffer (Sigma) and 15.5 µL of water (PCR-grade water). PCR products were visualized by electrophoresis in a 2% agarose gel and 1X TBE. These gels were stained with SYBR-Safe and photographed in an LM-20 transilluminator.

Selected samples and microsatellites were genotyped in a SeqStudio Genetic Analyzer (Applied Biosystems by Thermo Fisher Scientific) and data was analysed using GeneiousPrime v.11.03 (www.geneious.com).

2.2.10. LONG-READ SEQUENCING

To characterize the higher coverage regions previously identified (see *Section 2.2.7.1*) that suggested mitochondrial copy number (CN) amplifications, two long-read sequencing strategies were used: (1) whole-genome sequencing on three tumoral samples as representatives of the three different mitochondrial CN amplifications identified and (2) amplicon sequencing of the region. After DNA purification with 0.4xAmpure XP Beads (Beckman Coulter Inc) and the repair and end preparation steps using NEBNext End Repair/dA-tailing module (NEB), the library was built using the Amplicons by Ligation (SQK-LSK109, Oxford Nanopore Technologies Ltd.). Libraries were loaded into R9.4 MinION sequencing cells (FLO-MIN106, Oxford Nanopore Technologies Ltd.) and sequencing readouts were controlled using Oxford Nanopore MinKNOW v18.01.6 software. Fastq files were generated using ON basecaller v2.0.1 and minimap2 (Li 2018) was used to map the sequencing reads against the mitochondrial reference genome when appropriate or against simulated genomes with tandem amplifications to investigate on the number of CN amplifications. SAM files were converted to BAM format and sorted and indexed with Samtools v1.7 (Li et al. 2009).

2.3. RESULTS AND DISCUSSION

2.3.1. DISTRIBUTION AND PREVALENCE OF COCKLE TRANSMISSIBLE CANCERS

While common cockles are distributed from Morocco to Russia along all the Atlantic Coast of Europe (FAO, 2019), we observe prevalence disparity of HN across cockle populations, with areas where the disease reaches high prevalence rates, and others with no disease at all. The overall prevalence of the disease was 5.3% (356/6,719; Figure 29A) however, it has only been diagnosed in the Southern regions of the European Atlantic Coast.

We have sampled 36 locations belonging to 11 different countries and HN has only been found on 19 sampling points belonging to 5 mostly southern European countries bathed by the Atlantic Ocean (Figure 29B): Portugal, Spain, France, England, and Ireland. Our results show a mainly continuous distribution of HN in southern Europe with some sporadic locations where no HN was found (Plymouth UDCE, Arcachon FACE, Bilbao EBCE, Grove EGCE, Placeres ELCE) which not necessarily means that there is no HN because, in the case of HN in Arcachon (France) and Grove (EGCE) has already been described in the literature (Le Grand *et al.*, 2010; Carballal *et al.*, 2001). Moreover, sampling was carried at different times of the year and the prevalence of neoplasia has been correlated to the reproduction cycle of cockles and, therefore, with the time of year (Diaz *et al.,* 2016). Note that no HN has been found on northern countries or in Morocco whose coastline is facing the Portuguese area of Algarve where the highest prevalence of the disease has been found. According to the population structure of cockles based on microsatellites and mitochondrial genes, the genetic variation of this species is characterized by two homogeneous and differentiated groups – southwestern and northern – and a heterogeneous central group (Krakau *et al.*, 2012; Martínez *et al.*, 2015; Vera *et al.*, 2021) which may be a barrier for HN spread as we do not find HN in northern populations. Along with the patterns of gene flow of cockle's populations, the density and distance of those populations, oceanic currents, or the marine physical conditions such as temperature, salinity, pH, pressure, CO_2 (Grossmann and Klotzbach, 2009) may also explain the distribution of HN.

Historically, HN was diagnosed in France, Ireland, Netherlands and Spain (Twomey and Mulcahy, 1984; Poder and Auffret, 1986; Carballal *et al.*, 2001; Díaz *et al.* 2016) with a wide range of HN prevalence reported depending on time and location (Table 2). Here we report HN also in two Portuguese locations (Algarve and Aveiro) and in Wales, United Kingdom (both sampling sets collected on 2017) from where, as far as we know, HN was not previously known probably due to the lack of pathological studies of this species in that area. A recent report has confirmed its existence in Portugal but not in the United Kingdom (Montaudouin *et al.*, 2021).

Our HN prevalence results compared to other HN reports available in the literature (Table 2) show some disparities. We report low HN prevalence (2.3%, 9/384) in one French Atlantic location (Roscoff) sampled in 2017 that contrasts with recent studies of the disease reporting a 28% of prevalence in this area (Montaudouin *et al.*, 2021) which could be pointing to an outbreak of HN in this region. Regarding Ireland, in old reports HN prevalence ranged from 22 up to 94% while in our collection we found a HN prevalence of 7.3% (41/563) which could be pointing to an HN incidence decrease in this region or it could just be highly variable from year to year and within seasons. In Spain, where this disease has been studied for decades, we see that the prevalence corresponds to the results of other publications (Carballal et al., 2001, 2015; Villalba, Carballal and López, 2001; Da Silva et al., 2005; Romalde et al., 2007; Diaz et al., 2013; Ruiz et al., 2013; Diaz et al., 2016; Montaudouin et al., 2021).

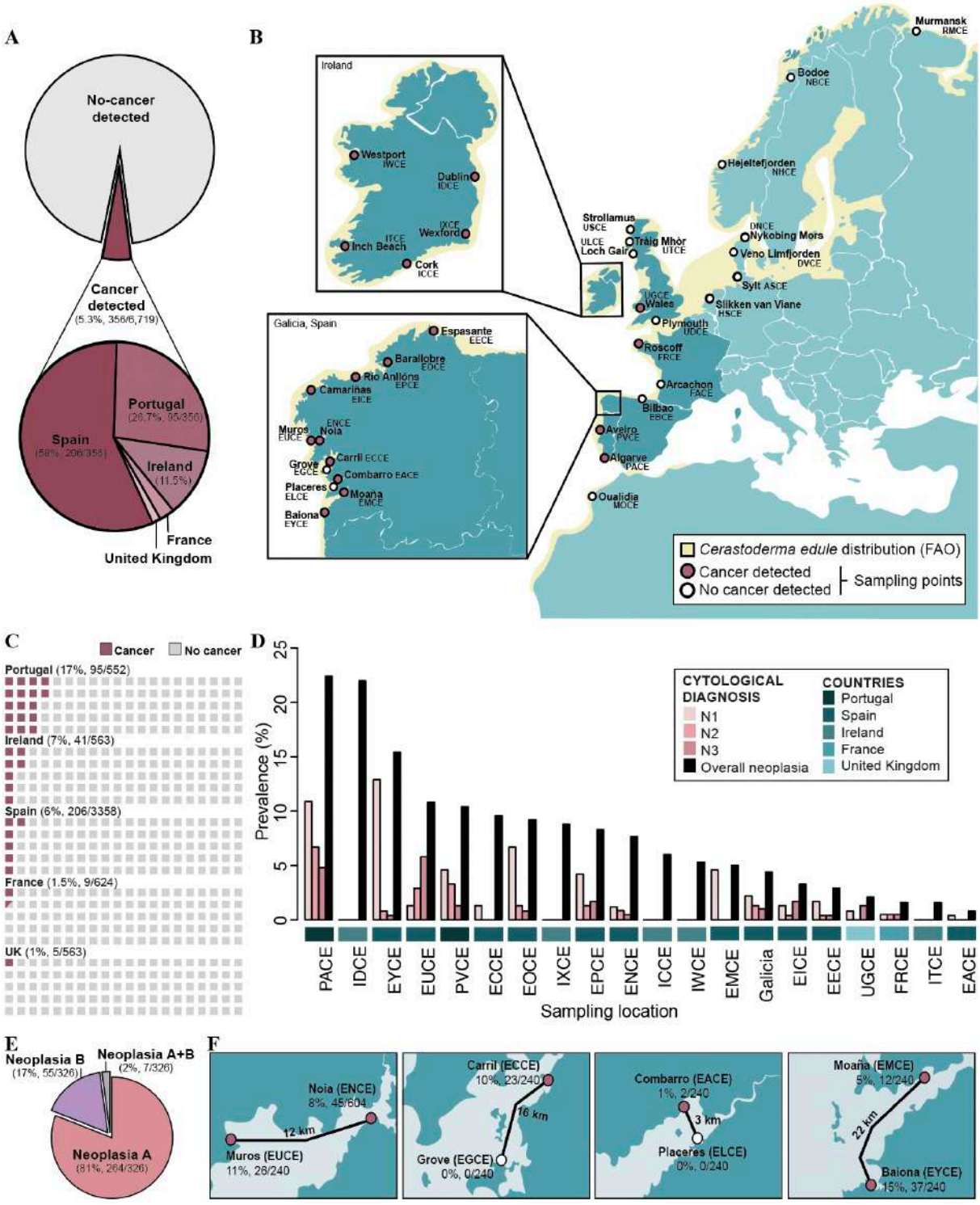


Figure 29. Distribution and prevalence of HN. **(A)** Pie charts illustrating the abundance of cancer detected in the cockle collection used for this study and the proportion of cancer samples by countries in our collection. **(B)** Geographical map showing in light yellow the distribution range of this species (FAO) and with points the sampling locations screened displaying an empty point when cancer was not detected and a filled point in locations where at least one cockle was diagnosed with HN; Ireland and Galicia (Northwest of Spain) are shown in zoomed maps due to the intensive sampling performed in those two regions. **(C)** Waffle plots featuring the prevalence of HN by country (*ie*. number of HN-diagnosed cockles found on that country divided by the total number of samples screened in that country). **(D)** Bar plot displaying in black the overall prevalence of each sampling location (*ie*. number of HN-diagnosed cockles found on that location divided by the total number of samples screened in that sampling location), for the cases where a severity stage was assigned (N1, N2, N3), the overall prevalence is

broken down; data from the 12 sampling points from Galicia are also shown merged; n can be found in Appendix A: Supplementary material. **(E)** Pie chart displaying the classification on type A and type B of neoplastic samples. **(F)** Zoomed maps of regions from Galicia (Spain) where two close sampling collections were assessed, distances between points were calculated on distance.to (accessed on December 26th, 2021) in kilometres (1 nautical mile = 1,852 kilometres), local prevalence is displayed for each sampling point; sole intention of showing nearby sampling locations with different prevalence, marine currents have not been taken into account.

Reports of HN in cockles from the Netherlands were published in 2021 by Montaudouin *et al.,* nevertheless, we screened 144 cockles from Slikken Van Vianne in 2017 (Netherlands, Appendix A: Supplementary material - sampling summary table) and no HN was found which could be due to (1) sampling on a year of low prevalence with a small sample size that did not allow to detect HN, or (2) the dynamics of this infectious disease which are heavily dependent on the rate of transmission from infectious to susceptible hosts (Real and Biek, 2007).

The highest HN prevalence was found in Portugal (17%, 95/552), followed by Ireland and Spain while the lowest prevalence was found in France and the United Kingdom (Figure 29C). When breaking down the data by sampling points (Figure 29D, black bars), three locations show a HN prevalence greater than 15% in Portugal (Algarve, PACE), Ireland (Dublin, IDCE) and Spain (Baiona, EYCE). A recent study trying to understand the impact of global warming on marine bivalves brought to light that in temperature stress conditions, circulating haemocytes leave the haemolymph to gain access to the intervalvar fluid before being released in seawater (Caza *et al.*, 2020) which might explain why we found more HN prevalence in southern regions as cancer cells will more often be released to the seawater causing contagion.

The sampling summary table included in the Appendix A: Supplementary material shows precises timings of sampling collection because temporal cycles have been previously described to affect HN prevalence in Spanish cockle populations, suggesting that the drops in HN prevalence could be due to the death of the severe diseased individuals evaluated in the previous month (Díaz, 2015). In addition, HN causes inhibition of gametogenesis in cockles, which could result in a decrease in population size (Díaz *et al*., 2016).

Table 6. Cytological severity of the HN diagnosis in cockles. Cell counting of 500 cells per sample.

Diagnosis stage	**Number of samples**	**Proportion of cases on this stage**	**Definition of the stage** (Cooper et al., 1982a; Farley et al., 1986; Barber, 1990; Brousseau and Baglivo, 1991)
N3	40	15%	**Severe stage** because the proportion of cancer cells was higher than 75% in the haemolymph.
N2	59	22%	**Medium stage** when the proportion of cancer cells ranged from 15% to 75% in the haemolymph.
N1	156	58%	**Early stage** if the proportion of cancer cells was lower than 15% in the haemolymph
NA	16	6%	*Not applicable* if the haemolymph cell monolayer was not good enough to perform cell counting.
Total	271*		

* Not all HN cockles collected were diagnosed through haemocytology due to logistic restrictions.

In terms of HN severity stage, we never found severe (N3) or medium (N2) stages if there were not early (N1) stages of cancer in a given sampling location, however, sometimes the three stages were found (*eg*. Algarve in Portugal, PACE; Roscoff in France, FRCE) or at times only the early (N1) stage (*eg*. Moaña in Spain, EMCE) was found (Figure 29D). This could be due to (1) cockle's immune system delaying cancer progression or (2) the rapid death of individuals in advanced stages in their habitat or during sampling or (3) the month of sample collection (Díaz *et al*., 2016). In general, 58% of all the cancer samples collected for this study were categorized as early stage (N1) and only 15% were in a severe stage (N3) which are the better samples for high throughput sequencing as more than 75% of haemolymph cells would be

cancerous (Table 6). Unfortunately, cytological diagnosis of HN was not always performed due to logistic difficulties (*ie*. Irish samplings).

In terms of HN type (Figure 29E), the majority of cancer samples were classified as type A (81%, 264/326) and they belonged to France, Ireland, Spain and Portugal. As type B only 17% (55/326) of cancer samples were classified as type B and they belonged to United Kingdom, Ireland, Spain and Portugal. In France and United Kingdom only one type of neoplasia was found in our samplings. In addition, seven cancer samples (2%, 7/326) were classified as type A and B, further investigations on these samples can be read in *Section 2.3.4* of this doctoral thesis.

In the region of Galicia (Spain), an intensive sampling was carried out given the geographical ease to access these samples. Close located points give us an idea of how the prevalence varies across short distances in the marine environment (3-22 km). For instance, prevalence rises from 0% in O Grove to 9.6% in Carril being just 16 km apart while in Muros (11%) and Noia (8%) that are 12 km apart it is a quite stable value (Figure 29E). Therefore, these prevalence values have been taken at a particular timepoint and vary widely even in short distances, so estimates should not be assumed to be constant, and the incidence rate of the disease should be studied to get a better idea of the epidemiology of HN. In addition, other parameters such as sea currents or changes in salinity due to the arrival of large amounts of water from rivers at certain times may also play a role. Moreover, we should consider that countries where fewer localities were sampled could have HN prevalence underestimated.

A sampling size of 240 cockles per location was estimated to be capable of detecting pathological relevant differences without compromising the ethical acceptability of sampled populations; based on an overall prevalence of 2% used to compute the number of samples desired for the genomic analysis of HN. In addition, the proximity of certain regions (Galicia, Spain) made possible the access to more locations so, our HN collection is overcomposed by Spanish cockles that represent the 58% of the samples (Figure 29A).

2.3.2. COCKLE SPECIES OF TRANSMISSIBLE CANCERS

Although common cockle is the most abundant species across the sampled area, it overlaps in several localities of the Atlantic coast with the lagoon cockle, being often difficult to distinguish both species morphologically. Thus, for all the samples from individuals where cancer cells were found, and for several healthy individuals, a species determination was performed by species-specific PCR amplification of their ribosomal DNA ITS region as described in Freire *et al.*, 2011. Results showed that all the cancer samples used for subsequent sequencing were common cockles (Figure 30).

Coexistence of common and lagoon cockles was found in several Spanish sampling locations (Carril, ECCE; Placeres, EPCE; Combarro, EACE and Espasante, EECE) although all individuals diagnosed with HN tested positive for common cockles and negative for lagoon cockles (Appendix A: Supplementary material). However, the latter species has also been reported to be affected by HN (Rodriguez *et al.*, 1997; Carballal *et al.* 2016) but whether it is a new transmissible cancer lineage, an interspecies transmission of common cockle cancer lineages or a non-contagious HN remains unknown.

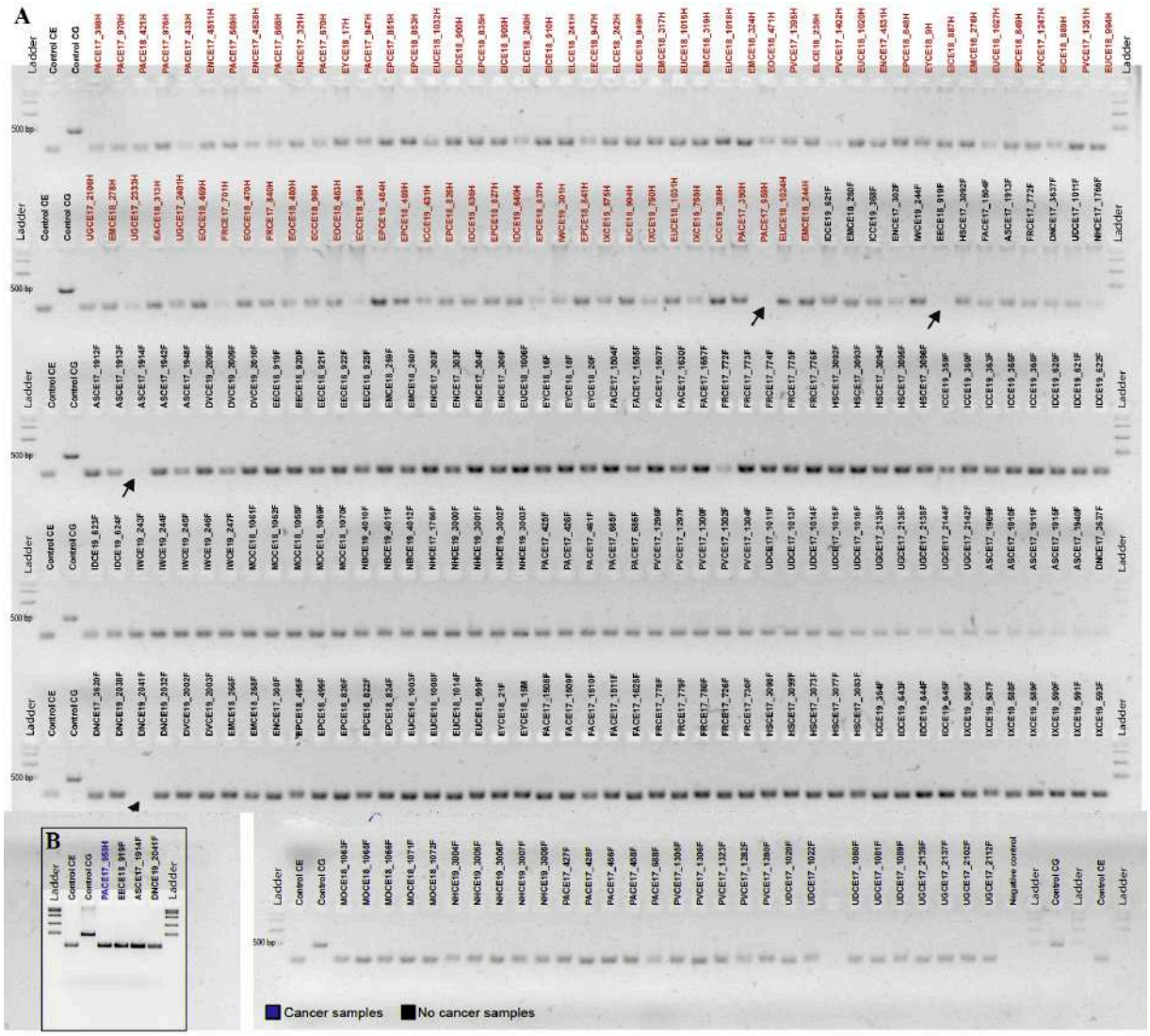


Figure 30. Species determination of 259 samples. **(A)** All cancer (red) and non-cancer (black) samples amplified the 185bp bank of *C. edule* (CE) while not a single sample was determined to be *C. glaucum* (CG) as they do not amplify a band at 470bp. Some samples (arrows) did not amplify any band. **(B)** Repetition of species-specific PCR for the four samples (arrows) that showed no band in the previous gel, all of them resulted to be *C. edule*.

2.3.3. DEVISING THE SEQUENCING DATASET TO STUDY THE EVOLUTION OF COCKLE TRANSMISSIBLE CANCERS

2.3.3.1. Cockle reference genome

Scuba Cancers project, within which this doctoral thesis was framed, aimed to obtain the reference genome of common cockle *Cerastoderma edule* in which I actively participated in several steps. A first initial sequencing of a male cockle with no pathologies or chromosomal abnormalities allowed us to (1) get an overview of cockle genome and (2) design in-house microsatellites previously described in *Section 2.2.8.* The analysis of that sequencing data characterized the repetitive content and heterozygosity degree of cockle genome, as well as estimated its size around 1.5 Gb, approximately half size of a human genome. This size value differed by more than 160 Mb from that previously established by flow cytometry on this species (Rodriguez-Juiz, Torrado and Mendez, 1996).

Paired-end Illumina sequencing reads are randomly generated therefore, k-mer frequency follows a Poisson distribution, except a high proportion at low frequency due to sequencing errors (Zhang *et al.*, 2012; He *et al.*, 2016). In our case, a bimodal distribution was obtained, showing two peaks at 16 and 35 indicating that the genome of the sequenced cockle had a high level of heterozygosity (He *et al.*, 2016). Generally, parthenogenetic or inbreed lines of individuals are selected for genome sequencing to avoid genome heterozygosity (Ekblom and Wolf, 2014), but in bivalve molluscs this is not possible.

The analysis of k-mers usually characterizes the repetitive content of sequences which often hamper severely the assembly of most genomes (Williams *et al.*, 2013). To overcome these problems (Pendleton *et al.*, 2015) and obtain a good common cockle reference genome, a combination of different sequencing technologies was performed to ensure a high-quality assembly (Oxford nanopore long-reads, Illumina mate pair reads, Illumina paired-end short reads and Hi-C Illumina reads).

Genome size estimation was corrected with the definitive animal sequenced ending in a 0.8Gb genome (Figure 31A) which represents a third of the human genome and it is within the range of bivalve genome sizes. Common cockle has 38 chromosomes in both sexes, consisting of 19 pairs of chromosomes; a chromosome-level assembly (Figure 31B) of the genome was achieved for this reference genome by Jorge Zamora (unpublished data of Scuba Cancers project).

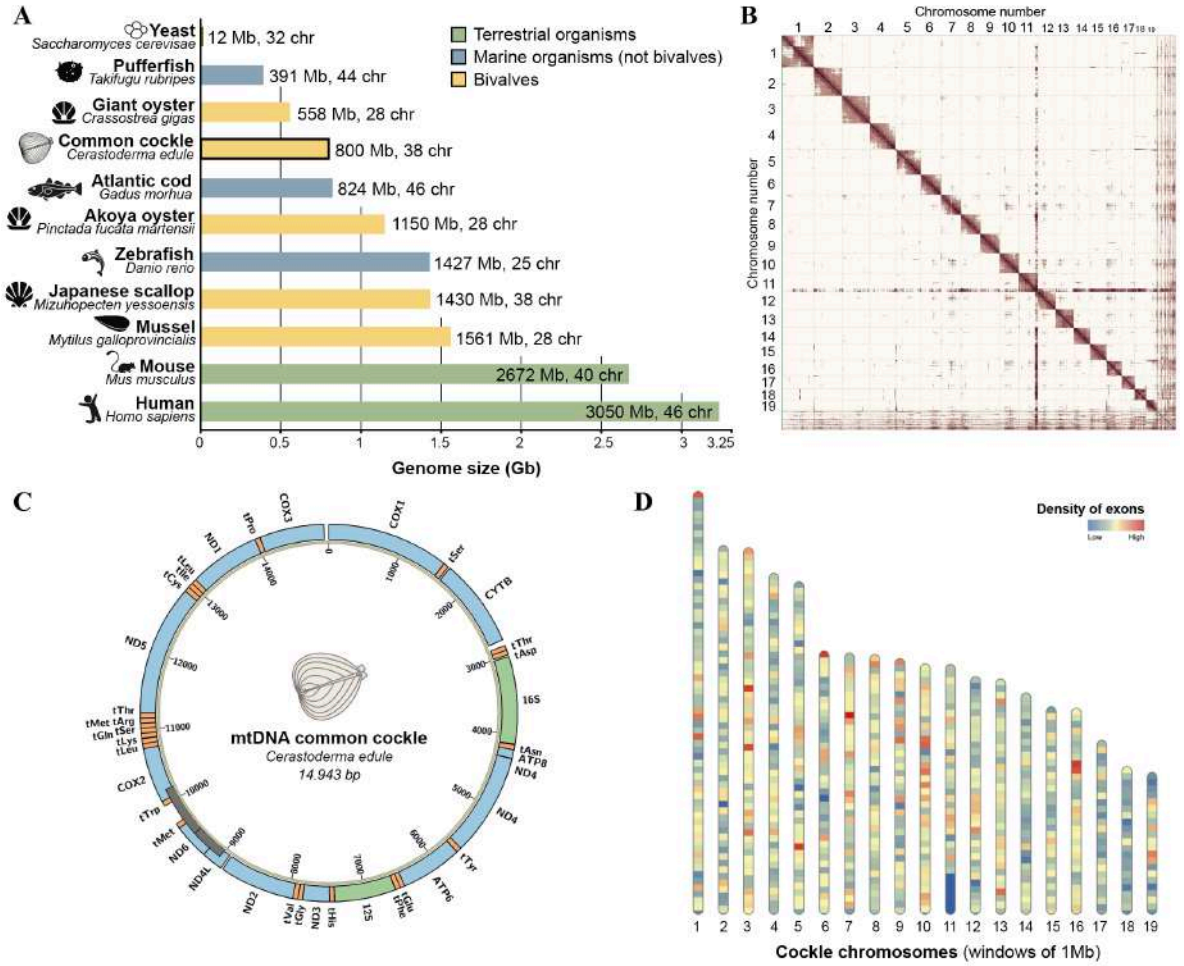


Figure 31. Cockle reference genome. **(A)** Genome size comparison of common cockle (0.8Gb) against terrestrial organisms (human, mouse, yeast) and marine organisms (several fish and bivalves), references can be found in

Figure 20 of *Section 1.3.4.3*. **(B)** Hi-C chromosome contact map where each block represents a Hi-C contact between two genomic loci, darker colour of a block indicates higher contact intensity (Courtesy of Jorge Zamora). **(C)** Annotation of mitochondrial chromosome of the reference genome. **(D)** Landscape of exome regions across each chromosome.

The hybrid assembly of the genome resulted on 19 chromosomal scaffolds, 1690 additional short scaffolds and the mitogenome with a N50 contig length of 1.28 Mb and a N50 scaffold length of 39.6 Mb. The largest chromosome length is 64.6 Mb, much smaller than the largest human chromosome which is 246Mb (Yunis, 1976). Mitochondrial genome was also recovered from the scaffolds by Jorge Zamora and annotated by Daniel García Souto (Figure 31C) showing high similarities with the published version (Quinteiro and Rey-Mendez, 2017).

Heterozygosity was estimated to be at 1.86 % (humans: 1%, Schneider *et al*., 2017) and G+C content of 35.6%, similar to that of the Tasmanian devil (36.4%, Murchison *et al*., 2012) but lower than that of domestic dogs (41%, Wang *et al*., 2021) or humans (45.2%, Schneider *et al*., 2017). At least 48% of the common cockle genome assembly is composed of transposable elements, a type of repetitive sequences. The prediction of coding genes and their functional annotation revealed 17,693 genes and an exome of 42Mb accounting for the 5% of the cockle genome (Figure 31D), bigger than the human exome that is about 30Mb constituting 1.1% of the human genome (Nurk *et al*., 2022).

2.3.3.2. Cockle transmissible cancers

To elucidate the evolutionary history of cockle transmissible cancers, 20% (n=70) of our tumour collection (Figure 32A) was sequenced including at least one sample of each population where cancer was diagnosed. This meant the creation of a sequenced tumours dataset in which more than half of the samples belong to Spain, followed by Portugal and Ireland (Figure 32B). In fact, the regions with low prevalence corresponded to the least represented in the sequencing dataset (*i.e*., French and English tumours have only 3 samples of each one) and samples of early (N1) stages of cancer were included for some sampling points of England, Spain and Portugal (Figure 32C).

Isolation of pure and high-quality DNA in sufficient amounts was challenging probably because most of the protocols were developed for vertebrates and do not perform well in mollusc tissues due to the content of mucopolysaccharides that tend to co-purify with DNA (Adema, 2021).

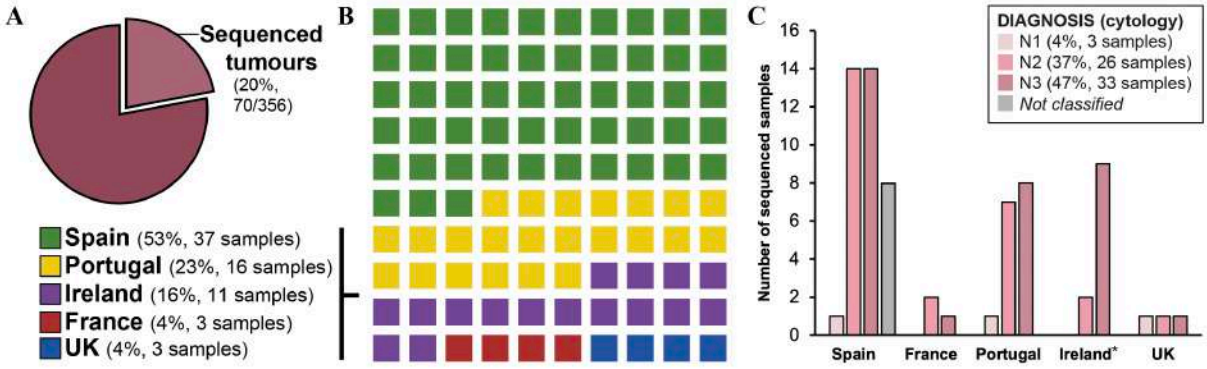


Figure 32. Sequencing dataset of cockle transmissible cancers. **(A)** Pie chart illustrating the proportion of tumours sequenced out of the total number of tumours collected. **(B)** Waffle plot featuring the representation of tumour locations within the sequenced dataset. **(C)** Bar chart broken down by country and by diagnosis stage using the percentages of 500 cell counted.

To compare tumour genomes to the host genome, for 63% (44/70) of the tumours, a matched-normal tissue was sequenced. By definition, a matched-normal is a sample of healthy tissue of the same individual, however HN is a cancer affecting the haemolymph, that is the circulatory fluid of this animals, which is bathing all tissues, making it difficult to select a healthy tissue with no cancer cells. Among the tissues showing less infiltration of cancer cells according to histological inspections, foot, adductor muscle and mantle were the tissues selected.

To estimate the evolutionary history of the cancer and, hereafter be able to filter as much germline variation as possible to study the somatic variation of the tumours, the genetic background of the host species was needed. Therefore, we built a panel of 481 normal individuals (PoN) from 34 different locations (Figure 33A) covering the distribution range of the host species, representing 7% (481/6,719) of the collected individuals (Figure 33B). Populations where no cancer was diagnosed were also included because the origins of these cancers are unknown, and these populations could be key to unveil their origins. Unfortunately, some populations are underrepresented in the PoN due to technical difficulties resulting in an overrepresentation of southern cockles. These 481 healthy cockles added to the 70 tumour individuals and its 44 matched-normal tissues make a total dataset of 595 samples sequenced (Figure 33C) that, to our knowledge, it is the most complete genomic dataset of marine contagious cancers produced to date.

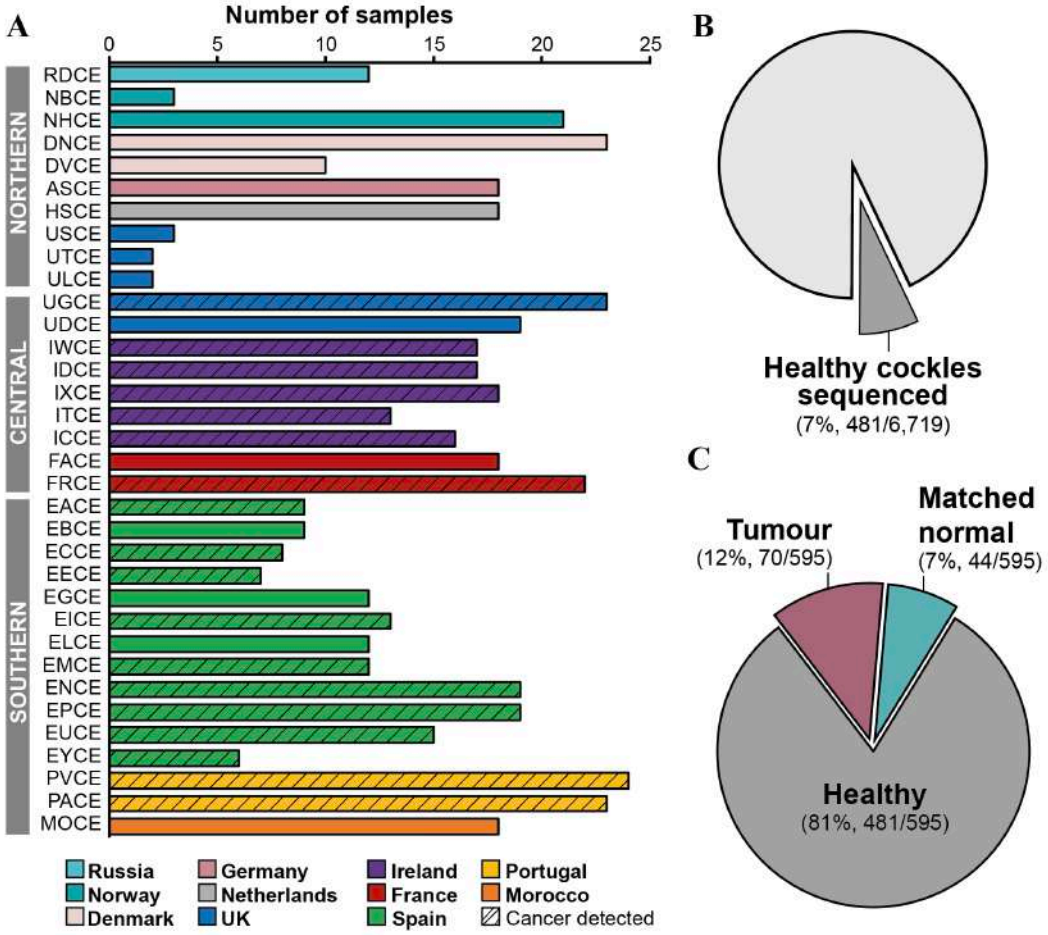


Figure 33. Sequencing dataset of healthy cockles (PoN). **(A)** Number of samples sequenced per location (both cancer and healthy cockles); location codes are given in Supplementary material. **(B)** Pie chart indicating the proportion of healthy cockles sequenced out of the total number of cockles collected. **(C)** Proportions of sample types out of all sequenced samples.

2.3.4. EVOLUTION OF COCKLE TRANSMISSIBLE CANCERS

2.3.4.1. Mitochondrial evolutionary history: multiple lineages

Throughout the evolution of Metazoa, gene content of mitochondria-genomes is highly conserved, as is the close packing of genes in contrast to nuclear chromosomes that have regions with no known genes. Animal cells carry tens to thousands of copies of the mitochondrial genome (mtDNA), an autonomously replicating circular chromosome encoding genes essential for oxidative energy metabolism (Wolstenholme, 1992). In general, mitochondrial DNA is normally inherited from the mother although doubly uniparental inheritance (DUI) is a major exception found in many bivalve species. Nonetheless, the occurrence of DUI has not been reported in common cockles despite being studied (Lucentini *et al.*, 2020) and our analysis of 481 healthy individuals support this evidence.

The coverage analysis of this haploid chromosome showed higher coverage in tumours than in healthy cockles (Figure 34A) but when grouping by type of nucleic acid sequenced, samples undergone a WGA protocol prior to sequencing are the main cause of this highest coverage (Figure 34B). Few healthy cockles (16/465, Figure 33B) were sequenced with WGA but half of the tumours (44/70, Figure 32D) did need WGA. WGA protocols were not needed in our sampling dataset to obtain good coverage of the mitochondrial chromosome due to the large number of copies present in a cell, but the availability of large amounts of nuclear DNA was of critical importance for this study. In previous studies inspecting cancer cells in electronic micrographs, high number of mitochondria were seen compared to haemocytes (Díaz *et al.*, 2011) which also helps to explain the differences of coverage between tumour and healthy samples (Figure 34A).

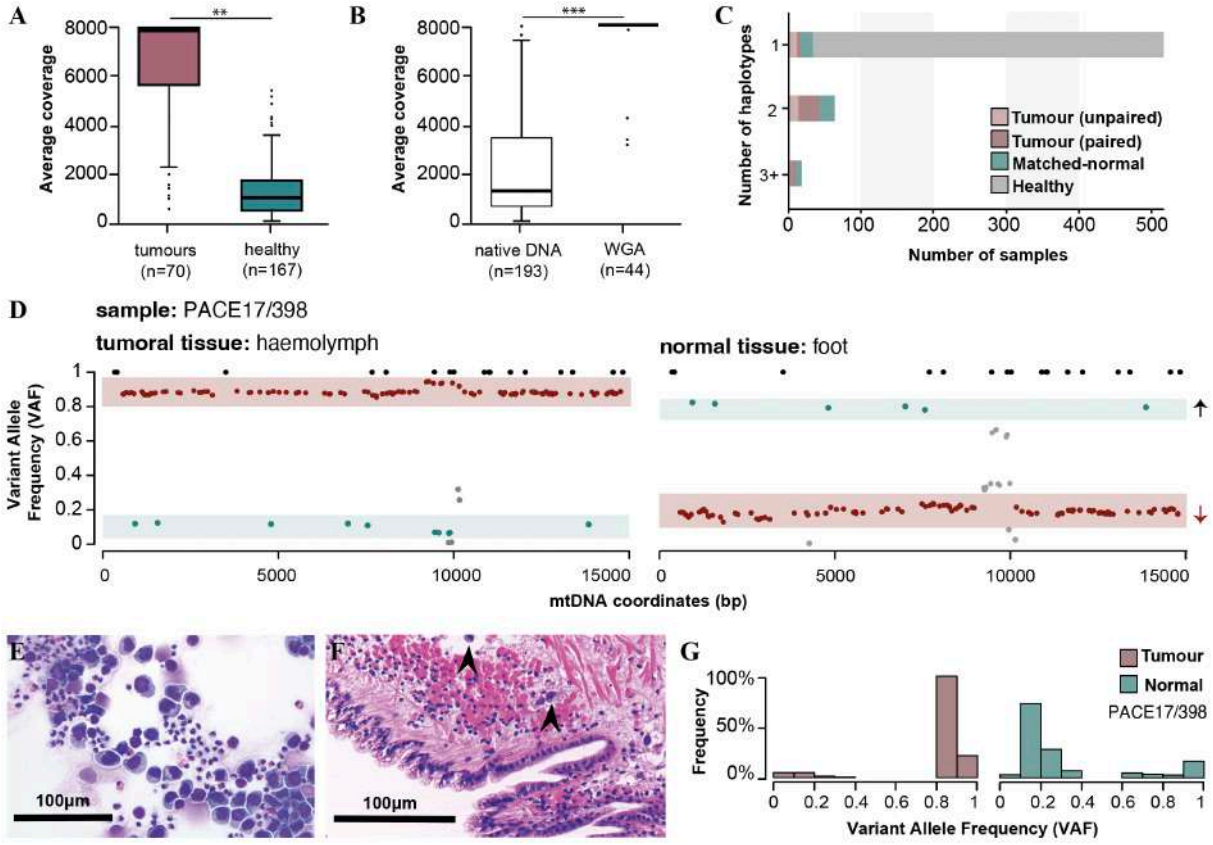


Figure 34. Mitogenome alignment, variants and deconvolution. **(A)** Average coverage on the mitochondrial genome of tumour samples and healthy cockles. **(B)** Average coverage on the mitochondrial genome of samples in which libraries were prepared using native isolated DNA and samples that undergone a whole-genome

amplification (WGA) protocol prior to library preparation. **(C)** Number of samples showing one, two, three or more haplotypes in the VAF plots broken down by sample type. **(D)** Variant allele frequency of variants across the mitochondrial genome sequenced from the tumoral tissue (left) and a normal tissue (right) of sample PACE17/398. Haplotype of tumour cells is highlighted in red while the haplotype of host cells in blue; fixed variants, that is variants present in both haplotypes, are coloured in black and grey variants were excluded of the interpretation for being in an area known to have structural variants. Arrows represent the movement of haplotypes from tumour tissue (haemolymph) to normal tissue (foot) featuring the decrease of the tumoral haplotype in the normal tissue. **(E)** Haemocytological preparation of the haemolymph, cell counting performed on this sample showed 89% of cancer cells. **(F)** Histological section of the foot showing low infiltration of cancer cells (arrowheads). **(G)** Histogram of the variant allele frequencies for that same sample.

To unravel whether several mitogenomes were present in cockles diagnosed with cancer (host and tumour mitogenomes), we analysed the allele frequency of all variants called in every sequenced sample (Figure 34C-D). All healthy cockles showed only one haplotype (Figure 34C, grey, VAF ~ 1) while only 24% (17/70) of tumour samples showed a unique haplotype. The most common case in tumoral samples was to find two haplotypes (63%, 44/70) or sometimes even more (13%, 9/70). In good correspondence with the haplotypes found in the matched-normal tissues in which 45% of the cases showed two haplotypes and 39% only one haplotype (Figure 34C). To set up the method, we started checking paired samples where usually the haemolymph was sequenced as tumoral tissue and the less infiltrated tissue as normal. Cell counting of haemolymph preparations (Figure 34E) was used to determine which mitogenome (*i.e.,* tumour or normal) was expected at higher/lower relative frequency in the haemolymph. We usually observed two haplotypes in at least one of the paired samples; when two haplotypes were present in both tissues, the behaviour was often opposite: the higher haplotype in the haemolymph decreased and vice versa (Figure 34C). The amount of cells that could be seen in the haemocytology (Figure 34E) or the histology (Figure 34F) corresponded roughly with the VAF values of the haplotypes (Figure 34G).

Once we had the haplotypes deconvoluted, we inferred a phylogeny to see the relationships between all healthy and cancer genomes. Four phylogenies were built using different methods, from more simple (genetic distances) to more complex (Maximum Likelihood and Bayesian). All the trees showed nine monophyletic lineages of cancer mtDNA haplotypes interspread within non-cancer genomes (Figure 35 and Appendix A: Supplementary material – mitochondrial phylogenies of tumours). Thus, matched-normal haplotypes did not group with the tumour haplotype of that sample, instead, they are distributed along the phylogenies clustering with other healthy samples (Figure 35). Any cancer sample clustered with their matched-normal, therefore no cases of non-transmissible HN were found in our dataset.

Healthy cockles mtDNA genealogies confirmed the geographical patterns of genetic variation previously described in the literature (Krakau *et al.*, 2012; Martínez *et al.*, 2015; Vera *et al.*, 2021) with northern, central and southern groups (Appendix A: Supplementary material – mitochondrial phylogenies of healthy cockles).

All nine mitochondrial cancer lineages have at least two samples and the lineage with more samples was the *HT-Nabia* lineage with fifteen samples (Figure 36A). Some lineages had samples from a single population (*i.e., HT-Sequana, HT-Sucellus* or *HT-Lugus* lineages) while others have up to eight different populations (Figure 36A, bars). *HT-Nabia* lineage is widespread along the Atlantic coast of Spain and Portugal, *HT-Sulis* lineage has also been found in two countries (UK and Spain) while the rest of lineages were found in one country only (Figure 36B). Forty percent (6/15) of the populations where cancer was found had more than one mitochondrial cancer lineage (Figure 36C) being the *HT-Cissonius* lineage the most widespread one (five populations).

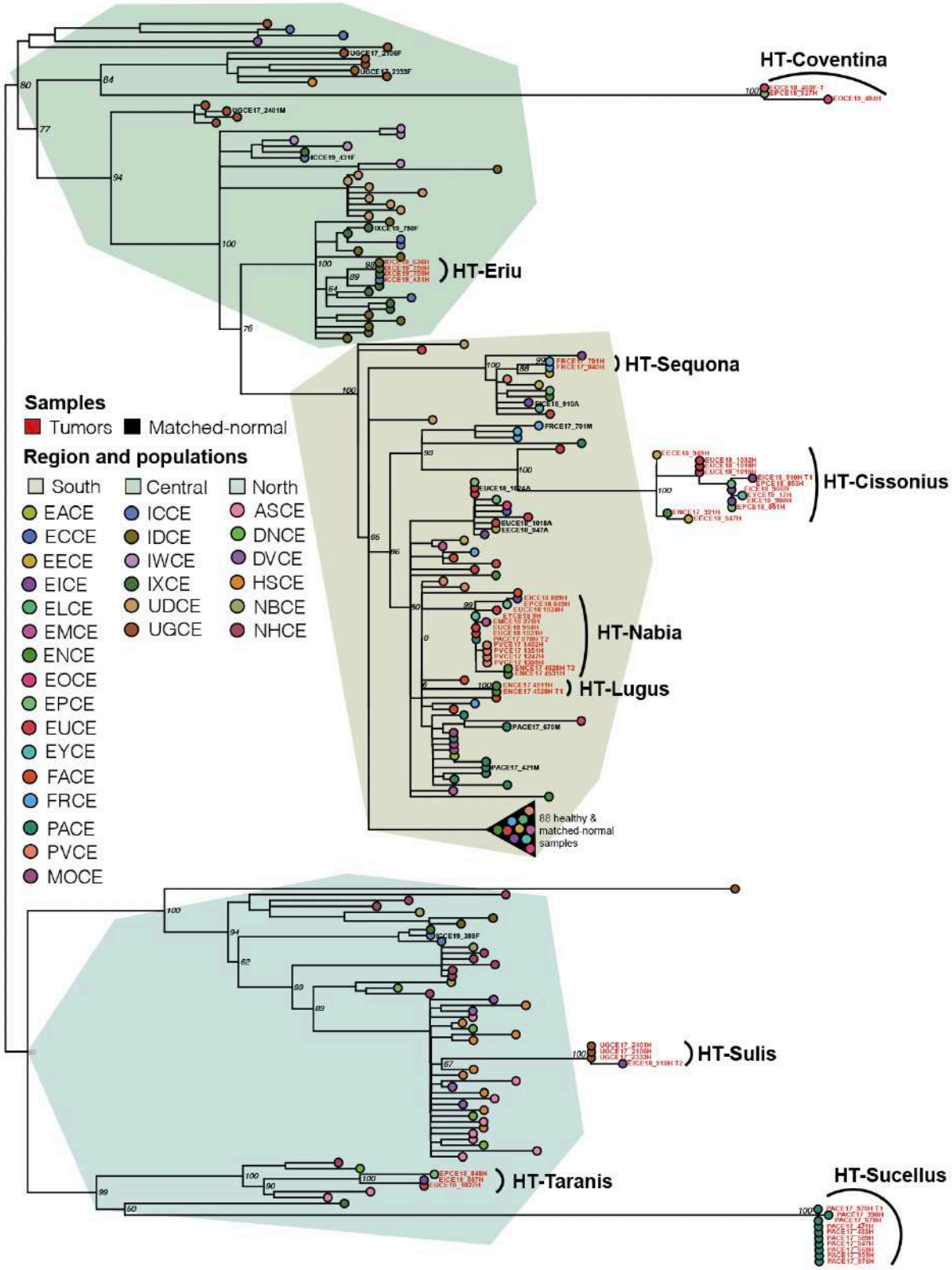


Figure 35. Maximum-likelihood (ML) phylogeny of cockle transmissible cancers amd their hosts based on mitogenomes. Sample codes of tumours (red) and several matched-normal (black) samples are provided. Numbers at nodes are statistical support values (bootstrap proportions) shown for relevant nodes only. Nine major lineages are recovered and named based on geographical or genetic characteristics. Samples are coloured by location and three main clusters (north-central-south) are shown to highlight the geographical patterns previously described in the literature. Tree is midpoint rooted.

Not always the sister taxa of cancer lineages are from its geographical region (Figure 36D); all cancer samples were found on central and southern areas but some lineages closest relatives are northern individuals (*i.e.*, *HT-Sulis*, *HT-Sucellus*, *HT-Taranis*), some are samples of the same region where they were found (*i.e.*, *HT-Eriu*) and the majority are southern samples (*i.e.*, *HT-Lugus*, *HT-Sequana, HT-Cissonius*, *HT-Nabia*). It should be noted that some lineages (*i.e.*, *HT-Sucellus*, *HT-Coventina*) are separated from their sister-taxa by long branches possibly reflecting that they are older lineages (*i.e.*, more mutations have been accumulated) and maybe originated in other locations while nowadays only persist in those where they are found. In other words, this probably reflects expansion and extinction of cancer lineages along the oceans.

Time estimations were performed with a standard substitution rate (0.01 substitutions per million year) of invertebrate mitochondrial genomes (Allio *et al.*, 2017) and using a midpoint rooted tree as a fixed tree and uncorrelated relaxed clock model (Drummond *et al.*, 2006). Fixing or letting the tree to be estimated made no significant difference in the estimated ages as well as using one or three partitions. However, Yule tree prior estimated always younger ages but we show the coalescent ages because this model better fits this kind of intraspecific data. As a consequence of being one locus, the accuracy of time estimates is low, but they can give us an idea of the relative ages of mitochondrial cancer lineages (Table 7). Notably, *HT-Sequana* and *HT-Eriu* lineages are the most recent while *HT-Coventina* and *HT-Sucellus* the oldest (Figure 36E).

Table 7. Time estimates of mitochondrial cancer lineages origin (*i.e.,* TMRCA cancer lineage and closest normal samples). Three partitions, coalescent tree prior.

Cancer MT lineages	MRCA estimated age (MY) Mean [HPD*]
HT-Coventina	0.269 [0.115 - 0.533]
HT-Sucellus	0.261 [0.101 - 0.531]
HT-Taranis	0.1 [0.033 - 0.183]
HT-Cissonius	0.072 [0.024 - 0.133]
HT-Nabia	0.062 [0.021 - 0.092]
HT-Sulis	0.054 [0.013 - 0.079]
HT-Lugus	0.051 [0.011 - 0.078]
HT-Eriu	0.037 [0.005 - 0.049]
HT-Sequana	0.033 [0.004 - 0.039]

*Highest Posterior Density

Topology testing on both ML and Bayesian phylogenies supported eight out of nine cancer lineages being independent (Table 8). For two lineages (i.e., *HT-Lugus* and *HT-Nabia*) the hypothesis that they are in fact a single lineage could not be rejected. Both were found in Noia, Spain (ENCE) but one of them is more widely distributed towards the south (Figure 36B).

Table 8. Topology test results. Different hypothesis tested are listed through the implemented constraints. *(A)* Results of Shimodaira-Hasegawa (SH) test (Shimodaira, Hasegawa, 1999) and of approximately unbiased (AU) test (Shimodaira, 2002). *(B)* Posterior probability of hypothesis based on frequency of compatible trees.

	(A)				*(B)*		
	Maximum likelihood						
Hypothesis	SH p-value		AU p-value		Posterior Probabilities Odds	Trees supporting hypothesis	Significantly worse than unconstrained tree?[1]
Unconstrained tree	1	+	0.502	+	*NA*	*NA*	*NA*
All cancer lineages are monophyletic	0	-	4.11e-05	-	0	0/18000	Strongly significantly worse
HT-Sucellus + *HT-Taranis* are monophyletic	0.0015	-	0.00025	-	0	0/18000	Significantly worse
HT-Nabia + *HT-Lugus* are monophyletic	0.712	+	0.548	+	0.06861	1235/18000	Not significantly worse
HT-Cissonius + *HT-Nabia* + *HT-Lugus* are monophyletic	0.0006	-	4.99e-13	-	0	0/18000	Significantly worse

[1]The constraint tree is considered to be significantly worse if the P value is lower than 0.05 and posterior probabilities odds lower than 0.05.

Unrooted or midpoint trees are shown because the closest-related species *Cerastoderma glaucum* was not a good outgroup as we lose the interspecific structure of our species of study (*Cerastoderma edule*).

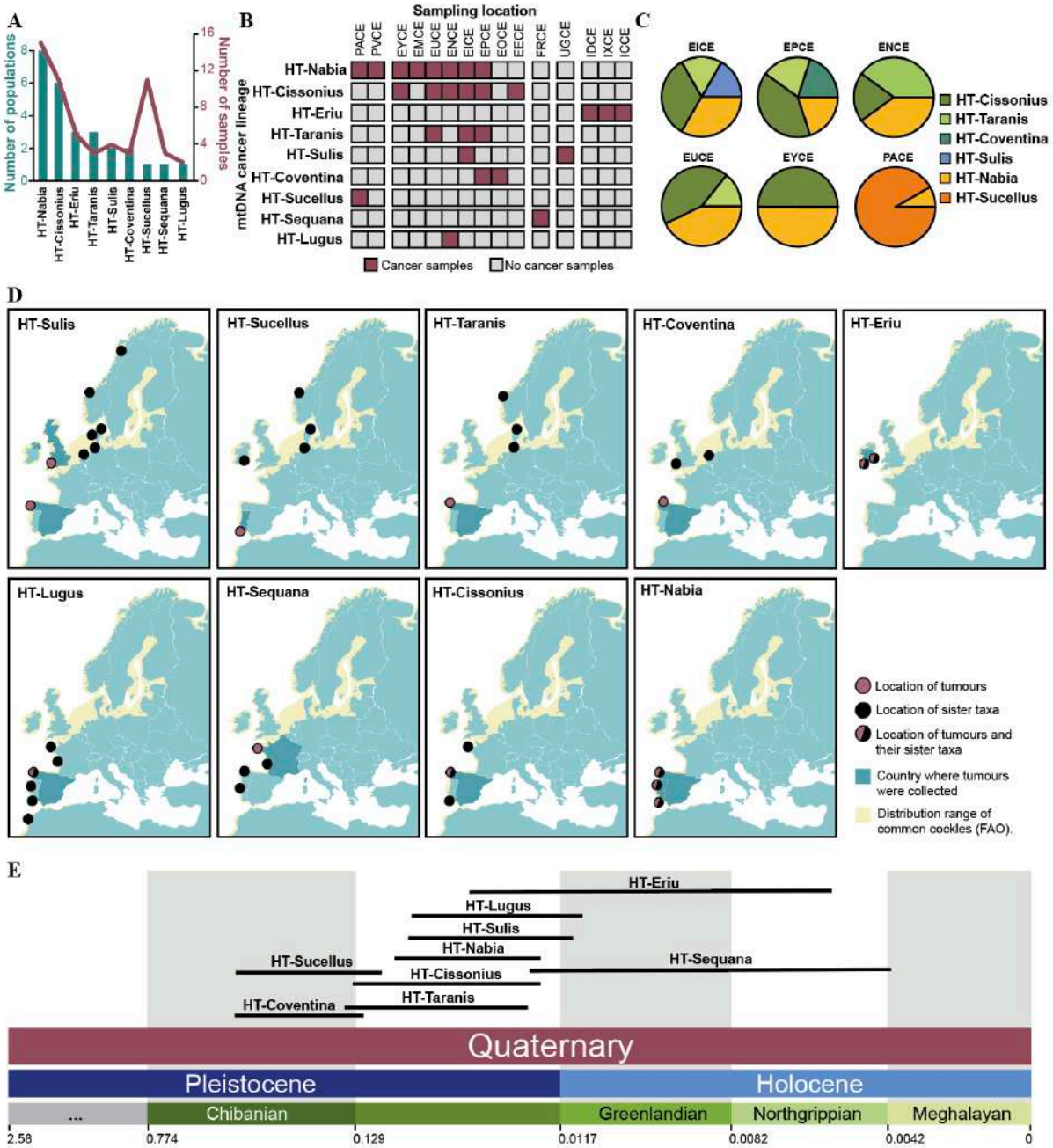


Figure 36. Structure of mitochondrial clonal lineages. **(A)** Number of populations (bars) and samples (line) per mitochondrial cancer lineage. **(B)** Sampling locations where a mitochondrial cancer lineage was found. **(C)** Pie charts of mitochondrial cancer lineages per populations where more than one mitochondrial cancer lineage was found. **(D)** Maps per lineage indicating the locations of the sister taxa to each cancer lineage that can be identified in the phylogeny. **(E)** Schematic representation of the age ranges estimated for each cancer lineage.

At this point, these nine cancer lineages were consistent with two hypothesis: (1) multiple cancer origins or (2) several horizontal transfers or mtDNA captures from healthy cells as it was described for the transmissible cancer of dogs (Rebbeck, Leroi and Burt, 2011).

2.3.4.2. Histopathological and nuclear makers: two lineages

Two nuclear cancer lineages of cockle transmissible cancers from one Spanish population (Noia, ENCE) were previously reported by Metzger et *al.* (2016) using nine polymorphic microsatellite loci and a 3kb region of the *EF1α* gene. Those cancer lineages were usually classified as neoplasia A and B reflecting different features of cells observed through histological sections (Figure 15, Carballal *et al.*, 2001) and 82% of our sequenced tumour samples were classified under the category of type A. No additional phenotypes on HN samples were discovered across the distribution range of these cockle transmissible cancers.

Phenotypically we had two groups (type A and B), even in this large HN collection (Figure 29E). To investigate if nuclear cancer lineages as in Metzger *et al.* (2016) or if more lineages could be found with nuclear markers in agreement with mtDNA lineages, we initially screened tumours and healthy samples with 19 microsatellites, 12 already published (Martinez et *al.* 2015) and 7 additional identified bioinformatically (5 trinucleotides, 2 tetranucleotides). Microsatellites are short DNA sequences consisting of tandem repeated motifs that vary in length, typically from 1 to 6 bp long commonly present in non-coding regions and are characterized by high levels of repeat length polymorphism that are the result of two mutation mechanisms; replication slippage and unequal crossover (Munchen, 1992). It seemed a good approach to differentiate tumour cells (mainly found in haemolymph – H, and gills B) from healthy host cells (two tissues were used: adductor muscle A, mantle M) as it can be observed for four microsatellites in Figure 37A.

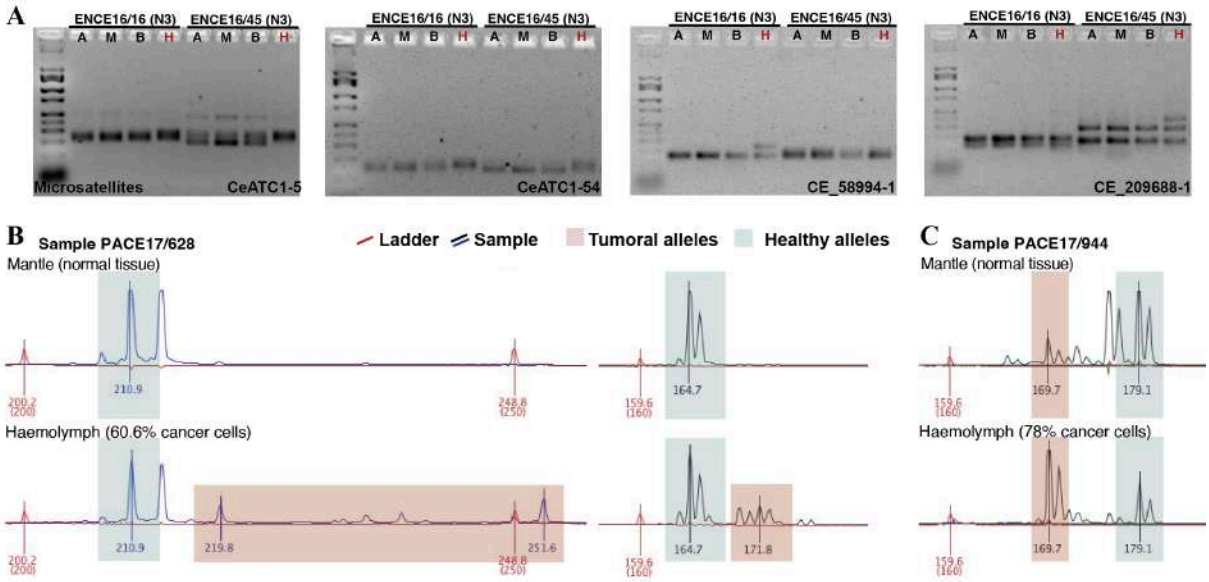


Figure 37. Microsatellite analysis of cockle transmissible cancers. **(A)** Electrophoresis gels of four microsatellites amplified in four tissues (A=muscle, M=mantle, B=gills, H=haemolymph) of two highly neoplastic (N3) cockles. **(B)** Electropherogram of two microsatellites genotyped in two tissues (normal and tumour) of a neoplastic cockle, two genotypes are shown in the tumoral tissue (two alleles shaded in red and one allele shaded in blue). **(C)** Electropherogram of one microsatellite genotyped in two tissues (normal and tumour) of a neoplastic cockle, both tissues show two genotypes, but peak heights are opposite.

As the electrophoresis gel bands did not have enough resolution to identify the alleles, we decided to genotype them through Sanger sequencing. Certain microsatellites in some paired samples behaved as expected, two or four alleles in the haemolymph and only one or two in the matched-normal tissue (Figure 37B); for some cases in which HN had already infiltrated tissues we were able to see tumour and healthy alleles in both samples (Figure 37C). However, most samples gave unexpected results, missing peaks, not comparable data between the pairs and/or

same alleles. As an exception, there was one microsatellite that seem to mark very well all the cancers classified as type B (Figure 38A). In general, microsatellites differentiate healthy cockles from cancer as more than the expected 1-2 genotypes could be seen (Figure 37B-C) but they did not mark all the cancer samples as one (Figure 38A-B). Several factors could be influencing these results such as the annealing of primers in these potential old lineages as they were designed for contemporary cockle DNA and the primer could be preferentially annealing in healthy cells instead of cancer cells due to old polymorphisms present in that region and therefore the sensitivity of the amplifications might be amplifying preferentially healthy cells; the ploidy of cancer cells that it is higher in cancer cells or systematic laboratory errors. For this reason, we discarded the use of this approach to investigate the cancer lineages and we continued with the histological classification under two categories.

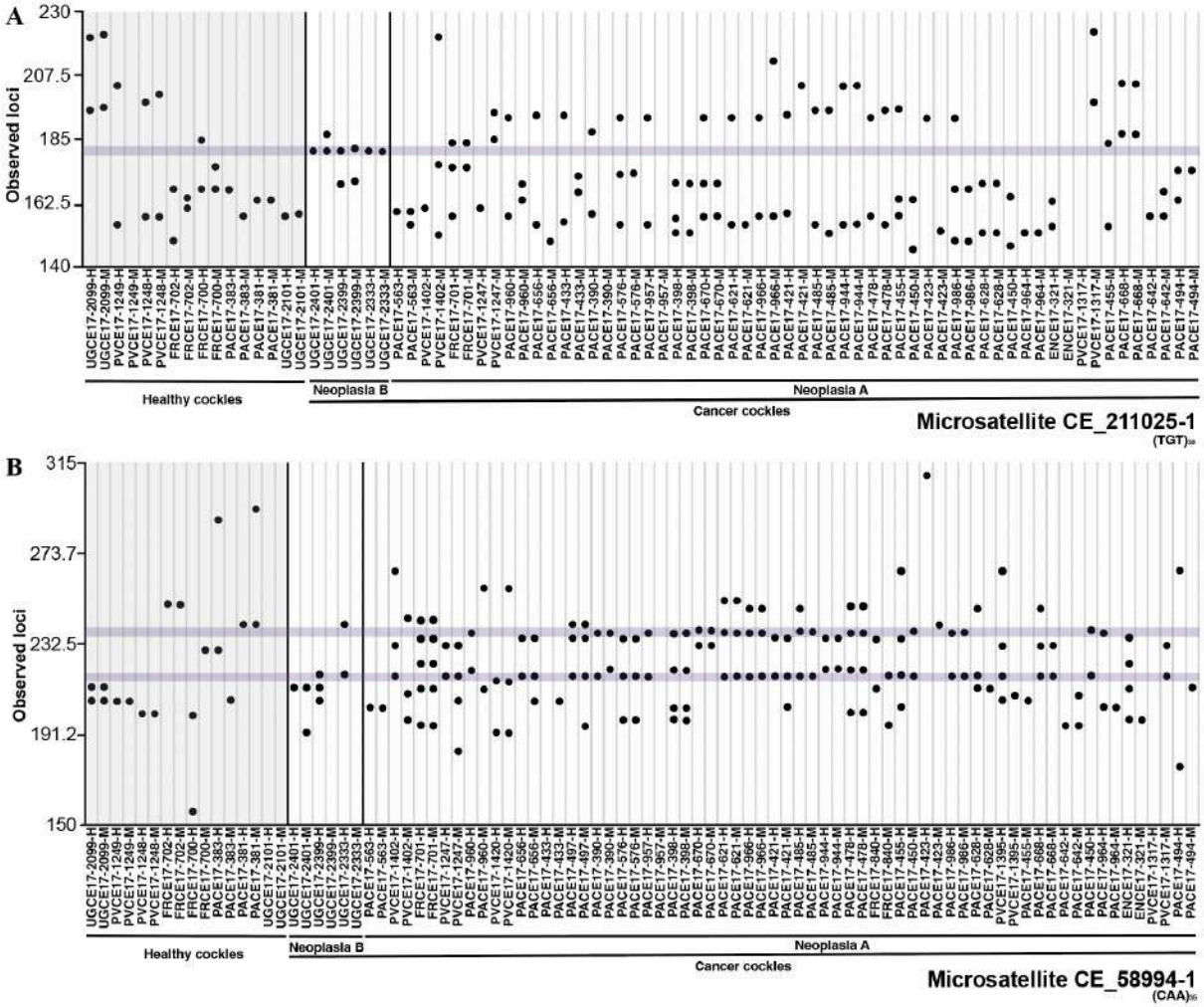


Figure 38. Two microsatellite loci across the dataset. **(A)** Alleles observed for one in-house designed microsatellite, samples diagnosed by histology as neoplasia B show a unique repetition locus (shaded in purple). **(B)** Alleles observed for one in-house designed microsatellite, no clear conclusions can be drawn but several cancer samples share unique repetition locus.

2.3.4.3. Evolutionary history of cockle transmissible cancers

In a nutshell, mitochondrial and nuclear DNA seem to have different evolutionary histories. While nine cancer lineages have been found by analysing the mitochondria, no hints of those lineages were seen when observing their phenotype (Table 9) or analysing microsatellite data. These results cannot be explained simply by a high mutation rate in the mitochondria of HN samples and suggest that HN lineages periodically acquire the mitochondria of their hosts as it has already been proven to happen in a different transmissible cancer (Rebbeck, Leroi and Burt, 2011). Cancers are usually characterized by a high metabolic rate (and thus mutation rate) and, in the case of transmissible cancers that have a longer lifespan, mitochondria accumulates deleterious mutations allowing cell-selection of cancer cells that capture mitochondria from its host (Rebbeck, Leroi and Burt, 2011).

Table 9. Histological diagnosis versus mt cancer lineage.

Cancer MT lineages	Histological diagnosis
HT-Sulis	B
HT-Sequana	A
HT-Eriu	A
HT-Coventina	B
HT-Taranis	B
HT-Lugus	A
HT-Cissonius	A
HT-Nabia	A
HT-Sucellus	A

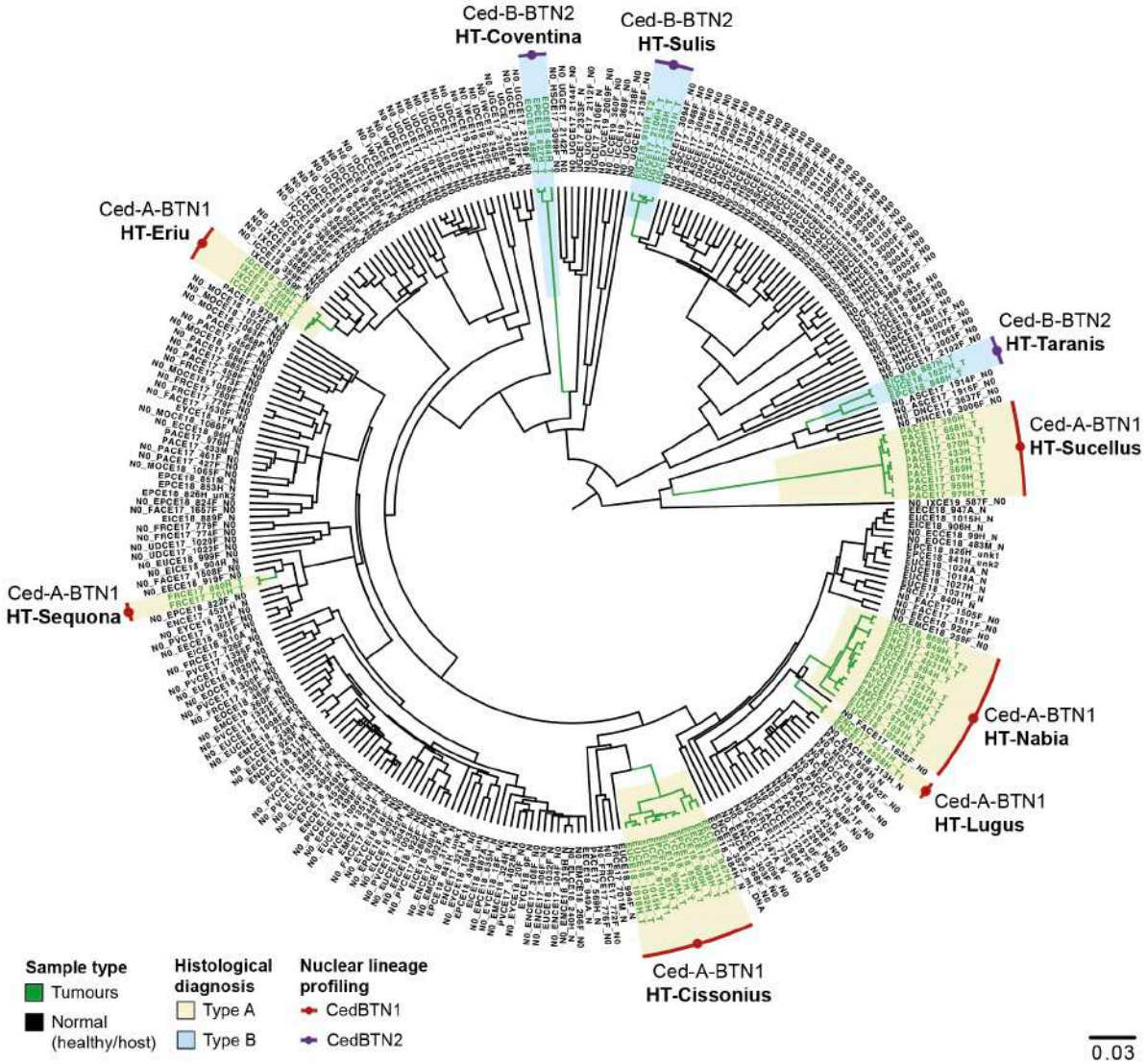


Figure 39. Bayesian phylogeny of cockle transmissible cancers based on mitogenomes. Sample codes and branches of tumours are coloured in red while matched-normal and healthy samples in black. Nine major mitochondrial cancer lineages are highlighted in yellow/blue depending on the histological diagnosis of the

samples. No outgroup was included; therefore, an unrooted tree is shown. This phylogeny is included in the Appendix A – Supplementary material where posterior probability values can be observed.

Mitochondrial cancer lineages did not correspond to histological phenotypes (nine versus two) although no mixture of phenotypes was found in any mitochondrial cancer lineage, that is, each cancer lineage always showed a single phenotype (Table 10). Sister taxa of the three mitochondrial cancer lineages classified as type B are all clustering with northern and central healthy cockles suggesting that type B might have been usually on northern regions and therefore captured the mitochondria on those regions while type A mitochondrial cancer lineages are found through all the phylogenetic tree (Figure 39).

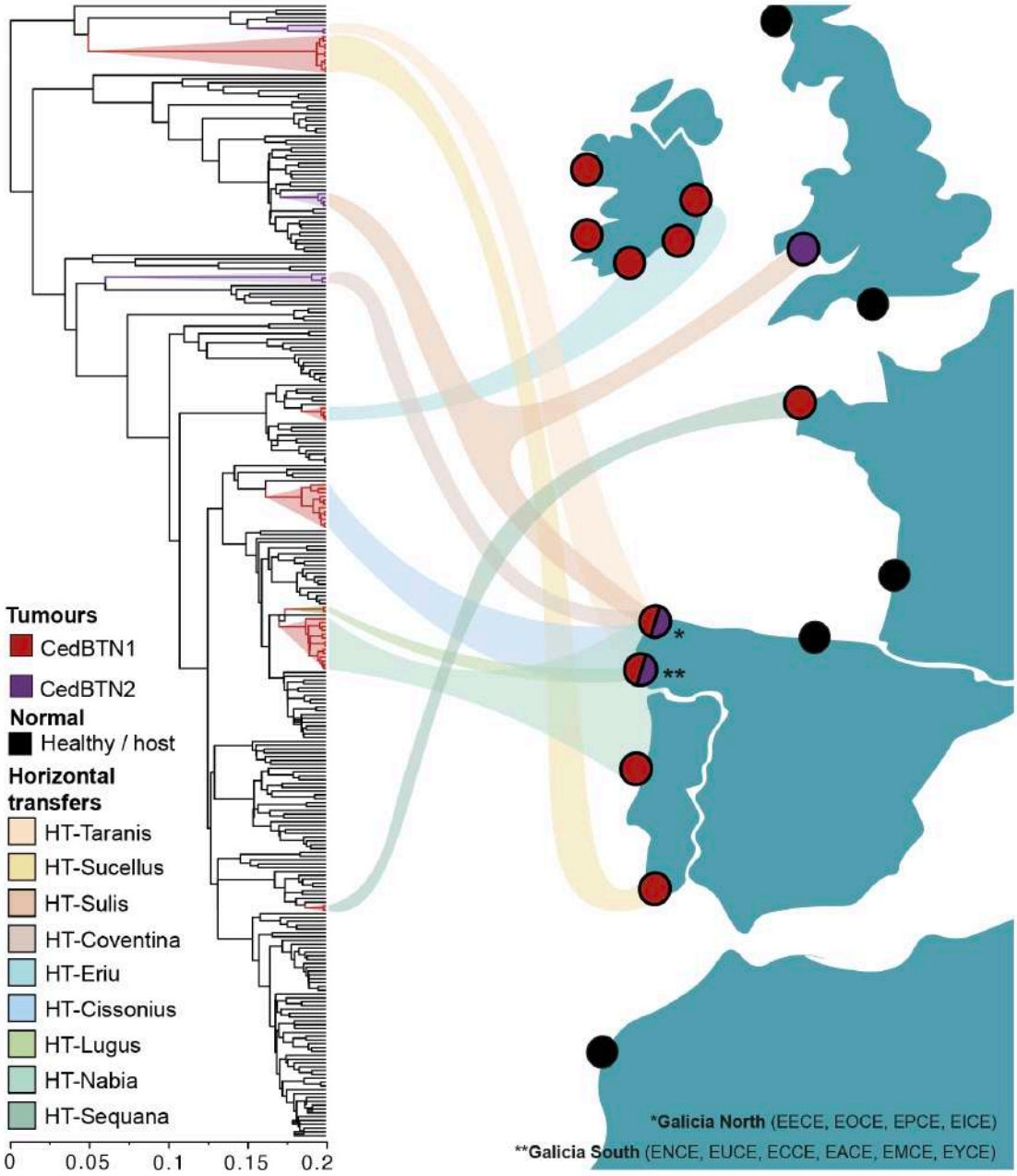


Figure 40. Geographical distribution of horizontal transfers and CedBTN lineages. Points of Galicia (Northwest of Spain) have been merged into two categories: north comprising EECE, EOCE, EPCE, EICE and south comprising ENCE, EUCE, ECCE, EACE, EMCE, EYCE.

Mitochondrial cancer lineages show a geographic distribution generally finding HTs in closed populations except for *HT-Sulis* that has been found in two distant locations, however, nuclear cancer lineage Ced-a-BTN1 is widely distributed while CedBTN2 has only been found in England and Spain surrounded by Ced-a-BTN1 (Figure 40).

2.3.4. COINFECTION OF TWO CANCER LINEAGES IN A SINGLE COCKLE

The groundwork for analysing cockle transmissible cancers allowed us to look for coinfections, that is infections of the same host by distinct tumour lineages.

As far as we know, two cancer co-infections have not been described in species affected by contagious cancers yet. However, the fact that these cancer lineages behave as parasites makes possible this situation because, in order to survive, parasites evolve to increase their ability to propagate in the next host; thus, the target of selection is transmission success (Murgia, 2006). In addition, the dispersal way in bivalves makes coinfection very possible.

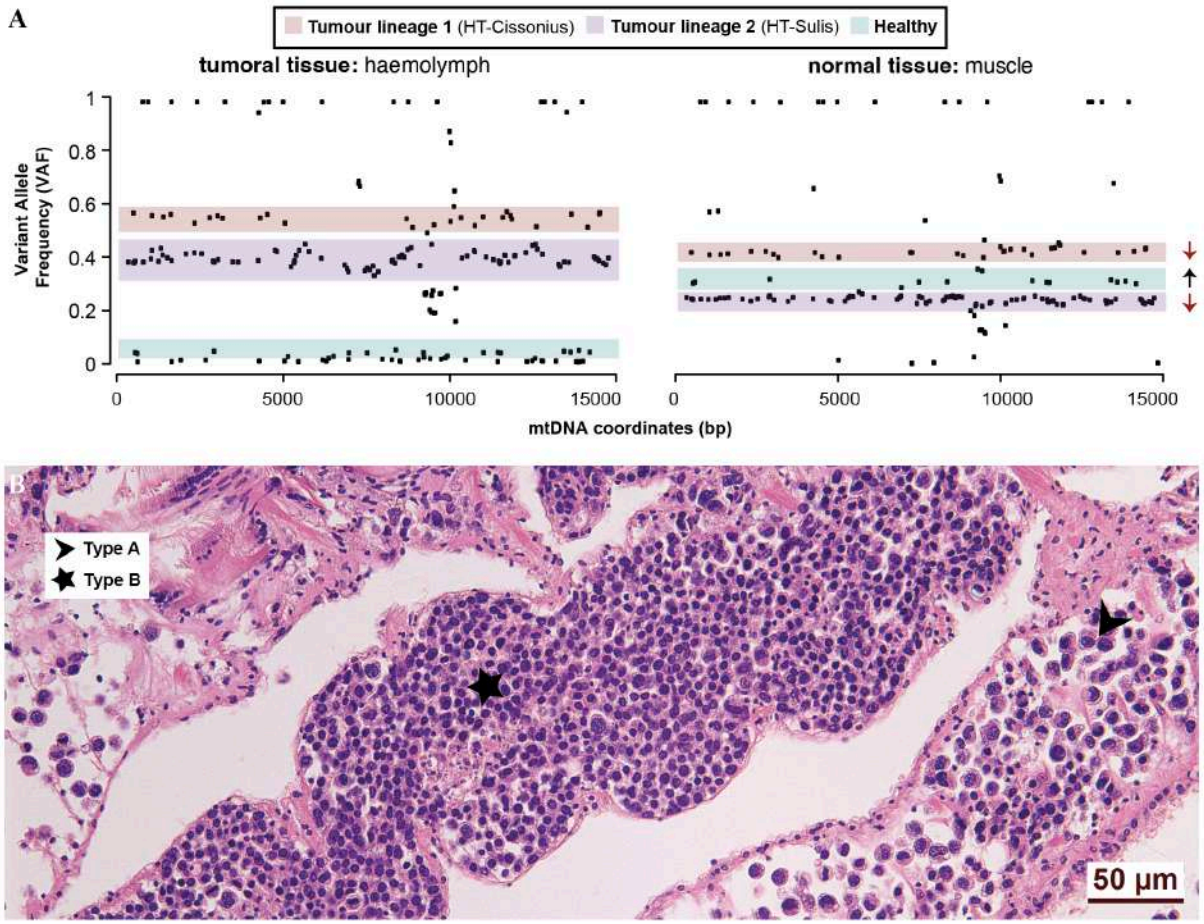


Figure 41. Coinfection of type A and B in a single cockle (sample EICE18/910, cell counting of 100% cancer cells). **(A)** VAF plots showing three haplotypes in the tumoral and normal tissue, tumoral haplotypes decrease in the normal tissue while healthy. **(B)** Histological section showing type A (arrow-head) and type B (star) cancer cells in the same individual (Courtesy of Seila Díaz).

To investigate coinfections, we focus on the samples that had three or more haplotypes in the mitogenome (Figure 34C). For those that were paired samples (two tissues were sequenced) we studied the behaviour of those haplotypes finding that the two haplotypes were higher in the tumoral tissue (i.e., haemolymph) usually decreased in the normal tissue while the

third haplotype behaves in reverse (Figure 41A, Figure 42B). In addition, the percentage of visually counting cancer cells in the haemolymph did not have any correspondence in the tumoral VAF plot unless two haplotypes were sum together. We decided to perform triple clonal deconvolution on these samples, and we added them to the phylogeny finding that two haplotypes of these samples clustered into two different mitochondrial cancer lineages whereas the last haplotype clustered with healthy cockles (Figure 35).

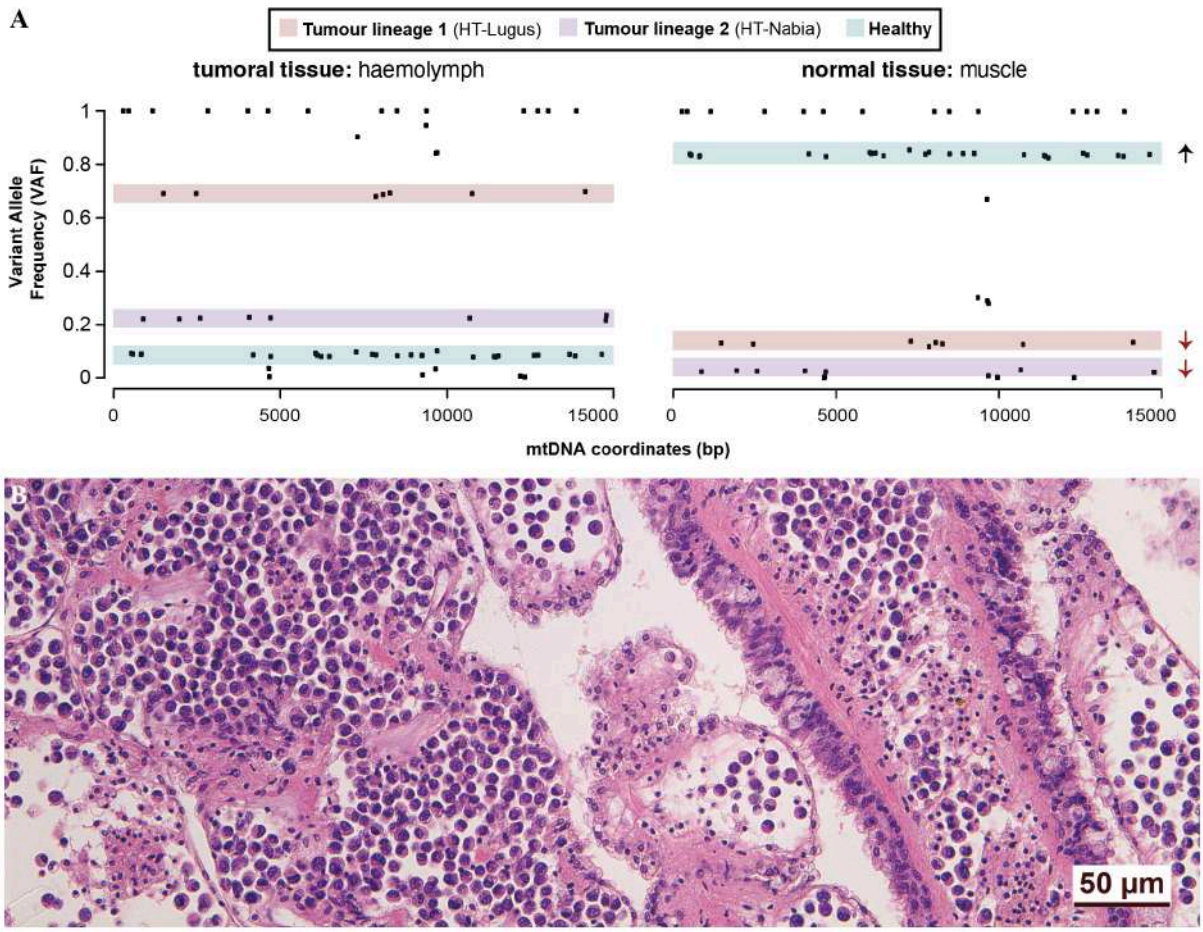


Figure 42. Coinfection of two type A cancer lineages in a single cockle (sample ENCE17/4528, cell counting of 100% cancer cells). **(A)** VAF plots showing three haplotypes in the tumoral and normal tissue, tumoral haplotypes decrease in the normal tissue while healthy. **(B)** Histological section showing type A cancer cells (Courtesy of Seila Díaz).

Nine samples were found to have coinfections of two mitochondrial cancer lineages which represents 13% of sequenced tumoral samples suggesting that coinfection is relative frequent in cockle transmissible cancers. Surprisingly, one coinfected samples had a mitochondrial cancer lineage belonging to the type A (*HT-Cissonius*) while the other belonged to type B (*HT-Sulis*) which was confirmed by histological methods (Figure 41B). The rest of the cases had coinfection of two type A mitochondrial cancer lineages (Figure 42A), therefore the histology was not useful to validate it (Figure 42B). However, histological sections of not-sequenced cancer samples were inspected looking for signs of coinfection and 2%[6] (7/326) of samples were diagnosed with type A and type B being all the cases found from Galicia, Spain.

[6] Samples not sequenced with coinfection of type A and type B are: EOCE18_473, ENCE17_316, ENCE17_4516, ENCE17_4519, EYCE18_44, EYCE18_50.

These results open many fundamental questions about the epidemiology of cockle transmissible cancers. It is usually assumed that the primary disease progresses slowly (Martcheva and Pilyugin, 2006), does it happen also in this case? Are some cancer lineages more aggressive than others? Does being infected by a contagious cancer lineage makes it more susceptible of being infected by a second one? Are there cockles infected by three contagious cancer lineages? Does infection by the same cancer lineage happens often? Is it an age-dependent pattern? Does coinfection depend on the density of affected individuals? Future research on this topic might answer some of the previous questions.

2.3.5. MITOCHONDRIAL EVOLUTIONARY FOOTPRINTS ON COCKLE TRANSMISSIBLE CANCERS

There are pieces of evidence from genomic footprints supporting the results presented in the previous section and being assets to develop markers to monitor the movement of mitochondrial cancer lineages in the following years.

2.3.5.1. Copy number gain in the mtDNA of three cancer lineages

All cancer samples belonging to three mitochondrial cancer lineages always showed a copy-number (CN) amplification in the same region supported by two clusters of matching reads associated with an increase of coverage (Figure 43A-B). As the CN amplifications did not appear in the matched-normal but they appeared in the tumour, we paid special attention to them as potential molecular markers for the diagnosis of those cancer lineages (Figure 43A).

When comparing the CN amplifications in all the samples, we realized that all the samples belonging to a cancer lineage shared the same start/end coordinates revealed by the coverage change and the cleavage mapping, supporting the fact that those samples belong to a single lineage. However, when comparing the three different cancer lineages showing CN amplifications, they had neither the same start/end coordinates nor the same amplification length (Figure 43C, Table 10).

Table 10. Description of the position, length and support of copy-number amplifications in the mitochondria of three cancer lineages.

Mitochondrial cancer lineage	Coordinates		Length (bp)	No. of samples
	Start	End		
HT-Sucellus	9019	10128	1109	11
HT-Lugus	9471	10167	696	2
HT-Cissonius	9019	10159	1140	12

By inspecting the sequence pattern of the affected region, we saw that all CN amplifications start with the exact same sequence pattern of CGGTGG (Figure 43D) suggesting the importance of this pattern to produce the amplification and therefore supporting how two independent events started in the same coordinate. However, it cannot be ruled out that this region could be susceptible to amplifications with a neutral effect during replication, and, in this case, it is possible that these events are not observed in healthy cockles due to the action of negative selection on the mitochondria.

Regarding the annotations of that region (Figure 43E), the biggest amplifications (*i.e., HT-Cissonius* and *HT-Sucellus* lineages) affected the *mt-ND6* gene, two mitochondrial partial genes (*mt-ND4L* and *mt-CO2*), two tRNAs and a non-functional region that could be the origin of

replication. Nevertheless, the smallest amplification of *HT-Lugus* lineage did not affect the *mt-ND6* but the non-annotated region.

To investigate the copy-number of those amplifications, we sequenced three representative samples with long-read technologies that allowed us to cover the whole region in individual reads. Two of them (*HT-Cissonius* and *HT-Lugus* lineages) had a triplication in the area while *HT-Sucellus* lineage is just a duplication (Figure 43F), these findings are also supported by the proportional increase of coverage (Figure 43C).

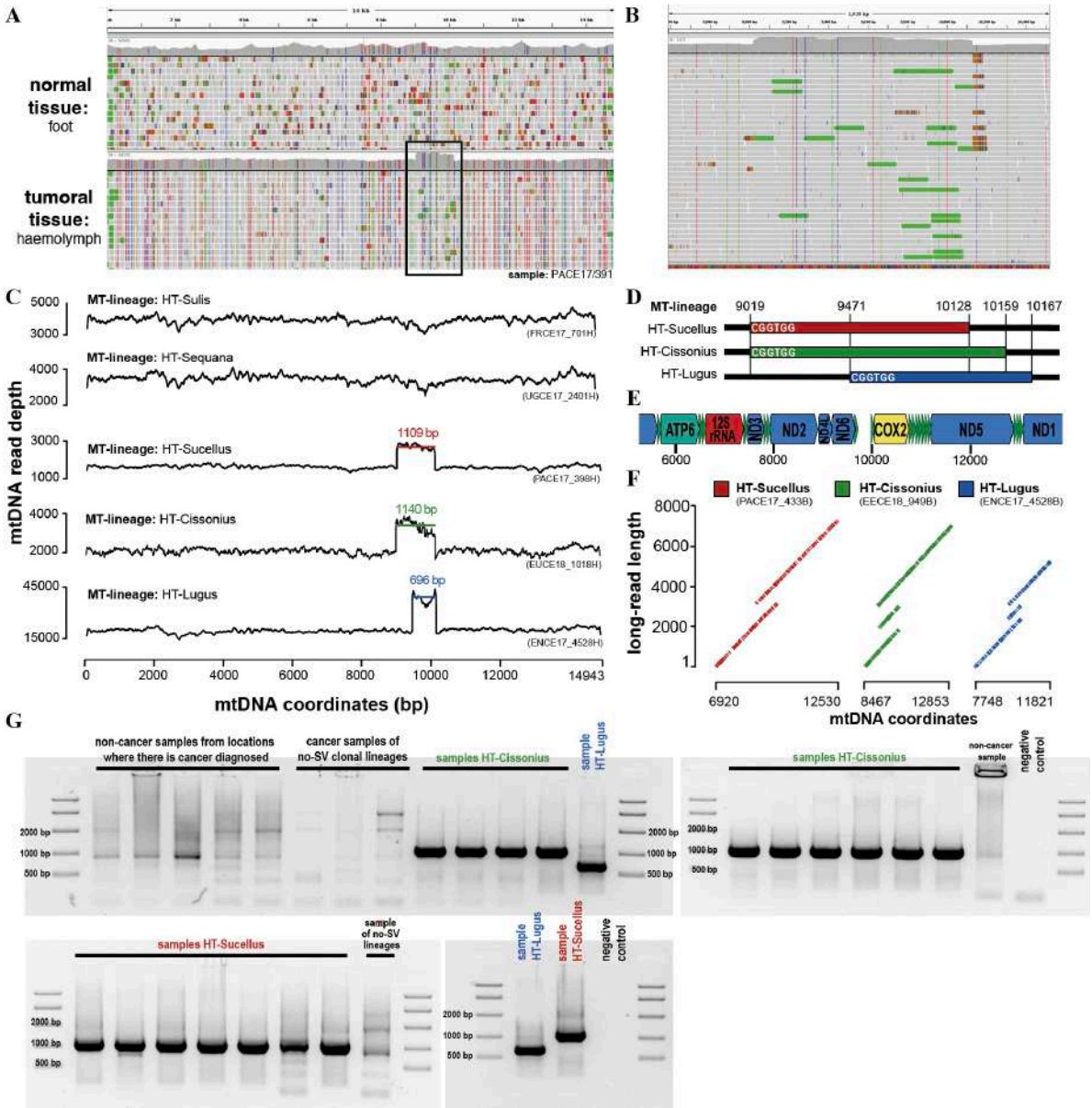


Figure 43. Copy number amplifications on cockle transmissible cancers. **(A)** Coverage increases (black square) in the tumoral tissue of a cancer cockle when aligning reads against the mitogenome. **(B)** Zoom in the area with coverage increase, two clusters of green reads pointing outside can be observed. **(C)** Coverage analysis of five representative samples of different mitochondrial cancer lineages, *HT-Sucellus*, *HT-Cissonius* and *HT-Lugus* show an increase around coordinates 9-10kb. **(D)** Schematic representation of the three cancer lineages with structural variation in the mitogenome, two of them share the starting coordinate and all of them start with the same sequence pattern. **(E)** Annotations and coordinates of the interest area of cockle's mitogenome. **(F)** Dot plot of long-reads generated with minion technology showing that *HT-Sucellus* has a duplication in that area while *HT-Cissonius* and *HT-Lugus* have a triplication. **(G)** Genetic test using these SVs as a molecular marker designed to differentiate these cancer clonal lineages from healthy and other cancer samples.

In humans, a recent study showed a similar event happening in a skin cancer (Yuan *et al*., 2020) which is surprising as the exogenous agents to which melanoma is exposed (*e.g*. ultraviolet light, chemicals…) could be the same as the ones mutating cockle cancer cells when they are being transmitted through water.

Many efforts were done to design diagnostic PCRs using these characteristic SVs of some mitochondrial cancer lineages. Initial results are shown in Figure 43G, same amount of DNA was loaded in every well and samples with the CN amplification show a thick band of the expected size (*i.e*., around 1kb for *HT-Cissonius* and *HT-Sucellus;* around 0.5 kb for *HT-Lugus* lineage). However, more tests must be performed before we can conclude that this is a good diagnostic PCR.

2.3.5.2. From single nucleotide variants to cancer detection

Whole-genome sequencing is expensive and requires high expertise for routine analysis. Molecular markers could be a widely and useful approach to monitor and study the dynamics of cockle contagious cancer lineages.

Currently, we classify the cancer lineages that affect cockles into type A and B according to morphological characteristics and nuclear markers (unpublished data) and we subclassify these types into nine lineages as described in the *Section 2.3.4* (Table 10). Unfortunately, there are no molecular markers described yet for these nine cancer lineages therefore, we have filtered the common SNVs present in each mitochondrial cancer lineage with the PoN to obtain potential somatic mutations that good targets as molecular markers. The number of common SNVs in a cancer lineage varies widely (Figure 44B) but with the filtration target mutations are significantly reduced (Figure 44C). Potential somatic mutations of each cancer lineage (Figure 44C) would be good biomarkers as even when amplifying in ancient samples, they would be able to mark the cancer lineage and not current extinct polymorphisms from the population.

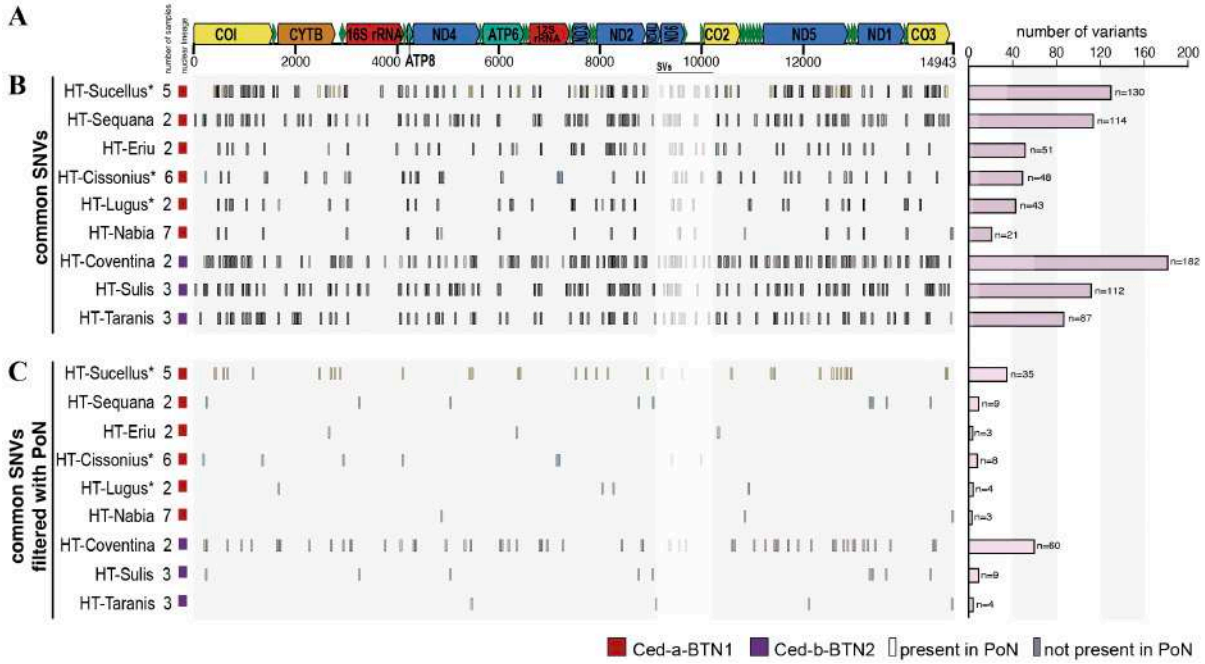
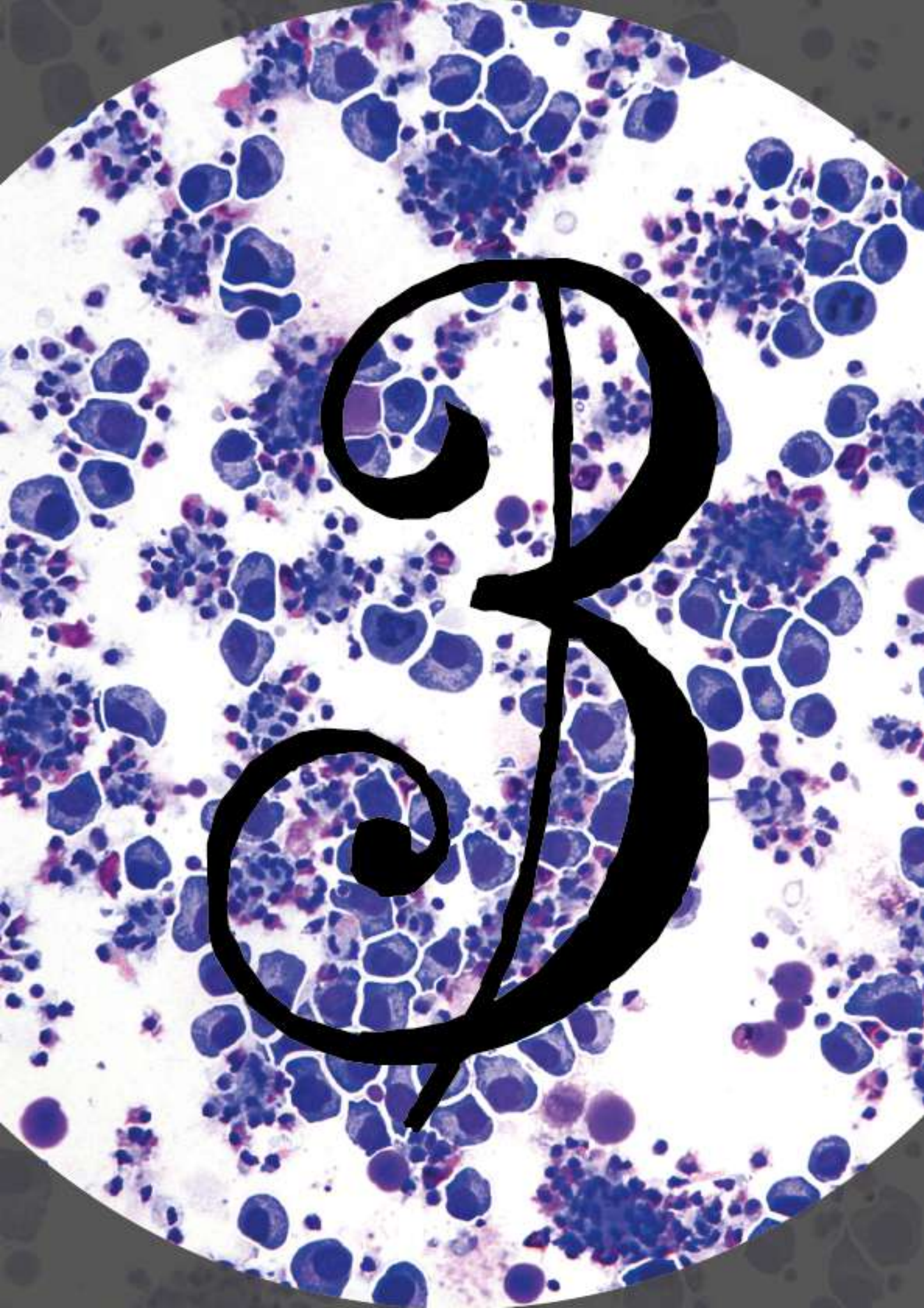


Figure 44. Overall variants called along cockle's mtDNA. **(A)** Mitochondrial genome and its gene annotation. All **(B)** and potential somatic **(C)** variants coloured by mitochondrial cancer lineages. Total number of samples per lineage and type of nuclear lineage are reported, potential somatic variants were called using a panel of normals to filter germline variation. Position of mitochondrial structural variants (SVs) is highlighted and cancer mitochondrial lineages with confirmed SVs in that area are marked with an asterisk (*). Bar plot displaying number of variants per lineage is located on the right or can be found in Appendix A: Supplementary material.

In the future, the validation of these SNVs as molecular markers could reliably distinguish cancer lineages in a simple PCR with no need for sequencing. This promising approach requires a deep understanding of the mtDNA evolution and dynamics of cockle contagious cancer lineages to see how reliable is to use mitochondrial molecular markers.


B

Chapter cover shows cancer and healthy cells from a cockle's haemolymph smear. The image was taken by the doctoral candidate for the research included in this chapter within the framework the European Research Council Starting Grant no. 716290 Scuba Cancers.

Acknowledgments. Yunah Lee and Prof. Young Seok Ju provided essential knowledge and resources for the transcriptomic analysis and interpretation of data of this chapter.

Chapter 3.
Histogenetic origin of cockle transmissible cancers revealed by transcriptomic profiling

In Dublin's fair city
Where the girls are so pretty
I first set my eyes on sweet Molly Malone
As she wheeled her wheelbarrow
Through streets broad and narrow
Crying, "Cockles and mussels, alive, alive, oh!"
Cockles & Mussels (Molly Malone)[7]

"Omnis cellula e cellula: *All cells come from cells*."
Rudolf Virchow

3.1. BACKGROUND

Two contagious cancer lineages are spreading among common cockles with an independent origin. Phylogenetic studies showed that HN has arisen at least twice throughout cockles' evolution and those two cancer lineages share common cyto-histological characteristics that classify them as a single disease but differ enough that two histologic phenotypes were described before their independent origin was discovered (Metzger *et al.*, 2016 and Chapter 2 of this thesis). Despite the discoveries of aetiology in the understanding of HN causation in cockles, that is the transmissible nature of this cancer, the cell type that originated cancer cells on the cancer founder remains unknown.

The nomenclature haemic neoplasia was used in the eighties (Elston *et al.*, 1988) and later it was deprecated in favour of the term disseminated that did not imply the histogenesis of the neoplasia which was unknown (Elston, Moore and Brooks, 1992). It is generally considered to be a sarcoma (neoplasia of mesoderm-derived tissues) although a haematopoietic and a gonadal origin have also been proposed (Alderman, Green and Balouet, 2017) hence, we cannot rule out the possibility of a non-haemocytic cell line being the ancestry of cancer cells.

[7] Popular song considered the unofficial anthem of Dublin that is sung at many events around. For the samplings of this thesis, the doctoral candidate spent one month in Ireland where she met Molly Malone in Grafton Street and enjoyed this song in Irish pubs.

3.1.1. ANATOMY AND REPRODUCTION CYCLE OF COCKLES

The reproduction cycle of common cockles is similar to many other bivalve species (Figure 45A). Common cockles display gonochorism which means that there are only two sexes, and each individual cockle is either male or female (Maia, Barroso and Gaspar, 2021). Sexual maturity and gametogenesis is dependent on the size of the bivalve along with temperature, quantity and quality of food and other environmental factors (Martínez-Castro and Vázquez, 2012). Both males and females show synchronism in gonadal development reaching high fecundity being their spawning season usually on spring to mid-autumn (Maia, Barroso and Gaspar, 2021). Eggs undergo meiotic division to reduce the number of chromosomes to a haploid number before the sperm and egg pronuclei can fuse to form the zygote or fertilized egg (Figure 45A). The larvae are part of the zooplankton community, so they drift in the water column for around 30 days, which allows for passive larval dispersal by ocean currents that drive connectivity and gene flow between populations spread along the Northeast Atlantic (Wilmes and Robins, 2020; Vera *et al.*, 2021). Very few of these larvae will survive their pelagic phase over the following month due to predation and food pressure as they drift through the ocean feeding on phytoplankton (Wilmes and Robins, 2020). When larvae approach maturity, larvae settle and use the foot to crawl on a substrate and is ready to metamorphose. Metamorphosis is a critical time during which cockle changes from a pelagic or planktonic to a sedentary benthic existence (Helm and Bourne, 2004). Once settled in the sea sand, they growth till they reach the adultness and close their life cycle. Growth can be observed on conspicuous rings in the external surface of shells formed every year during winter (Maia, Barroso and Gaspar, 2021).

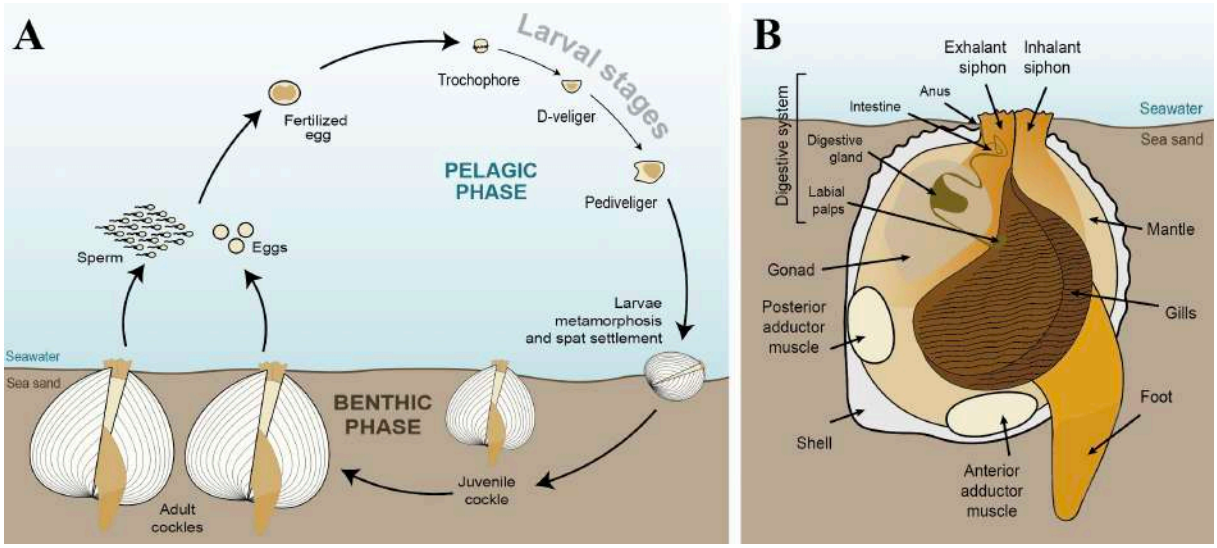


Figure 45. Life cycle and anatomy of common cockles. **(A)** The life cycle of common cockles *Cerastoderma edule* from fertilization to adults: the fertilized egg develops into a larva that undergoes several stages such as the bivalve typical D shape before the pediveliger larva that develops the foot, metamorphosed spat settles into the sea sand and grow into juveniles and eventually adults that will release sperm or eggs and close the cycle. **(B)** Internal anatomy scheme of common cockles' soft parts. Both infographic schemes were designed for this thesis by the doctoral candidate consulting the available literature and using her scientific experience dissecting these animals.

By opening shell valves of cockles, soft parts of the animal can be observed (Figure 45B). Soft parts are covered by the **mantle**, which is composed of two thin sheaths of tissue, thickened at the edges with small tentacles at the tips of the **siphons**. Water is drawn into the cockle through the inhalant siphon, through the gills and then is returned to the surrounding water through the exhalant siphon. **Adductor muscles** close the valves being the anterior termed as

the "quick muscle" because it contracts to close the valves shut, and the posterior as the "catch muscle" because it holds the valves in position when they have been closed. At the base of the visceral mass is the **foot**, an organ that is used to burrow into the substrate and anchor the animal in position. Two pairs of **gills**, an organ specialized for filter feeding from the water as well as for respiration, are located on each side of the body. Gills filter food from the water and direct it to the labial palps, which surround the mouth, a short oesophagus leads from the mouth to the stomach which is surrounded by the digestive gland, then it continues to a curled intestine that extends into the anus. A crystalline style can often be seen in histological sections of the **digestive system** which is believed to assist in mixing food and release enzymes. The **circulatory system** is simple but difficult to trace; the heart pumps the haemolymph; aortas carry it to all parts of the body and the venous system is a vague series of thin-walled sinuses through which haemolymph blood returns to the heart. The **nervous system** consists of three pairs of ganglia with connectives. **Gonads** occupy a major portion of the visceral mass and is generally only evident during the breeding season. Sperm is discharged in a thin, steady stream through the exhalent siphon while discharge of eggs is more intermittent, and they are emitted in clouds from the exhalent siphon (Helm and Bourne, 2004).

3.1.2. HISTOGENESIS OF MAMMAL CONTAGIOUS CANCERS

Contagious or transmissible cancers have been described to be naturally occurring in mammals such as dogs and Tasmanian devils (Murgia *et al.*, 2006; Pearse and Swift, 2006; Pye *et al.*, 2016). The histogenesis of the canine transmissible venereal tumour (CTVT) remains unclear although immunophenotypic suggested a histiocytic origin of CTVT (Murgia, 2006; Hendrick, 2017) and, in cell culture, tumour cells undergo a morphological transformation from round cells to fibroblast-like cells (Murgia, 2006).

Devil facial tumour disease (DFTD) have arisen at least twice throughout evolution (Murchison *et al.*, 2010, 2012; Pye *et al.*, 2016; Stammnitz *et al.*, 2018). Cancer lineage DFT1 has been proposed to be neural-crest-derived tumours originating from Schwann cells by analysing the differential expression of miRNAs and confirming it with quantitative PCR (Murchison *et al.*, 2010). Similarity of tissue markers suggest a similar cell of origin for the other cancer lineage, DFT2, however, the Schwann cell marker PRX is not expressed in DFT2 in contrast to DFT1 (Stammnitz *et al.*, 2018).

3.1.3. TRACING THE CELL-OF-ORIGIN OF CANCER

A tumour originates from a normal cell that has undergone tumorigenic transformation as a result of genetic mutations, the cell-of-origin of a tumour is the normal cell that receives the first cancer-causing mutations (Rycaj and Tang, 2015). Cancer cells may retain transcriptional features of the cells from which they derive. Therefore, it is possible to gain insights into the origin of tumour cells by identifying the cell type that cancers most closely resemble. To enable a systematic cancer and normal transcriptome comparison, all cell types of the individual have to be characterized at single-cell resolution. The challenge then lies in identifying cancer cells resemblance within the healthy data atlas, as cancer cells may resemble other non-neoplastic cells as an important caveat of this reasoning is that the plasticity of the cancer cell transcriptome may obliterate mRNA traces of the cancer cell of origin (Coorens and Behjati, 2022).

Histopathology and gene-expression profiles of tumours often remain relatively stable during progression from primary tumour to metastasis and even end-stage disease (Visvader,

2011) providing a good scenario to investigate the origin of these cancer cells by means of transcriptomic analysis.

3.2. MATERIALS AND METHODS

3.2.1. COCKLES COLLECTION AND CANCER DIAGNOSIS

Cockles *Cerastoderma edule* were collected from natural beds from four locations of the Atlantic coast of Spain (Noia, ENCE; Baiona, EYCE) and Portugal (Algarve, PACE; Aveiro, PVCE). Samplings were carried out from 2017 to 2021. All samples arrived at the laboratory alive and were maintained in a tank with closed-circuit of running seawater for 48 h before the diagnosis and further procedures. Same facilities and ethical approvals as in *Section 2.2.1*.

Disseminated neoplasia was firstly diagnosed by examination of haemolymph cell monolayers. Haemolymph was withdrawn from the adductor muscle of every bivalve sample using a 23-gauge needle attached to a 5 ml syringe. 50 µl of haemolymph were mixed with 150 µl of cold modified Alsever's anti-aggregate solution (Bachere, Chagot and Grizel, 1988) and cyto-centrifuged onto slides (130 g, 7 min, 4 °C). The haemolymph cell monolayers were fixed and stained with the kit Hemacolor (Merck) and examined on a Leica CTR6 LED light microscope for diagnosis. A manual counting of 500 cells was performed to obtain a parameter of purity for the subsequent analysis. Diagnosis was verified through histological sections and neoplasia types were differentiated by size and cell interaction where (i) type A were larger and more scattered and (ii) type B smaller, clustered and more compressed (Carballal *et al.*, 2001; Figure 15).

For each specimen, organs (visceral mass, gills, mantle, foot and gonad when available) were dissected, fixed in Davison's solution and embedded in paraffin. Then, 5 μm thick sections were micro-dissected and stained with Harri's haematoxylin and eosin and examined using a light microscope for histopathological analysis.

A species determination was performed by species-specific PCR amplification of their ribosomal DNA ITS region (Freire, Insua and Mendez, 2005). Amplifications were performed in a final volume of 25 µl; the reaction mixture contained 20 ng/ µl of genomic template DNA, 1 mol/L of each primer, 2.5 µl of dNTPs at 2 µM, 0.5 µl of Taq polymerase (Sigma Aldrich, 5 uds/µl) and 2.5 µl of the polymerase buffer. PCR products were checked on 2% agarose gels stained with SYBR-Safe.

3.2.2. COCKLES LARVAE PRODUCTION

Cockles collected in 2017 from Noia (NW Spain) were kept in 50 L tanks with filtered seawater at 20°C. Spontaneous spawning was observed and larvae were cultured in individual 150 L cylindrical-conical tanks at a density of 8 ± 3 larvae mL-1 with sea water filtered at 1 µm and treated with UV, slight aeration, temperature 19.0 ± 1.4 ºC, in an open circuit with a renewal of 5% volume / hour. The diet consisted of *Tisochrysis lutea* (ECC038), *Chaetoceros neogracile* (ECC007), *Phaeodactylum tricornutum* (ECC028) and *Rhodomonas lens* (ECC030) in a ratio of 1:1:1:1 (according to the cell count), and *Tetraselmis suecica* (ECC036) was included from the seventh day of culture. The daily diet was administered automatically every 4 hours in 6 daily intakes, maintaining a constant concentration in the tank of 20-40 cells µl-1. Larvae were collected from different spontaneous spawning and were classified by microscopical features. To obtain different stages of samples for RNA-seq, samples were collected at less than 24 hours (mainly trocophore stage), 4 days (trocophore stage and D-veliger stage), 11 days (mainly D-veliger stage) and 15 days (mainly pediveliger stage) postfertilisation.

3.2.3. RNA EXTRACTION AND SEQUENCING

Tissues, larval stages and haemolymph were preserved in RNAlater (Qiagen) first frozen with liquid nitrogen and then moved to a −80 °C freezer until extraction. Total RNA was extracted using the RNA extraction kit (RNeasy, Qiagen, Inc., Chatsworth, CA) following the manufacturer's protocol. RNA purity was evaluated with Nanodrop One (Thermo Fisher Scientific), RNA yield was measured in a Qubit fluorometer with the broad range kit (Thermo Fisher Scientific), and RNA integrity was evaluated in a 4200 TapeStation System (Agilent).

RNA sequencing was done in Macrogen Inc. (Seoul, South Korea) to where samples were shipped on dry ice. After quality check, libraries were prepared using the TruSeq RNA with Ribo-Zero library (Illumina). Amplified libraries were sequenced 100M reads/sample with 150 bp paired end reads (250 bp insert size) on an Illumina NovaSeq6000 platform.

3.2.4. REFERENCE GENOME AND PROCESSING OF RNA-SEQ RAW READS

Raw reads were assessed with FastQC (version 0.11.7) and mapped to the draft reference genome of *C. edule* provided by the Scuba Cancers Project (ERC-2016-STG) using STAR version 2.7.3a (Dobin *et al.*, 2013). Before mapping the reads of all samples, five out of thirteen parameters were tested in two samples (the healthy ENCE17_H_Pool and the cancer PACE17_656H) to optimize the mapping environment for this species. Forty eight combinations of parameters that were frequently reported as critical factors for the performance of STAR (Table 11, Veeneman *et al.*, 2016; Baruzzo *et al.*, 2017) were run and the combination showing the highest proportion of uniquely mapped reads and exonic reads was selected based on STAR and Qualimap version 2.2.1 (Okonechnikov, Conesa and García-Alcalde, 2016) results. Alignments were also visually checked using the Integrative Genomics Viewer (Robinson *et al.*, 2011).

Table 11. Tested parameters for the optimization of STAR mapping for RNA sequencing reads.

Parameter name	Description (STAR manual)	Default value	Tested parameter*
outFilterMismatchNmax	Maximum number of mismatches per pair. Alignment will be output only if it has no more mismatches than this	10	10; 33
seedSearchStartLmax	It defines the search start point through the read. The read is split into pieces no longer than this value. Maximum length of seed.	50	12; 30; 50
AlignSJoverhangMin	Minimum overhang (block size) for spliced alignments	5	5; 15
AlignSJDBoverhangMin	Minimum overhang for annotated junctions	3	1; 3
outFilterType	Type of filtering Normal: standard filtering using only current alignment BySJout: keep only those reads that contain junctions that passed filtering into SJ.out.tab.	Normal	Normal; BySJout

*Selected parameter for mapping all samples is underlined.

Mapped reads from all samples were quantified by RSEM version 1.3.1 (Parrish, Hormozdiari and Eskin, 2014), generating tables of read counts and TPM values. A total of 14067 genes were captured from the samples.

3.2.5. DIFFERENTIAL GENE EXPRESSION ANALYSIS

Overall raw read counts were normalized by regularized log transformation method using DESeq2 version 1.34.0 (Love, Huber and Anders, 2014). Genes that were significantly up-

regulated in specific tissue were identified by differentially expressed gene analysis and those with high p-value (>0.05) were filtered out from the list of the genes. The top 60 genes for each tissue (comparing each tissue against all the other tissues), genes that had lowest adjusted p-value were selected as ‘tissue specific genes’.

3.3. RESULTS AND DISCUSSION

3.3.1. A GENE EXPRESSION ATLAS OF COCKLE ORGANS AND ITS LARVAL STAGES

Transcriptome profiling has become widely adopted to characterize the status and diversity of biological samples (Cieślik and Chinnaiyan, 2018). To illuminate the cell of origin of cockle transmissible cancer lineages, we needed the profile of cockle healthy tissues. Therefore, we carried out transcriptome sequencing with Illumina paired-ends in RNA samples isolated from seven healthy cockle tissues/organs (Figure 46A): foot (Figure 46B), gills (Figure 46C), mantle (Figure 46D), digestive gland (Figure 46E), gonad (Figure 46F), adductor muscle and haemolymph (Figure 46G). In addition, we sequenced four larval stages ranging from the D-larva (Figure 46H) to early juveniles (Figure 46I).

Over 70% of reads were mapped in most samples except for the initial larval stage (LCE10_<24h) for which more than 50% of reads remain unmapped with the parameters used. To deeply characterize the relationships among the tissues/organs analysed, tissue specific genes were selected using DESeq2 ($log2FC > 0$ & $p\text{-value} < 0.05$). A total number of 480 genes were selected in a balanced way across the different samples and tissues (Adductor muscle: 60, Digestive: 60, Foot: 60, Gills: 60, Gonad-siphon: 60, Haemolymph: 60, Mantle: 60, Larvae: 60). Figure 46J shows the patterns organ and tissue-specific expressed genes, as we did not used a single-cell RNA seq approach, tissues/organs consist of an amalgamation of the cell types that integrate that tissue/organ. Therefore, similarities can be observed in highly muscular tissues/organs such as adductor muscle, foot or mantle. Digestive system and gonad are usually within the same area in the visceral mass so probably cell types of both organs might have been included in both dissections and similarities within the transcriptomic profile were expected. However, both tissues are embedded with connective tissue that could also explain the similarities. Gills and larvae show a particular unique profile. The transcriptomic profile of the haemolymph is clearly different from the rest of the tissues/organs, although it can be observed attenuated across all tissues/organs, as would be expected since the hemolymph bathes all the tissues of the cockle.

3.3.2. IDENTIFICATION OF THE HISTOGENETIC ORIGIN OF COCKLE TRANSMISSIBLE CANCERS

A cancer genome encompasses a wealth of information about the identity of its cell-of-origin (Polak *et al*., 2015). We investigated cockle transmissible cancer cells from the haemolymph of eight specimens with a severe stage of HN collected in three different sampling points from Spain and Portugal. Leukaemia-like tumours were diagnosed through cyto-histological examination and samples were classified into two types that were cyto-histologically different: type A (Figure 47A) characterized by a pleomorphic nucleus and a looser arrangement of cancer cells in the connective tissue, and HN haemolymph type B (Figure 47B) characterized with smaller cancer cells than type A, a tighter arrangement and rounded, smaller nucleus (Carballal *et al*., 2015; Metzger *et al.*, 2016). Previous studies have shown an independent origin of these two cancer lineages (Metzger *et al*., 2016).

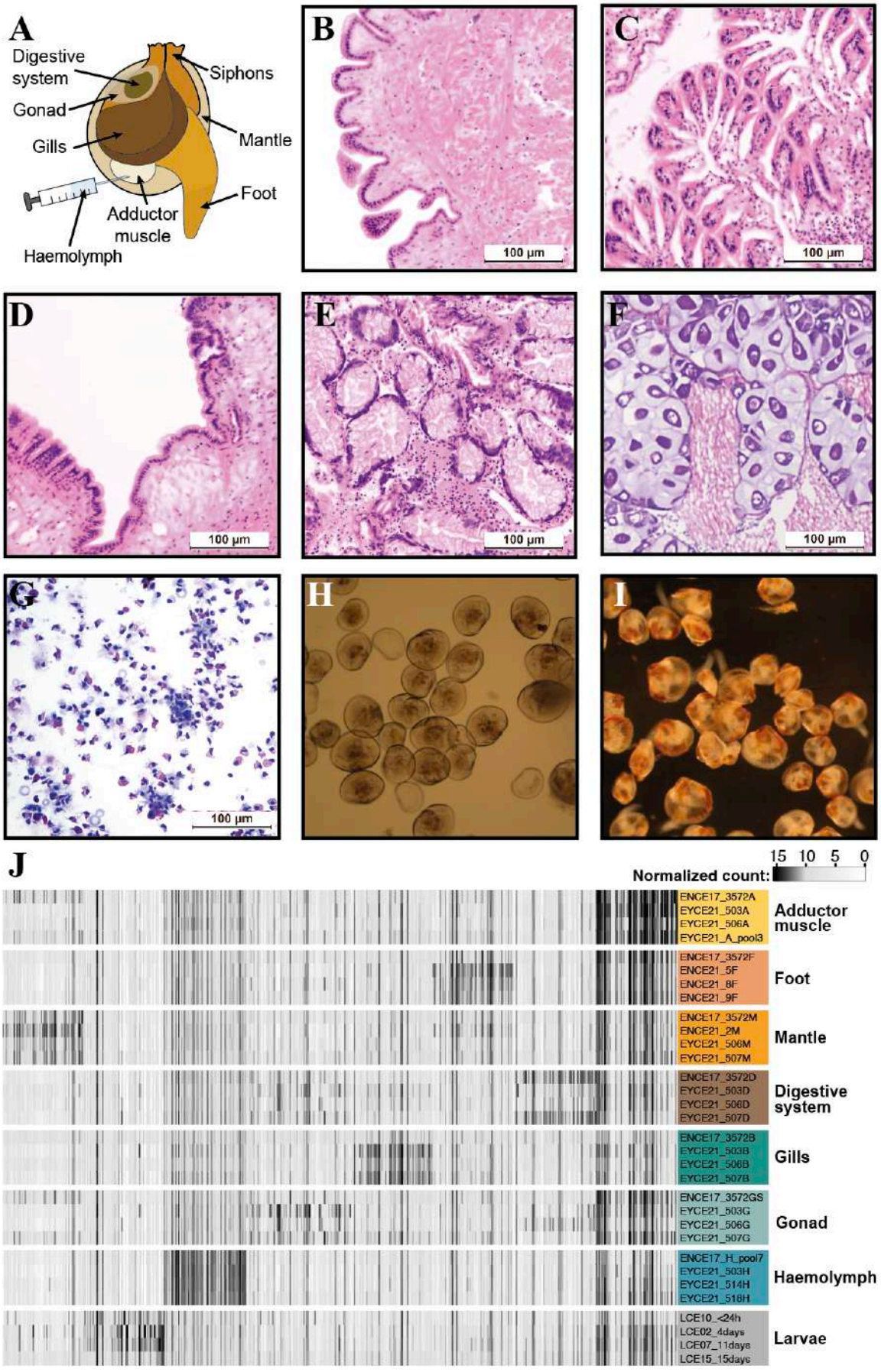

A
Digestive system
Gonad
Gills
Siphons
Mantle
Foot
Adductor muscle
Haemolymph
B
C
D
E
F
G
H
I
100 µm
J
Normalized count:
15 10 5 0
ENCE17_3572A
EYCE21_503A
EYCE21_506A
EYCE21_A_pool3
Adductor muscle
ENCE17_3572F
ENCE21_5F
ENCE21_8F
ENCE21_9F
Foot
ENCE17_3572M
ENCE21_2M
EYCE21_506M
EYCE21_507M
Mantle
ENCE17_3572D
EYCE21_503D
EYCE21_506D
EYCE21_507D
Digestive system
ENCE17_3572B
EYCE21_503B
EYCE21_506B
EYCE21_507B
Gills
ENCE17_3572GS
EYCE21_503G
EYCE21_506G
EYCE21_507G
Gonad
ENCE17_H_pool7
EYCE21_503H
EYCE21_514H
EYCE21_516H
Haemolymph
LCE10_<24h
LCE02_4days
LCE07_11days
LCE15_15days
Larvae

Figure 46. Gene expression atlas of healthy cockles. **(A)** Schematic cockle tissues and organs dissected for this study. Histological sections of healthy cockles showing **(B)** foot, **(C)** gills, **(D)** mantle, **(E)** intestine and **(F)** gonad. **(G)** Haemolymph cell monolayer from a healthy cockle consisting of three cell types: granulocytes, hyalinocytes and type III cells (Russell-Pinto et al., 1994). **(H-I)** Photomicrographs of the larval development showing the **(H)** initial stage with D-larva and **(I)** early juveniles. **(J)** Heat map of organ and tissue-specific expressed genes with a total of 60 genes clustered based on tissue and gene expression (high expression values in black); details of the samples used can be found in Table 11.

The integration of cancer samples into the heatmap of organ and tissue-specific expressed genes (Figure 46J, Figure 47C), pointed by similarities to a haemocytic origin of both cancer lineages (light read, light purple). We performed sample distances and unsupervised consensus clustering of tumour, healthy and larval samples by mRNA expression profiles revealing that both cancer lineages clustered together (Figure 47D). Initial stages of larval samples (LC02, LC07, LC10) grouped together although the late larval stage (LC15) lined up with healthy tissues in the Euclidian sample distance analysis. Similarly, organs with tissue similarities or proximity also grouped together (*i.e.*, gills; digestive system and gonads; mantle, adductor muscles and gonad with siphon; initial larval stages; healthy haemolymph). Both cancer lineages (*i.e.*, type A and B) clearly clustered with healthy haemolymph samples. Principal component analyses (PCA) were performed in both datasets, with all genes a cluster of cancer samples with healthy tissues was not clear (Figure 47E), however, when plotting the tissue specific genes cancer samples clustered closed to healthy haemolymph samples (Figure 47F).

Taken together, all our results from comparing transcriptomes of cancer samples to normal tissues/organs suggest that two distinct lineages of transmissible cancer, with distinct morphologies and genotypes, have a histogenesis of haematopoietic origin. Although the plasticity of the cancer cell transcriptome may obliterate mRNA traces of the cancer cell of origin or even change the cancer cell type, we prefer the more parsimonious view that the healthy tissues/organs describe the differentiation state of cancer cells (Coorens and Behjati, 2022). However, healthy haemolymph of cockles is composed of three types of circulating haemocytes – granulocytes, hyalinocytes and type III cells – (Russell-Pinto et al., 1994), whether both cancer lineages arose from the same cell type remains unknown. Future directions fall in distinguishing the haemocyte cell type that originated each of these cancer lineages.

Bivalve transmissible neoplasia (BTN) has been reported in several species (Carballal *et al.*, 2015) but different cell and tissue types show profound differences in their response to cancer driver mutations (Visvader, 2011). Therefore, transcriptomic studies need to be rolled out systematically in other BTN lineages to investigate their cell-of-origin because the histogenesis of two lineages affecting the same species (*i.e.*, cockles) may not reflect the cell-of-origin of other lineages affecting the same or different species. However, in this study we show the same origin for two independent cancer lineages which suggests haemolymph cells in bivalves might be prone to serve as the seed for a malignant cell to be able to colonize other individuals and avoid any immunological response. Further analysis needs to be performed to gain the most complete picture of the origins of contagious cancers in bivalves.

Rather than the disconnected findings on the cancer cell of origin of bivalve contagious cancers, our study provides a framework for comparing the origins of contagious cancers in both mammals and bivalves. A haematopoietic cell origin of cockle transmissible cancers contrasts with that of the canine clonally transmissible cancer which has been proposed to be of histiocytic origin (Ajayi *et al.*, 2018) or that of the Tasmanian devil transmissible cancer originated in a Schwann cell (Murchison *et al.*, 2010). While malignancies arising from the same anatomical site have traditionally been treated as a single disease, here we face cancers

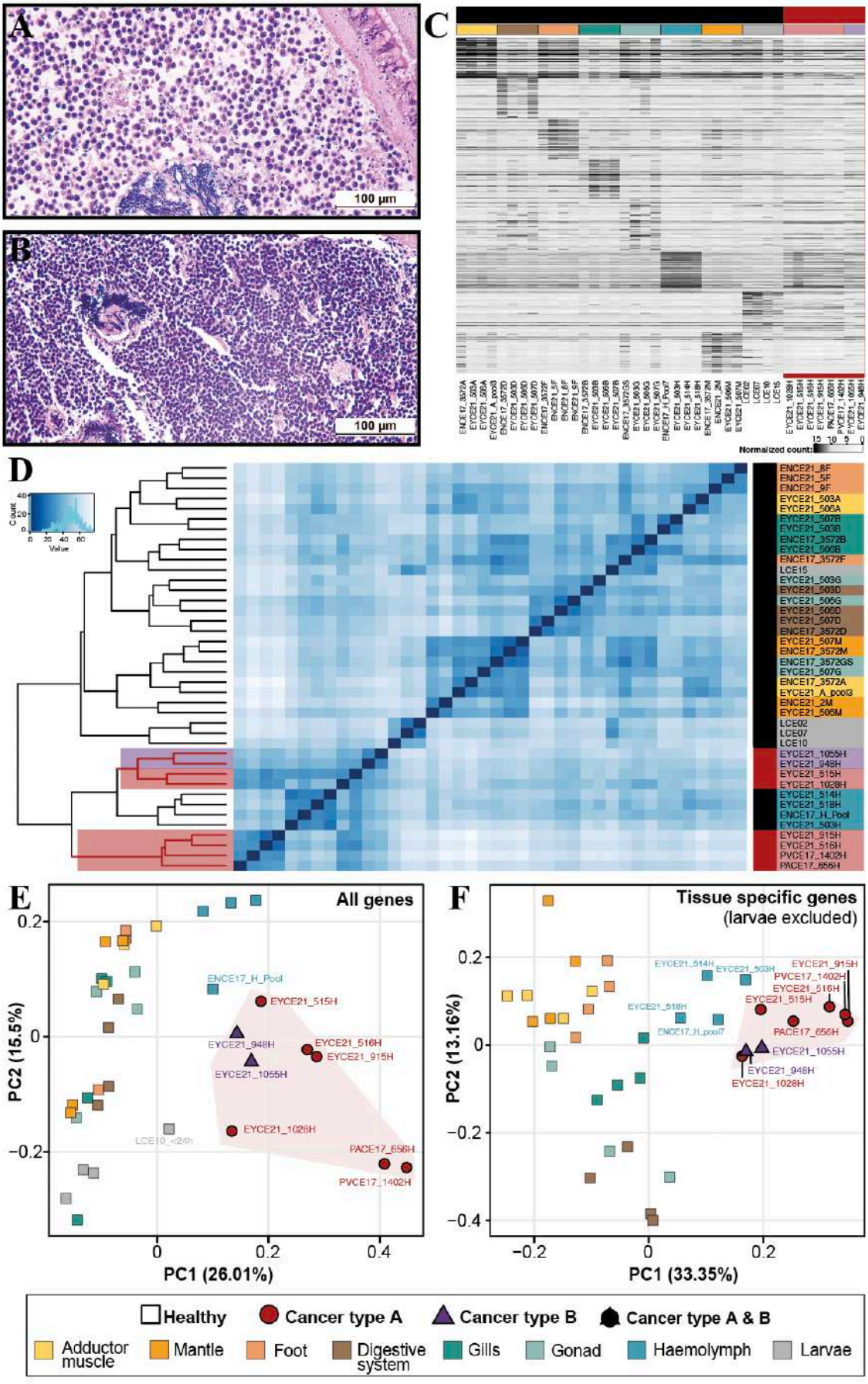

A
100 µm
B
100 µm
C
Normalized count:
D
Count
Value
E
All genes
PC2 (15.5%)
PC1 (26.01%)
ENCE17_H_Pool
EYCE21_515H
EYCE21_948H
EYCE21_516H
EYCE21_915H
EYCE21_1055H
EYCE21_1028H
LCE10_<24h
PACE17_656H
PVCE17_1402H
F
Tissue specific genes
(larvae excluded)
PC2 (13.16%)
PC1 (33.35%)
EYCE21_514H
EYCE21_503H
EYCE21_915H
PVCE17_1402H
EYCE21_516H
EYCE21_515H
EYCE21_518H
ENCE17_H_pool7
PACE17_656H
EYCE21_1055H
EYCE21_948H
EYCE21_1028H
Healthy
Cancer type A
Cancer type B
Cancer type A & B
Adductor muscle
Mantle
Foot
Digestive system
Gills
Gonad
Haemolymph
Larvae

Figure 47. Transcriptomic profiling of two histopathological and genetically independent lineages of cockle transmissible cancers. **(A)** Cancer cells of cancer lineage type A in a histological section. **(B)** Cancer cells of cancer lineage type B in a histological section. **(C)** Heat map of organ and tissue-specific expressed genes with a total of 60 genes clustered based on tissue and gene expression (high expression values in black); details of the samples used can be found in Table 11. **(D)** RNA-sequencing sample distance analysis, samples were clustered using hierarchical clustering analysis, and the dendrograms represent the clustering results; heatmap illustrates the pairwise distances between the indicated samples, with the colours indicating the distances (i.e., the more blue the square, the more similar the samples). **(E)** Principal component analysis (PCA) of all genes shows clustering of cancer samples (red area) near healthy haemolymph samples; while PC1, explaining 26% of the total variance, separates cancer samples from healthy samples, PC2, explaining 16% of the total variance, differentiates healthy tissues. **(F)** PCA of tissue specific genes excluding larvae samples shows again clustering of cancer samples (red area) near healthy haemolymph samples; while PC1, explaining 33% of the total variance, separates cancer samples from healthy samples, PC2, explaining 13% of the total variance, differentiates healthy tissues.

with a contagious behaviour with different cell-of-origin, it will be of interest to define common and unique features of them to understand the histogenesis of transmissible cancers. Our study provides a framework for this work.

Table 12. Cockle specimens and tissues sequenced. Forty tissues of thirty-two specimens (eight neoplastic and twenty-four non-neoplastic) of cockles were sequenced with Illumina paired-end reads. Column 5 shows the cell counting of tumour cells, larvae could not be diagnosed.

Individual	Samples	Description	Type	Tumor purity
Cancer tissues (various specimens)	PVCE17_1402H	Haemolymph, cancer type A	tumor	100 %
	PACE17_656H	Haemolymph, cancer type A	tumor	98.4 %
	EYCE21_515H	Haemolymph, cancer type A	tumor	75 %
	EYCE21_516H	Haemolymph, cancer type A	tumor	98 %
	EYCE21_915H	Haemolymph, cancer type A	tumor	99 %
	EYCE21_1028H	Haemolymph, cancer type A	tumor	85.7 %
	EYCE21_1055H	Haemolymph, cancer type B	tumor	100 %
	EYCE21_948H	Haemolymph, cancer type B	tumor	90 %
Larval stages (various healthy specimens)	LCE10	Larvae of <24 hours (trocophore stage)	larvae	*NA*
	LCE02	Larvae of 4 days (D-veliger & trocophore stages)	larvae	*NA*
	LCE07	Larvae of 11 days (D-veliger stage)	larvae	*NA*
	LCE15	Larvae of 15 days (pediveliger stage)	larvae	*NA*
ENCE17_3572	ENCE17_3572A	Adductor muscle of reference cockle	healthy	0%
	ENCE17_3572B	Gills of reference cockle	healthy	0%
	ENCE17_3572D	Intestine/digestive system of reference cockle	healthy	0%
	ENCE17_3572M	Mantle of reference cockle	healthy	0%
	ENCE17_3572F	Foot of reference cockle	healthy	0%
	ENCE17_3572GS	Gonad and siphons of reference cockle	healthy	0%
EYCE21_503	EYCE21_503A	Adductor muscle	healthy	0%
	EYCE21_503B	Gills	healthy	0%
	EYCE21_503D	Intestine/digestive system	healthy	0%
	EYCE21_503G	Gonad	healthy	0%
	EYCE21_503H	Haemolymph	healthy	0%
EYCE21_506	EYCE21_506A	Adductor muscle	healthy	0%
	EYCE21_506B	Gills	healthy	0%
	EYCE21_506D	Intestine/digestive system	healthy	0%
	EYCE21_506G	Gonad	healthy	0%
	EYCE21_506M	Mantle	healthy	0%
EYCE21_507	EYCE21_507B	Gills	healthy	0%
	EYCE21_507D	Intestine/digestive system	healthy	0%
	EYCE21_507G	Gonad	healthy	0%
	EYCE21_507M	Mantle	healthy	0%
Healthy tissues (various specimens)	EYCE21_514H	Haemolymph	healthy	0%
	EYCE21_518H	Haemolymph	healthy	0%
	ENCE17_H_pool7	Haemolymph of 7 cockles pooled	healthy	0%
	ENCE21_9F	Foot	healthy	0%
	ENCE21_2M	Mantle	healthy	0%
	ENCE21_5F	Foot	healthy	0%
	ENCE21_8F	Foot	healthy	0%
	EYCE21_A_pool3	Adductor muscle of 3 cockles pooled	healthy	0%

Chapter cover shows an infographic adaptation of the abstract video of the research included in this chapter created by Pix Videos and funded by the European Research Council Starting Grant no. 716290 Scuba Cancers. Principal investigator of the project and the company Pix Videos have granted written permission to use the infographic content in this thesis.

CHAPTER 4.
Interspecific cancer contagion between clam species in the Seas of Southern Europe[8]

"This overall flow of genetic information […] is the language used to communicate and express life." Jennifer A. Doudna

"It becomes quite a puzzle – one which evolutionary biologists like to solve, and one that illuminates a path toward new cancer therapies." Jeffrey Townsend

4.1. BACKGROUND

Cancers are clonal cell lineages that arise due to somatic changes that promote cell proliferation and survival (Michael R. Stratton, Campbell and Futreal, 2009). Although natural selection operating on cancers favours the outgrowth of malignant clones with replicative immortality, the continued survival of a cancer is generally restricted by the lifespan of its host. However, clonally transmissible cancers – from now on, transmissible cancers – are somatic cell lineages that are transmitted between individuals via the transfer of living cancer cells, meaning that they can survive beyond the death of their hosts (Murchison, 2008). Naturally occurring transmissible cancers have been identified in dogs (Murgia *et al*., 2006; Murchison *et al*., 2014; Báez *et al*., 2019), Tasmanian devils (Murchison *et al*., 2012; Pye *et al*., 2016) and, more recently, in marine bivalves (Metzger *et al*., 2015, 2016; Yonemitsu *et al*., 2019).

Hemic neoplasia (HN), also called disseminated neoplasia, is a type of leukaemia cancer found in multiple species of bivalves, including oysters, mussels, cockles, and clams (Carballal *et al*., 2015). Although these leukaemias represent different diseases across bivalve species, they have been classically grouped under the same term because neoplastic cells share morphological features (Carballal *et al*., 2015). Some HNs have been proven to have a clonal

[8] *This chapter is a partial reproduction of the published peer-reviewed article:* García-Souto, D.#, **Bruzos, A.L.#**; Díaz, S.#; Rocha, S.; Pequeño, A.; Roman-Lewis, C.; Alonso, J.; Rodriguez, R; Costas, D.; Rodriguez-Castro, J.; Villanueva, A.; Silva, L.; Valencia, J.; Annona, G.; Tarallo, A.; Ricardo, F.; Bratos Cetinic, A.; Posada, D.; Pasantes, J.J.; Tubío, J.M.C. (2022). "*Mitochondrial genome sequencing of marine leukemias reveals cancer contagion between clam species in the Seas of Southern Europe*." *eLife*. 11:e66946. (#equal contribution). doi.org/10.7554/eLife.66946, ISSN: 2050-084X. More information in Appendix B: Publications reproduced in this thesis.

transmissible behaviour (Metzger *et al*., 2015), in which neoplastic cells, most likely haemocytes (i.e. the cells that populate the haemolymph and play a role in the immune response), are likely to be transmitted through marine water. In late stages of the disease, leukemic cells invade the surrounding tissues and, generally, animals die because of the infection (Carballal *et al*., 2015), although remissions have also been described (Burioli *et al*., 2019). Despite the observation that leukemic cells are typically transmitted between individuals from the same species, on occasion they can infect and propagate across populations from a second, different bivalve species (Metzger *et al*., 2016; Yonemitsu *et al*., 2019). Hence, these cancers represent a potential threat for the ecology of the marine environment, which argues for the necessity of their identification and characterization for their monitoring and prevention.

4.1.1. THE WARTY VENUS CLAM (*Venus verrucosa*)

Venus verrucosa Linnaeus (1758), commonly known as the "warty venus" (Figure 48A), is a marine bivalve species characterized by a series of 20 or more prominent concentric ridges intersected by radiating grooves resulting in wart-like spines (Carrilho Rodrigues da Silva, 2012). It is distributed in the Mediterranean, in the Atlantic from Norway to South Africa (Durban) and further east in the Indian Ocean to Mozambique (Poppe and Goto, 2000).

This species is particularly appreciated in France, where it is known as "praire" where registered catches by trawlers reach of 3500 t per year. In the southern Adriatic (Italy) and Greece, fisheries are local, reaching about 500 t per year. In Spain, several scenarios occur. In Galicia (NW Spain), where it is known as "carneiro" is a species captured using traditional methods that are not harmful to the seabed. As it is not much appreciated in its local gastronomy, their average catches of 100 t per year (data Xunta de Galicia; www.pescadegalicia) are exported to France or Mediterranean regions of Spain. In the Balearic Islands (NE Spain), where it is known as "escupiña gravada", it has great gastronomic value and local fisheries are in decline even though the catches are made by apnoea divers. The great economic value that this species acquires in the market has made its illegal captures prosper in some regions as in Malaga (S Spain), where it is known as "bolo" and where its illegal market is of the same order of magnitude as the official catch (Tirado, Salas and Márquez, 2003).

No pathological studies have shown cytohistological characterization of hemic neoplasia (HN) in warty venus clams. However, an isolated case of warty venus clam collected in Galicia was reported abnormal karyotypes which usually is a remarkable feature of hemic neoplasia (Carrilho Rodrigues da Silva, 2012).

Figure 48. Clam specimens. **(A)** *Venus verrucosa*, commonly known as warty venus clam (Natural History Museum Rotterdam, CC-BY SA), **(B)** *Chamelea gallina*, commonly known as striped venus clam (Natural History Museum Rotterdam, CC-BY SA) and **(C)** *Chamelea striulata* (Natural History Museum Rotterdam, CC-BY SA). See Appendix H.

4.1.2. THE STRIPED VENUS CLAM (*Chamelea gallina*)

Chamelea gallina Linnaeus (1758), commonly known as the "striped venus" (Figure 48B), is a subtidal (5-20 m) marine bivalve species characterized by broadly triangular but asymmetrical shells with a round anterior margin but a somewhat elongated posterior. It is distributed in the Mediterranean, in the Black Sea and in the Atlantic from the Portuguese coast to the Gulf of Cádiz (Kosyan and Divinsky, 2019). In Galicia (NW Spain), this species is not found (Trigo *et al.*, 2018).

This species is largely commercialized in the Mediterranean which explain the catches of 55,486 t per year (average of 2010-2015) being Italy (36,462 t per year), Spain (4,803 t per year) and Turkey (3,585 t per year) are the countries with more catches (FAO). However, high fishing pressure enhanced by several irregular mortality events has led to a sharp decrease in abundance of their populations.

4.1.3. THE STRIPED VENUS CLAM (*Chamelea striulata*)

Chamelea striulata Linnaeus (1758), commonly known as the "striped venus" (Figure 48C) like the clam described in *Section 4.1.2* and both species, *C. gallina* and *C. striulata* are very similar, in fact, most morphological differences are only identifiable by experienced observers (type of shell ridges, the lunular shape, shell outline). It is widely distributed from the Lofoten Islands (Norway), throughout the North Sea and the British Isles, south to the western Mediterranean (Iberian Peninsula) and along the Atlantic coast of Morocco, down to Madeira and the Canary Islands. Its distribution overlaps with that of *C. gallina* into the coast of the Algarve (S Portugal), the Gulf of Cadiz, the Strait of Gibraltar, and the Alboran Sea (Rufino *et al.*, 2006).

The taxonomy of these two bivalve species has been an issue of discussion among researchers: some authors considered them a single polymorphic species while others two species or subspecies separated geographically (García-Souto, Qarkaxhija and Pasantes, 2017).

4.2. MATERIALS AND METHODS

4.2.1. SAMPLING OF CLAM SPECIMENS

We collected 570 clam specimens from three different species, from the following countries and locations (Figure 49):

- *V. verrucosa* clams were collected in Spain (Ferrol, Ribeira and Canido, n = 90; Mahón, n = 67), France (Granville, n = 100), Croatia (Split, n = 18), Portugal (Oeiras, n = 19), and Ireland (Carna, n = 50).
- *C. gallina* clams were collected in Spain (Cadiz, n = 50; Mallorca, n = 50) and Italy (Naples, n = 50; Cattolica, n = 57).
- *C. striatula* clams were collected in Spain (Combarro, n = 9).

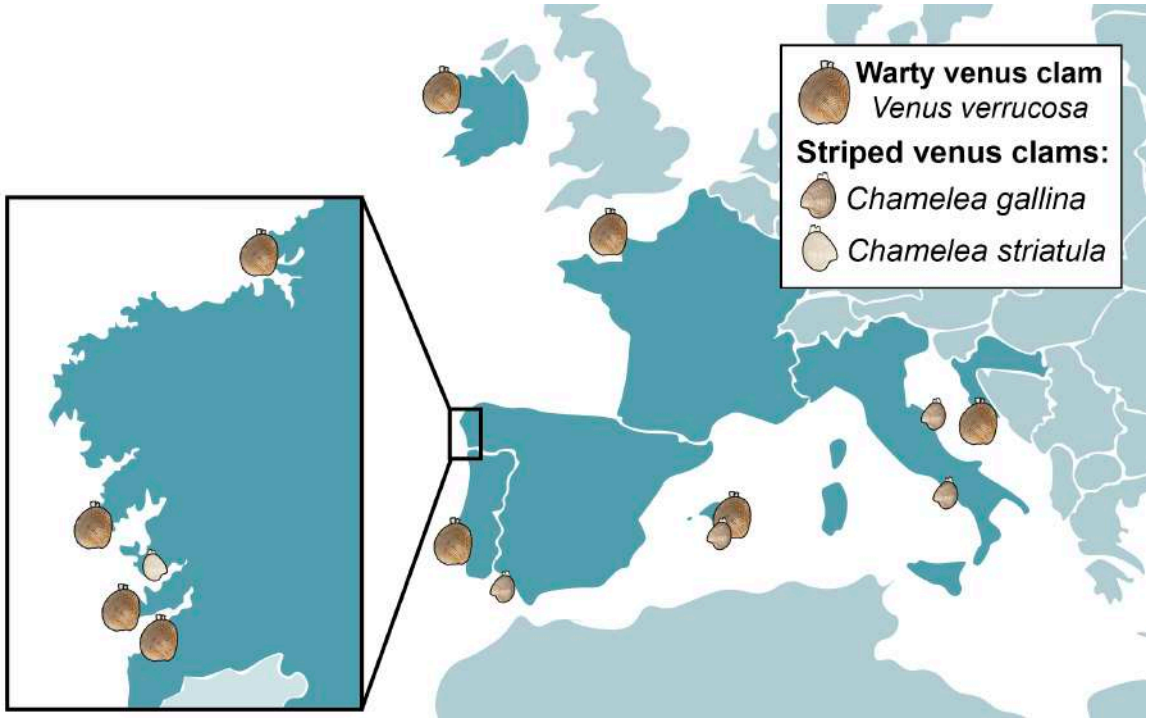


Figure 49. Geographical location of collected clam specimens.

Additionally, we recruited samples from the following specimens from private collections: one *V. verrucosa* clam collected in 2011 in Spain (Islas Cies), four *C. gallina* collected in 2015 in Italy (San Benedetto de Tronto), five *C. gallina* collected in 2015 in Spain (Huelva), and one *C. striatula* collected in 2014 in Spain (Marín, private collection of Dr. Juanjo Pasantes). Same facilities and ethical approvals as in *Section 2.2.1*.

4.2.2. DIAGNOSIS OF HN

We followed standard cytological and/or histological protocols to test and diagnose HN in the clam specimens. However, only histological examination resulted decisive for the diagnosis, particularly in early stages of the disease.

Briefly, for each animal, we extracted 300–2000 ml of haemolymph from the posterior adductor muscle using a 5 ml syringe with a 23 G needle. The haemolymph (50 ml) was diluted in cold Alserver's antiaggregant solution to a 1:4 concentration, and spotted by centrifugation

(130 × g, 4°C, 7 min) onto a microscope slide using cytology funnel sample chambers to produce a cell monolayer. Haemolymph smears were fixed and stained with Hemacolor solutions from Sigma-Aldrich and subsequently examined with a light microscope for the diagnosis of HN.

Tissues (visceral mass, gills, mantle, and foot) were dissected, fixed in Davidson's solution and embedded in paraffin. Then, 5-mm thick sections from each tissue were microdissected and stained with Harris' haematoxylin and eosin and examined using a light microscope for histopathological analysis.

HN was diagnosed and classified according to four disease stages (*i.e.*, N0, N1, N2, or N3) as follows.

- **N0 stage**: no signs of leukemic cells are found[9].
- **N1 stage**: small groups of leukemic cells were detected only in the vessels of the gills and in the connective tissue surrounding the digestive tubules.
- **N2 stage**: leukemic cells spread to different organs, conforming small groups in the connective tissue that surrounds the digestive gland and the gonadal follicles, branchial sinuses, and mantle.
- **N3 stage**: leukemic cells invade the filaments, completely deforming the plica structure in the gill, invade the connective tissue surrounding the gonadal follicles and the digestive gland; in the mantle, they invade the connective tissue, but in the muscle fibres of the mantle and foot, cells appear isolated or in small groups and in lower intensity than in other tissues.

Morphometric analysis (area and radio of cytoplasm and nucleus) of 200 circulating cells per individual (6 N0 and all cancer individuals) was performed on histological sections using NIS-Elements software. ANOVA test was used to differentiate morphometric cell populations.

4.2.3. ELECTRON MICROSCOPY ANALYSIS

Four *V. verrucosa* specimens (two non-neoplastic, ERVV17-2993 and ERVV17-2992, and two with high grade of HN, ERVV17-2995 and ERVV17-3193) were processed for transmission electron microscopy as follows: 2 mm sections of gills and digestive glands were fixed in 2.5% glutaraldehyde seawater for 2 hours at 4°C. Then, tissues were post-fixed in 1% osmium tetroxide in sodium cacodylate solution and embedded in Epon resin. Ultrathin sections were stained with uranyl acetate and lead citrate and examined in a JEM-1010 transmission electron microscope.

4.2.4. FLOW CYTOMETRY AND CYTOGENETICS

Hemolymph of 25 N0 and 3 cancer individuals from Galicia was fixed in 100% ethanol and stored at 20°C. Samples were centrifuged (800g, 10 min, 4 °C), pellets were resuspended in 0.01 M Phosphate Buffer Saline (PBS), then treated with PI (50 um/mL) and Dnase-free

[9] Sometimes referred as 'healthy' on this Chapter although it could happen that the specimens were affected by other pathologies.

Rnase A and finally incubated for 60 minutes at room temperature in the dark. Ploidy of hemolymph cells was analyzed by flow cytometry, 10.000 events of each sample were analysed using a BD Accuri C6 Flow Cytometer, data were analyzed with Flowlogic software.

Mitotic chromosomes of a neoplastic *V. verrucosa* specimen (EVVV11-02) were obtained as follows. After colchicine treatment (0.005%, 10 hr), gills were dissected, treated with a hypotonic solution, and fixed with ethanol and acetic acid. Small pieces of fixed gills were disaggregated with 60% acetic acid to obtain cell suspensions that were spread onto preheated slides. Chromosome preparations were stained with DAPI (0.14 mg/ml) and PI (0.07 mg/ml) for 8 min, mounted with antifade medium, and photographed.

A comparative screening of tandem repeats was performed on the genomes of *C. gallina* and *V. verrucosa* using RepeatExplorer (Novák, Neumann and Macas, 2010) on a merged short-read dataset of both species (500,000 reads each). Short reads of healthy and neoplastic animals were mapped onto both satellite consensus sequences using BWA, filtered according to their mapping quality (MAPQ > 60 and AS >70) and their abundance assessed by means of samtools/bamtools. Satellites *CL4* and *CL17* were selected for FISH purposes and FISH probes were PCR amplified (Table 13) and directly labelled with digoxigenin-11-dUTP (10× DIG Labeling Mix, Roche Applied Science). FISH experiments were performed as described in reference García-Souto *et al*., 2015.

Table 13. Primers used to amplify the satellites CL4 and CL17 on clam specimens for the FISH experiment.

Forward primer	Primer sequence (5' → 3')	Reverse primer	Primer sequence (5' → 3')
CL4F	TCAGAAACCGCTATTTTTCAC	CL4R	AAATGATGCTACGAACCTCC
CL17F	ATTCCAGAAATGTACATGAACAC	CL17R	ATTTTTGCACCAGATGTTCAC

Chromosome preparations, tandem repeats identification and FISH experiments were performed by Daniel García-Souto.

4.2.5. DE NOVO ASSEMBLY OF MITOGENOMES AND ANNOTATION

In total, we performed whole-genome sequencing on 23 samples from 16 clam specimens, which includes 8 neoplastic and 8 non-neoplastic animals by Illumina paired-end libraries of 350 bp insert size and reads 150 bp long.

First we assembled the mitochondrial genomes of one *V. verrucosa* (FGVV18_193), one *C. gallina* (ECCG15_201), and one *C. striatula* (EVCS14_02) specimens with MITObim v1.9.1 (Hahn, Bachmann and Chevreux, 2013), using gene baits from the following *mt-COI* and *16S* reference genes to prime the assembly of clam mitochondrial genomes: *V. verrucosa* (*mt-COI*, with GenBank accession number KC429139; and *16S*: C429301), *C. gallina* (*mt-COI*: KY547757, *16S*: KY547777), and *C. striatula* (*mt-COI*: KY547747, *16S*: KY547767). These draft sequences were polished twice with Pilon v1.23 (Walker *et al*., 2014), and conflictive repetitive fragments from the mitochondrial control region were resolved using long read sequencing with Oxford Nanopore technologies (ONT) on a set of representative samples from each species and tumours. ONT reads were assembled with Miniasm v0.3 (Li, 2016) and corrected using Racon v1.3.1 (Vaser *et al*., 2017).

Protein-coding genes, rDNAs and tDNAs were annotated on the curated mitochondrial genomes using MITOS2 web server (Bernt *et al*., 2013), and manually curated to fit ORFs as predicted by ORF-FINDER (Rombel *et al*., 2002). Then, we employed the entire mtDNAs of *V. verrucosa* (FGVV18_193) and *C. gallina* (ECCG15_201) as 'references' to map reads from

individuals with neoplasia, filter reads matching either mitogenome and assemble and polish their two (healthy and tumoral) mitogenomes individually as above.

Assembly and annotation were performed by Daniel García-Souto.

4.2.6. ANALYSIS OF *mt-COI* SEQUENCES

We retrieved a dataset of 3745 sequences comprising all the barcode-identified venerid clam *mt-COI* fragments available from the Barcode of Life Data System (BOLD, http://www.boldsystemns.org/). Redundancy was removed using CD-HIT (Fu *et al*., 2012), applying a cut-off of 0.9 sequence identity, and sequences were trimmed to cover the same region. Whole-genome sequencing data from both healthy and tumoral warty venus clams were mapped onto this dataset, containing 118 venerid species-unique sequences, using BWA-mem (Li and Durbin, 2009), filtering out reads with mapping quality below 60 (-q60), and quantifying the overall coverage for each sequence with samtools idxstats (Li *et al*., 2009). PCR primers were designed with Primer3 v2.3.7 (Kõressaar *et al*., 2018) to amplify a fragment of 354 bp from the *mt-COI* mitochondrial gene of *V. verrucosa* and *C. gallina* (Table 14), these analyses were performed by Daniel García-Souto. PCR products were purified with ExoSAP-IT and sequenced by Sanger sequencing.

Table 14. Primers designed to amplify *mt-COI* mitochondrial gene in the warty and striped venus clams.

Forward primer	Primer sequence (5' → 3')	Reverse primer	Primer sequence (5' → 3')
mt-COI-F	CCTATAATAATTGGKGGATTTGG	*mt-COI*-R	CAGCTACACACCAWACAAATATA

4.2.7. MITOGENOME COVERAGE ANALYSIS

We further mapped the paired-end sequencing data from healthy and neoplastic tissues from all neoplastic samples onto the 'reference' mitochondrial genomes of *V. verrucosa* and *C. gallina* (two of the previously assembled ones, FGVV18_193 and ECCG15_201) using BWA-mem v0.7.17-r1188 (Li and Durbin, 2009) with default parameters. Duplicate reads were marked with Picard 2.18.14 and removed from the analysis. Read coverage depth was computed with samtools v1.9 (Li *et al*., 2009), summarized by computing the average in windows of 100 bp size and plotted with R v3.5.3.

4.2.8. DRAFT ASSEMBLY OF NUCLEAR REFERENCE GENOMES

We ran the MEGAHIT v1.1.3 assembler (Li *et al*., 2015) on the Illumina paired-end sequencing data to obtain partial nuclear genome assemblies of *V. verrucosa* (FGVV18_193), C. gallina (ECCG15_201), and *C. striatula* (EVCS14_02). Then, single copy genes were predicted with Busco v.3.0.2 (Seppey, Manni and Zdobnov, 2019).

Candidate genes were considered if they (1) were present in the genomes of the three species, and (2) showed variant allele frequencies (VAFs) at exclusively 0, 0.5, or 1.0 in all the sequenced healthy (non-neoplastic) specimens. Under this criteria, two loci were finally selected: a 3914-bp long fragment of *DEAH12,* a gene encoding for an ATP-dependent RNA helicase and a 2.2-kp length fragment of the Transcription Factor II Human-like gene, *TFIIH*.

These analyses were performed by Daniel García-Souto.

4.2.9. IDENTIFICATION OF SNPs ON VARIABLE SINGLE-COPY ORTHOLOGOUS NUCLEAR LOCI

PCR primers (Table 15) were designed with Primer3 v2.3.7 (Kõressaar *et al*., 2018) to amplify and sequence a 441-bp region of the *DEAH12* nuclear gene and a 559-bp fragment of the *TFIIH* gene on neoplastic specimens from *V. verrucosa* and healthy animals from both species (*DEAH12*: 11 *V. verrucosa* and 9 *C. gallina*; *TFIIH*: 15 *V. verrucosa* and 12 *C. gallina*).

We screened for differentially fixed SNVs between both species using the *dapc* function in the R package Exploratory Analysis of Genetic and Genomic Data adegenet (Jombart and Ahmed, 2011).

Table 15. Primers designed to amplify *DEAH12* and *TFIIH* on clam specimens.

Forward primer	Primer sequence (5' → 3')	Reverse primer	Primer sequence (5' → 3')
DEAH12_F	AGGT ATGCTGAAACAAACACTT	DEAH12_R	ACGACAAATTTGATACCTGGAAT
TFIIH_F	TGGCATCTTTGTTACGGAC	TFIIH_R	CTTGTGRTTCTGTATCTGATCAATAA

These variants were later employed to filter the Illumina short reads matching either *V. verrucosa* or *C. gallina* genotypes from the neoplastic animals, and to obtain consensus sequences from tumour and healthy tissue in each sequenced specimen. Read filtering was performed with samtools fillmd (Li *et al*., 2009), while GATK mutect2 (Benjamin *et al*., 2019) was used for variant calling. Only variants with VAFs close to fixation (>0.9) were considered when building the consensus sequences.

These analyses were performed by Daniel García-Souto.

4.2.10. PHYLOGENETIC ANALYSES

Mitochondrial sequences for 13 coding genes and 2 rDNA genes from the 23 recovered mitogenomes (6 neoplastic, 17 from host and healthy specimens) were extracted from the paired-end sequencing data by mapping reads onto the previously reconstructed canonical mtDNAs for *V. verrucosa* and *C. gallina*, concatenated, and subjected to multiple alignment with MUSCLE v3.8.425 (Edgar, 2004).

The best-fit model of nucleotide substitution for each individual gene was selected using JModelTest2 (Darriba *et al*., 2012) and a partitioned Bayesian reconstruction of the phylogeny was performed with MrBayes v3.2.6 (Ronquist *et al*., 2012). Two independent Metropolis-coupled Markov Chain Monte Carlo (MCMC) analyses with four chains in each were performed. Each chain was run for 10 million generations, sampling trees every 1000 generations. Convergence of runs was assessed using Tracer (Rambaut *et al*., 2018).

DEAH12 and *TFIIH* sequences were subjected to multiple alignment using MUSCLE v3.8.425 (Edgar, 2004). Then, a 'species/population tree' was inferred with the starBEAST multispecies coalescent model, as implemented in BEAST v2.6.2 (Bouckaert *et al*., 2019). This analysis was performed using a Yule speciation prior and strict clock, with the best-fit model of nucleotide substitution obtained with jModelTest2 (Darriba *et al*., 2012) on both the concatenated mitochondrial haplotypes (13 protein-coding and 2 rRNAs genes) and unphased data from *DEAH12* and *TFIIH* nuclear fragments. The four mitochondrial groups observed on the mitogenome analysis (*V. verrucosa*, *C. gallina*, *C. striatula*, and tumour) were defined as tips for the species tree. A single MCMC of 10 million iterations, with sampling every 1000 steps, was run. A burn-in of 10% was implemented to obtain ESS values above 200 with Tracer v1.7.1 and the resulting posterior distributions of trees were checked with DENSITREE v2.1

(Bouckaert, 2010). A maximum clade credibility tree was obtained with TreeAnnotator (Bouckaert *et al*., 2019) to summarize information on topology, with 10% burn-in and Common Ancestors for the node heights.

These analyses were performed by Daniel García-Souto.

4.3. RESULTS AND DISCUSSION

4.3.1. LEUKAEMIA-LIKE CANCER IN WARTY VENUS CLAMS

To our knowledge, HN was not reported previously in warty venus clams although a cytogenetic study pointed it out (Carrilho Rodrigues da Silva, 2012). We investigated the prevalence of HN in the warty venus clam (*V. verrucosa*), a saltwater bivalve found in the Atlantic Coast of Europe and the Mediterranean Sea. We collected 345 clam specimens from six sampling regions in the Atlantic and the Mediterranean coasts of Europe across five different countries, including Spain, Portugal, France, Ireland, and Croatia (Figure 50A; Table 16).

Table 16. Sampling data of 571 specimens analysed in this study.

Country	Site	Code	Sampling coordinates	Samp. date	Sample size	Neoplastic specimens			
						N1	N2	N3	Total
Warty venus clam (*Venus verrucosa*)									
Spain	Ferrol**	EFVV	43.48325;-8.187209	Oct-17	30	0	0	0	0
	Ribeira**	ERVV	42.52870;-8.995799	Oct-17	30	1	0	2	3
	Vigo**	ECVV	42.19595;-8.80129	Oct-17	30	0	0	0	0
	Islas Cies**	EVVV	42.22258;-8.89383	Jun-11	1	*	*	*	1
	Mahon	EMVV	39.89029;-4.287749	Feb-18	67	2	1	2	5
France	Ganville	FGVV	48.85191;-1.694879	Jan-18	100	0	0	0	0
Croatia	Split	CSVV	43.54745;-16.33529	Apr-18	18	0	0	0	0
Portugal	Oeiras	PLVV	39.69316;-9.287104	Jul-18	19	0	0	0	0
Ireland	Carna	ICVV	53.30359;-9.863126	May-19	50	0	0	0	0
Striped venus clam (*Chamelea gallina*)									
Spain	Cadiz	ECCG	36.78597;-6.375165	Jun-20	50	0	0	0	0
	Huelva	ECCG	37.16459;-6.96620	Mar-15	5	0	0	0	0
	Mallorca	EMCG	Unknown	Jun-20	50	0	0	0	0
Italy	Naples	INCG	40.79773;-14.346828	Oct-20	50	0	0	0	0
	S. Benedetto	IMCG	42.91817;-13.90652	Jan-17	4	0	0	0	0
	Cattolica	IVCG	44.03711;-12.655139	Jun-20	7	0	0	0	0
	Cattolica	IVCG	44.03711;-12.655139	Dec-20	50	0	0	0	0
Striped venus clam (*Chamelea striulata*)									
Spain	Combarro	ECCS	42.432994;-8.689014	Jul-20	9	0	0	0	0
	Marin	EVCS	42.37626;-8.73773	Aug-13	1	0	0	0	0
Total:					**571**				**9**

* Hemic neoplasia stage was not determined because cytohistological examination was not possible in this individual, which was diagnosed by cytogenetics. ** Galicia region is the term used to agglomerate results from these three locations in this Chapter.

Cytohistological examination identified HN-like tumours in eight specimens from two sampling points in Spain (Figure 50B-C). Three HN-positive specimens (ERVV17-2995, ERVV17-2997, and ERVV17-3193) were collected in Galicia, northwest of the Iberian Peninsula in the Atlantic Ocean, and another five specimens (EMVV18-373, EMVV18-376, EMVV18-391, EMVV18-395, and EMVV18-400) were collected in Mahón, bathed by the Mediterranean Sea (Figure 50A).

Four of these specimens (ERVV17-2995, ERVV17-3193, EMVV18-391, and EMVV18-395) showed a severe form of the disease – classified as **N3 stage** – which is characterized by high levels of neoplastic cells infiltrating the gills, different levels of infiltration of the digestive gland and gonad, and low/very low infiltration of the mantle and foot (Figure 51, Appendix A: Supplementary material).

One specimen (EMVV18-400) was found that was affected with an intermediate form of the disease – **N2 stage** – characterized by low levels of neoplastic cells infiltrating the gill vessels, digestive gland, and gonad, but not the foot (Figure 51).

Three specimens (ERVV17-2997, EMVV18-373, and EMVV18-376) were diagnosed with a light form of the disease – **N1 stage** – characterized by low levels of neoplastic cells infiltrating the gills vessels only, and no infiltration in the remaining tissues (Figure 51, Appendix A: Supplementary material).

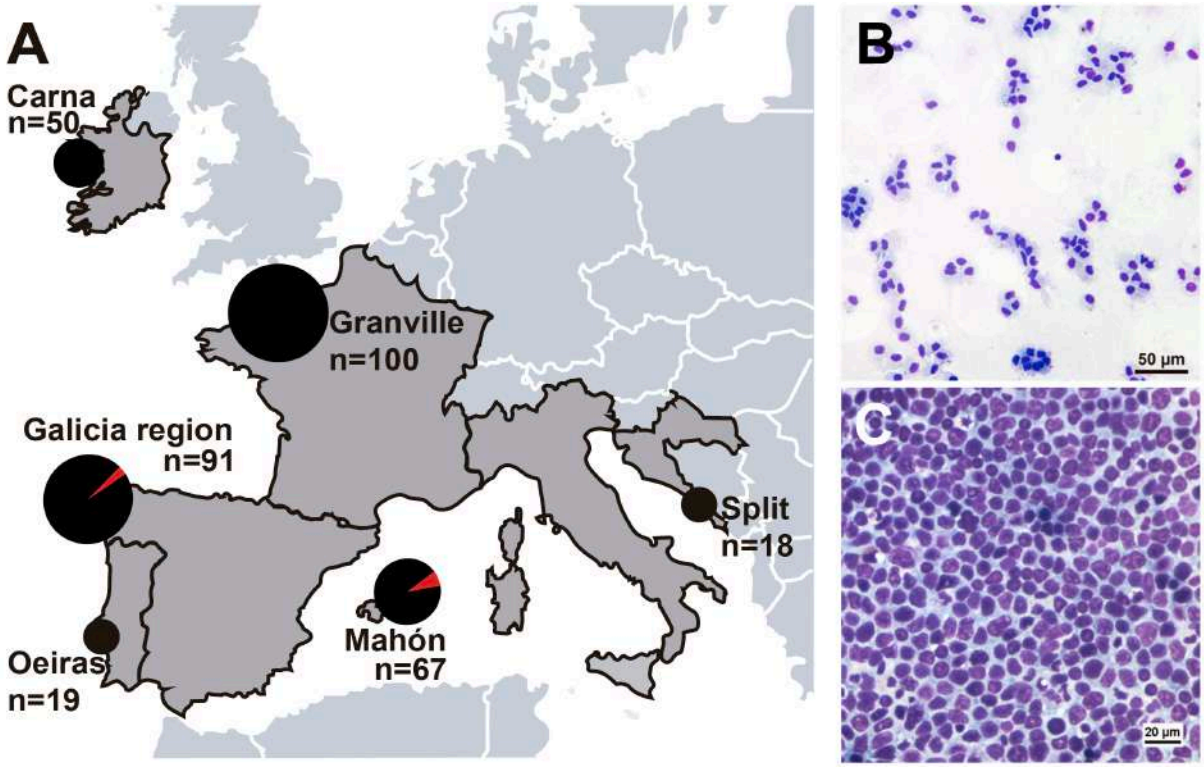


Figure 50. Diagnosis of HN in warty venus clams (adapted from García-Souto et al. 2022, reprinted with permission from eLife, CC-BY 4.0, see Appendix H). **(A)** Sampling map of warty venus clams collected for and specimens diagnosed with hemic neoplasia. Size of the pie charts correlates with the number of samples collected (number of samples 'n' is shown together with each pie chart). Pie charts show the proportion of samples with hemic neoplasia (black, no neoplastic specimens; red, neoplastic specimens). **(B)** Cytological examination of haemolymph smear from a healthy (N0) specimen, ERVV17-2963, shows normal haemocytes. **(C)** Haemolymph smear of a warty venus clam with high-grade (N3 stage) hemic neoplasia, ERVV17-3193, shows neoplastic cells that replaced normal haemocytes.

A statistically significant difference ($p < 0.001$, Table 17) was observed between the mean of nucleus-cytoplasm of healthy (Figure 50B) and neoplastic cells (Figure 50C); however, no statistically significant difference was observed on the mean of cell diameter between healthy and neoplastic cells (Table 17).

Nuclei morphology of neoplastic cells is usually circular, oval or kidney-shaped and no significant differences ($p < 0.001$) were detected between neoplastic cells from different populations in terms of cell diameter (that correlates with size) or nucleus-cytoplasm ratio.

Table 17. Cell measurements from preparations of non-cancer haemolymph and cancer haemolymph.

	Neoplastic cells	Haemocytes
Cell diameter	7,18 μm ± 1,21	6,78 μm ± 1,16
Nucleus-cytoplasm ratio	0,85 μm ± 0,26	0,76 μm ± 0,10

HN individuals were normally in the gametogenic cycle phase of post-laying or reabsorption of the gonad and both sexes have been found (4 males, 3 females) suggesting that HN on this species is not related to the sex of the individual. However, largest numbers should be screened for cancer to draw conclusions.

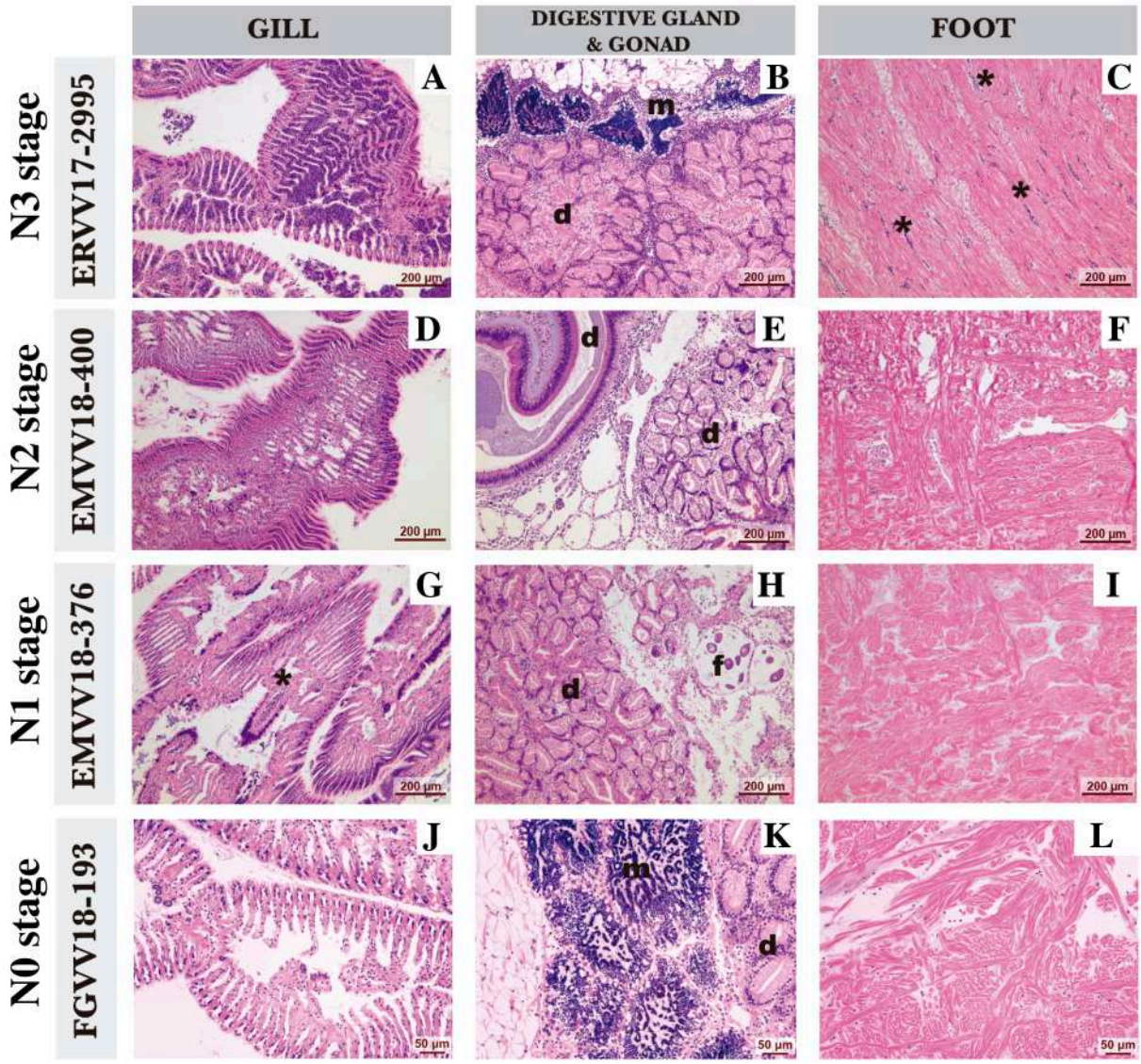


Figure 51. Histological diagnosis of hemic neoplasia in warty venus (*V. verrucosa*) specimens. Hematoxylin and eosin-stained photomicrographs of gill, digestive (d), gonad (male (m) & female (f)) and foot of warty venus specimens diagnosed with different stages of hemic neoplasia: high (N3), medium (N2), light (N1) and non neoplastic (N0). In the N3 stage, neoplastic cells infiltrate the connective tissue and vessels of different organs (A,B), and show low infiltration of foot (C). In N2 stage, cell groups are observed in different organs such as gills (D) and are not detected in the foot (F). In N1 stage, groups of neoplastic or isolated cells are detected in gill sinuses (G). N0 stage is completely devoid of any trace of hemic neoplasia at either gill, digestive gland and gonad and foot (J,K,L). Asterisks show groups of neoplastic cells. (Adapted from García-Souto et al. 2022, reprinted with permission from eLife, CC-BY 4.0, see Appendix H).

Electron microscopy analysis through gill's ultrathin sections from two neoplastic warty venus specimens (ERVV17-2995 and ERVV17-3193) revealed tumour cells with a round shape and a pleomorphic nucleus, which are morphological features that generally characterize bivalves' HN (Figure 52).

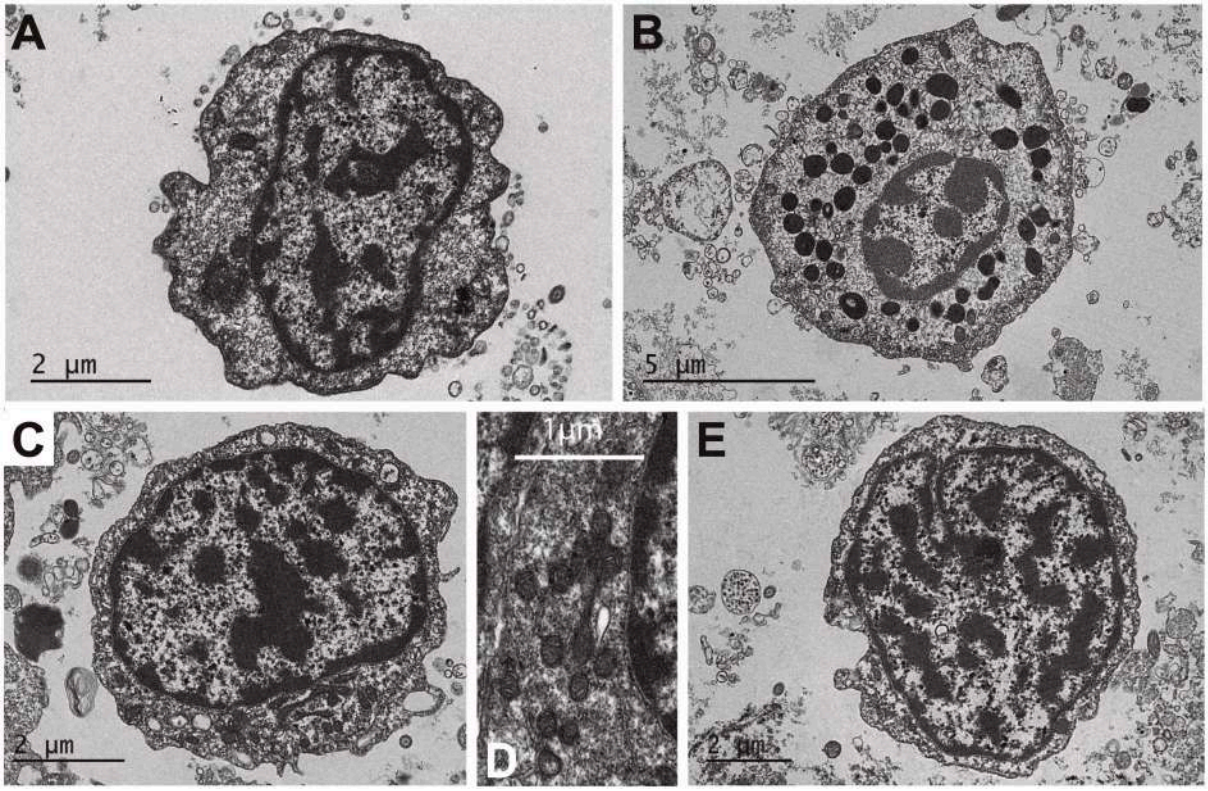


Figure 52. Transmission electron microscopy (TEM) of healthy and neoplastic *V. verrucosa* clams (Source: García-Souto et al. 2022, reprinted with permission from eLife, CC-BY 4.0, see Appendix H). **(A)** Hyalinocyte of healthy warty venus specimen ERVV17-2992. **(B)** Granulocyte of healthy warty venus specimen ERVV17-2993. **(C)** Neoplastic cell of specimen ERVV17-2995. **(D)** Mitochondrias in detail of neoplastic cell of specimen ERVV17-3193. **(E)** Neoplastic cell of specimen ERVV17-3193.

Ploidy analysis revealed stability in the DNA content of all neoplastic cells being most of them triploids when compared to diploid normal haemocytes (Figure 53). Furthermore, haemolymph from severe-affected clams (*i.e.*, N3 stage) showed higher percentages of triploid cells while early-stage clams (*i.e.*, N1 stage) showed less, which supports that flow cytometry would be a good diagnostic method for HN in the warty venus clam.

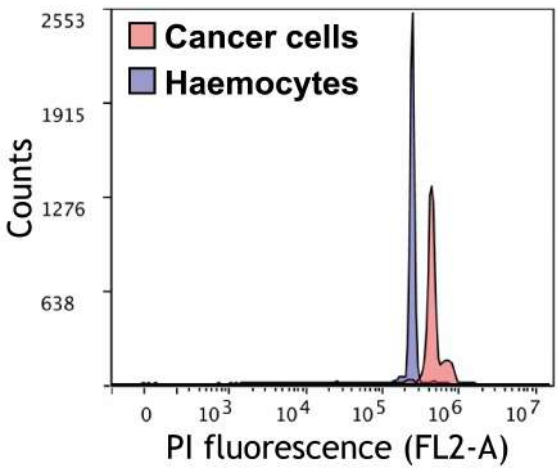


Figure 53. Ploidy analysis by flow cytometry of cancer and non-cancer cells. Histogram of DNA content showing a 2n peak of fluorescence (blue) that was with fluorescence mean of haemocytes from 25 non-neoplastic individuals (266.77± 5.64 fluorescence units with a variation coefficient range between 6-18%). Ploidy of neoplastic cells (red) showed stable values close to triploids (3.3n; 3.4n; 3.5n) with coefficients of variation between 7 and 18%. Severe affected individuals showed peaks of 81-85% of triploid cells while early-stage individuals showed 4.89% of circulating triploid cells (Courtesy of Seila Díaz).

Finally, one additional neoplastic warty venus specimen (EVVV11-02) was included in the study. The animal, which was sampled in 2011 in Galicia and came from a private collection, showed abnormal metaphases in the gills that were suggestive of HN. Although the species typically shows a 2n = 38 karyotype with metacentric chromosomes that are homogeneous in size (García-Souto *et al*., 2015), the tumoral metaphases from this individual showed around 100 chromosomes that were variable in size and shape (Figure 54).

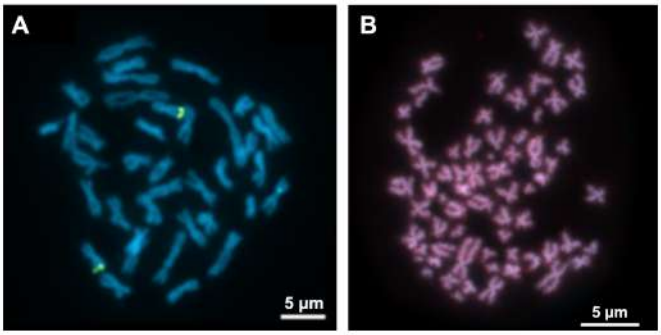


Figure 54. Chromosomes of healthy and tumoral cells of warty venus clams. **(A)** Mitotic chromosomes of warty venus clam (*V. verrucosa*) with the H3 histone gene probe mapped by FISH (green). Adapted from García-Souto *et al.*, 2015; copyright 2015, Springer Nature. **(B)** Metaphase chromosomes from a neoplastic cell found in the gills of the *V. verrucosa* specimen EVVV11-02, showing abnormal chromosome number (>19 pairs) and abnormal chromosome morphology. Chromosomes stained with 4′,6-DiAmidino-2-PhenylIndole (DAPI) and Propidium Iodide (PI), (Source: García-Souto et al. 2022, reprinted with permission from eLife, CC-BY 4.0, see Appendix H, generated by D. García-Souto).

4.3.2. MITOCHONDRIAL SEQUENCING REVEALS CANCER CONTAGION

To obtain some biological insights into the clonal dynamics of this cancer, we carried out whole-genome sequencing with Illumina paired-ends in DNA samples isolated from the tumoral haemolymph from eight out of nine neoplastic specimens mentioned above. Their feet were also sequenced, as foot typically represents the tissue with lower infiltration of neoplastic cells, making it a good candidate tissue to act as 'matched-normal' (*i.e*. host tissue). As for the animal with an abnormal karyotype (EVVV11-02) that was compatible with HN, we sequenced the only tissue available, which were gills (Table 18).

Table 18. Clam specimens and tissues sequenced. Sixteen specimens (eight neoplastic and eight non-neoplastic) from three different clam species (*V. verrucosa, C. gallina*, and *C. striatula*) were sequenced with Illumina paired-end reads. Columns 4 and 5 show the number of reads generated for the host tissue (when neoplastic, matched-normal tissue was foot) and the tumoral haemolymph, respectively.

Specimen origin	Specimen code	Diagnosis	Foot reads	Haemolymph reads
Warty venus clam (*Venus verrucosa*)				
Ribeira, Spain	ERVV17-2995	N3	833 M	919 M
Ribeira, Spain	ERVV17-2997	N1	766 M	598 M
Ribeira, Spain	ERVV17-3193	N3	739 M	850 M
Mahón, Spain	EMVV18-376	N1	784 M	849 M
Mahón, Spain	EMVV18-391	N3	617 M	623 M
Mahón, Spain	EMVV18-395	N3	697 M	679 M
Mahón, Spain	EMVV18-400	N1	782 M	1133 M
Vigo, Spain	EVVV11-02	N#	743 M*	-*
Split, Croatia	CSVV18-1052	Healthy	161 M	-
Mahón, Spain	EMVV18-385	Healthy	143 M	-
Granville, France	FGVV18-183	Healthy	752 M	-
Carna, Ireland	IGVV19-666	Healthy	155 M	-
Oeiras, Portugal	PLVV18-2249	Healthy	163 M	-
Striped venus clam (*Chamelea gallina*)				
S.Benedetto, Italy	IMCG15-69	Healthy	147 M	-
Cadiz, Spain	ECCG15-201	Healthy	752 M	-
Striped venus clam (*Chamelea striulata*)				
Vigo, Spain	EVCS14-09	Healthy	706 M	-

* The only available tissue from this neoplastic animal, collected in 2011, were gills. # Hemic neoplasia stage was not determined because cytohistological examination was not possible in this individual, which was diagnosed by cytogenetics.

Only one neoplastic specimen (EMVV18-373) that had a very low proportion of tumour cells in its haemolymph was excluded from the sequencing. Then, we mapped the paired-end reads onto a dataset containing non-redundant mitochondrial Cytochrome C Oxidase subunit 1 (*mt-COI*) gene references from 118 Venerid clam species. In six out of eight sequenced neoplastic specimens, the results revealed an overrepresentation (>99%) of reads in the sequenced tissues mapping to *mt-COI* DNA sequences that exclusively identified two different clam species (Figure 55A): the expected one, warty venus clam (*V. verrucosa*), and a second, unexpected one, the striped venus (*C. gallina*), a clam that inhabits the Mediterranean Sea (Figure 55B).

Preliminary analysis by PCR and capillary sequencing of *mt-COI* in the haemolymph of two neoplastic specimens, EMVV18-373 and EVVV11-02, revealed an electropherogram with overlapping peaks apparently containing two different haplotypes that match the reference *mt-COI* sequences for warty and striped venus (Figure 55B).

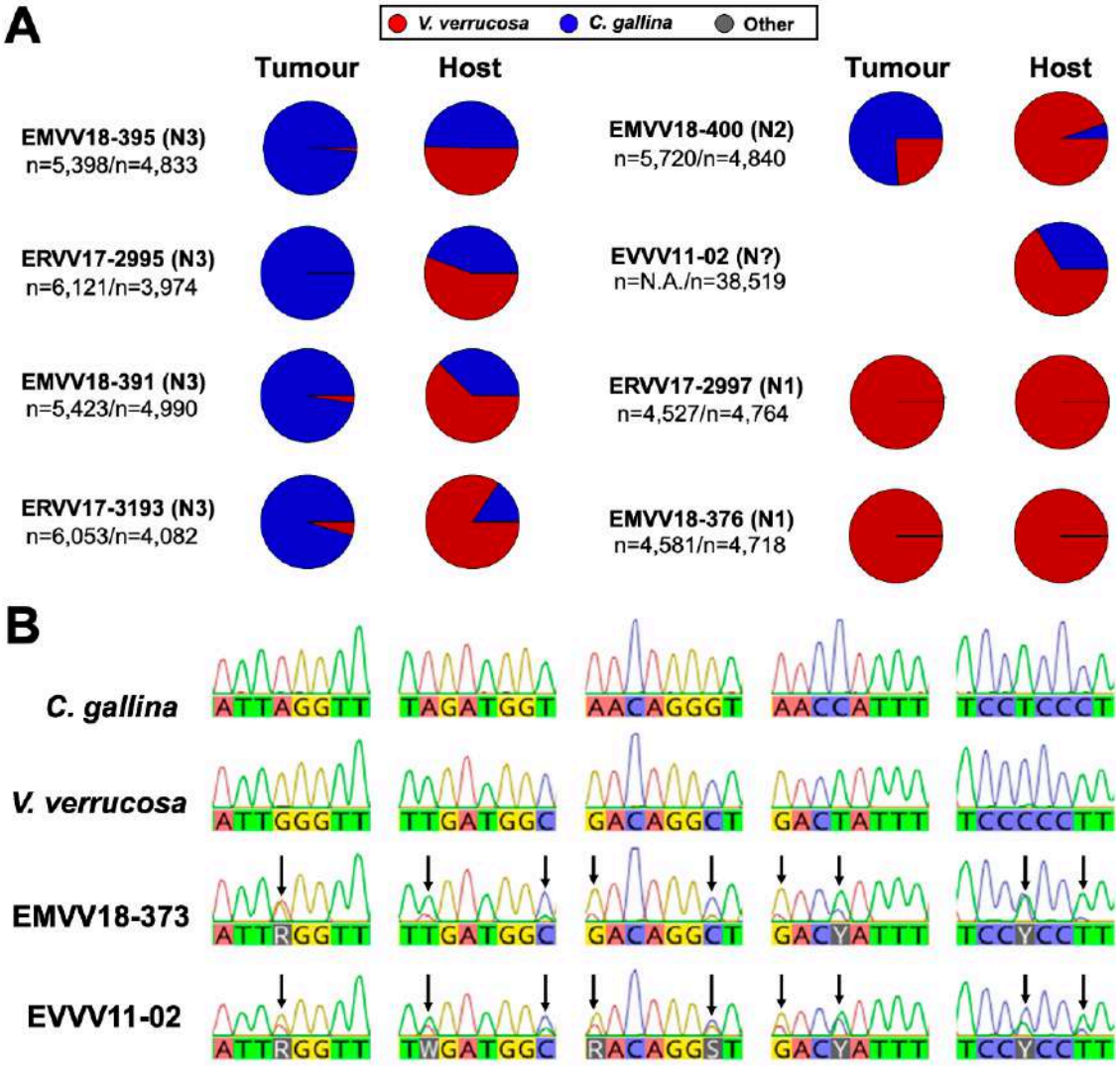


Figure 55. Mitochondrial *mt-COI* gene sequencing reveals cancer contagion between warty venus (*V. verrucosa*) and striped venus (*C. gallina*) clam species (Source: García-Souto et al. 2022, reprinted with permission from eLife, CC-BY 4.0, see Appendix H). **(A)** In eight warty venus specimens sequenced with Illumina paired-end reads, the pie charts show the proportion of reads mapping *mt-COI* reference sequences from 137 different Verenidae species, including *V. verrucosa* (red), *C. gallina* (blue), and the remaining

species (grey). Two different tissues were sequenced: the tumour tissue, typically haemolymph, and the host/matched-normal tissue, typically foot. Note that for specimen EVVV11-02 only the host/matched-normal tissue (gills) was available. 'n' denotes the total number of reads mapping the *mt-COI* reference for the tumour tissue (left), and the host tissue (right). **(B)** Capillary sequencing electropherograms of mitochondrial *mt-COI* gene fragments from two neoplastic *V. verrucosa* specimens (EMVV18-373 and EVVV11-02) and two healthy reference specimens from *V. verrucosa* and *C. gallina*. The results show overlapping peaks (arrows) in the sequenced tissues from the neoplastic animals, which suggest coexistence of mitochondrial DNA (mtDNA) haplotypes from two clam species (lab work done by S. Díaz and figure generated by D. García-Souto).

These results suggested cancer contagion between the two clam species of the family Veneridae. Hence, to decipher the origins of this clam neoplasia, we further analysed the mitochondrial DNA (mtDNA) from the two species involved and the tumours. Firstly, we performed multiplatform genome sequencing, including Illumina short reads and Oxford Nanopore long reads, on canonical individuals from the two species to obtain a preliminary assembly of the mitogenomes of *V. verrucosa* and *C. gallina*. These reconstructions resulted in 18,092- and 17,618-bp long mtDNA genomes for the warty venus and the striped venus clam, respectively (Figure 56). The comparative analysis of the nucleotide sequences from both mitogenomes confirms that, although both species are relatively close within the subfamily Venerinae (Canapa *et al*., 1996), they represent distinct sister species, showing a Kimura's two-parameter nucleotide distance (K2P) equal to 21.13%.

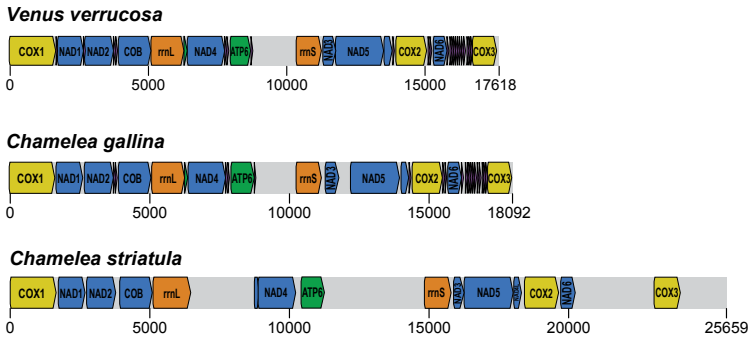
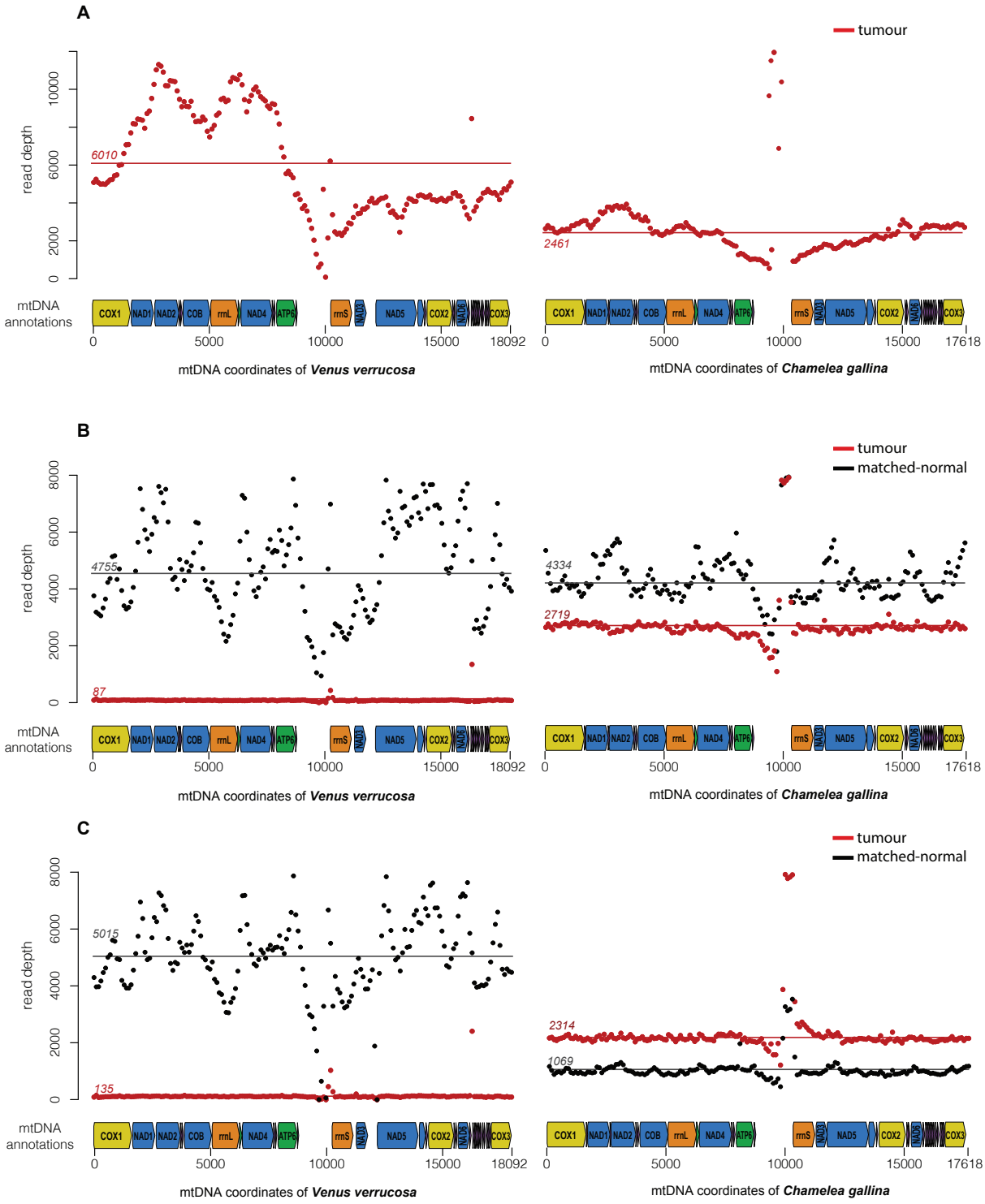


Figure 56. Draft reference mitochondrial DNA (mtDNA) genome assemblies reconstructed for *V. verrucosa*, *C. gallina*, and *C. striatula* (Source: García-Souto et al. 2022, reprinted with permission from eLife, CC-BY 4.0, see Appendix H).

Then, we mapped the paired-end sequencing data from the six neoplastic specimens with evidence of interspecies cancer transmission onto the two reconstructed species-specific mtDNA genomes. This approach confirmed the coexistence of two different mtDNA haplotypes in the six examined neoplastic samples, matching the canonical mtDNA genomes from the two clam species. For example, in a N2-stage specimen (EMVV18-400), this analysis revealed different proportion of tumour and host mtDNA molecules in the two tissue types sequenced (Figure 57F). Here, the striped venus mtDNA results the most abundant in the haemolymph, in which tumour cells are dominant over the remaining cell types, and the lower in the matched-normal tissue (*i.e*. infiltrated foot), where tumour cells represent a minor fraction of the total. Similar results were obtained for the remaining five neoplastic individuals (Figure 57A-E).


A
tumour
read depth
0
2000
6000
10000
6010
2461
mtDNA annotations
COX1
NAD1
NAD2
COB
rrnL
NAD4
ATP6
rrnS
NAD3
NAD5
COX2
NAD6
COX3
0
5000
10000
15000
18092
17618
mtDNA coordinates of Venus verrucosa
mtDNA coordinates of Chamelea gallina
B
tumour
matched-normal
read depth
0
2000
4000
6000
8000
4755
87
4334
2719
mtDNA annotations
mtDNA coordinates of Venus verrucosa
mtDNA coordinates of Chamelea gallina
C
tumour
matched-normal
read depth
0
2000
4000
6000
8000
5015
135
2314
1069
mtDNA annotations
mtDNA coordinates of Venus verrucosa
mtDNA coordinates of Chamelea gallina

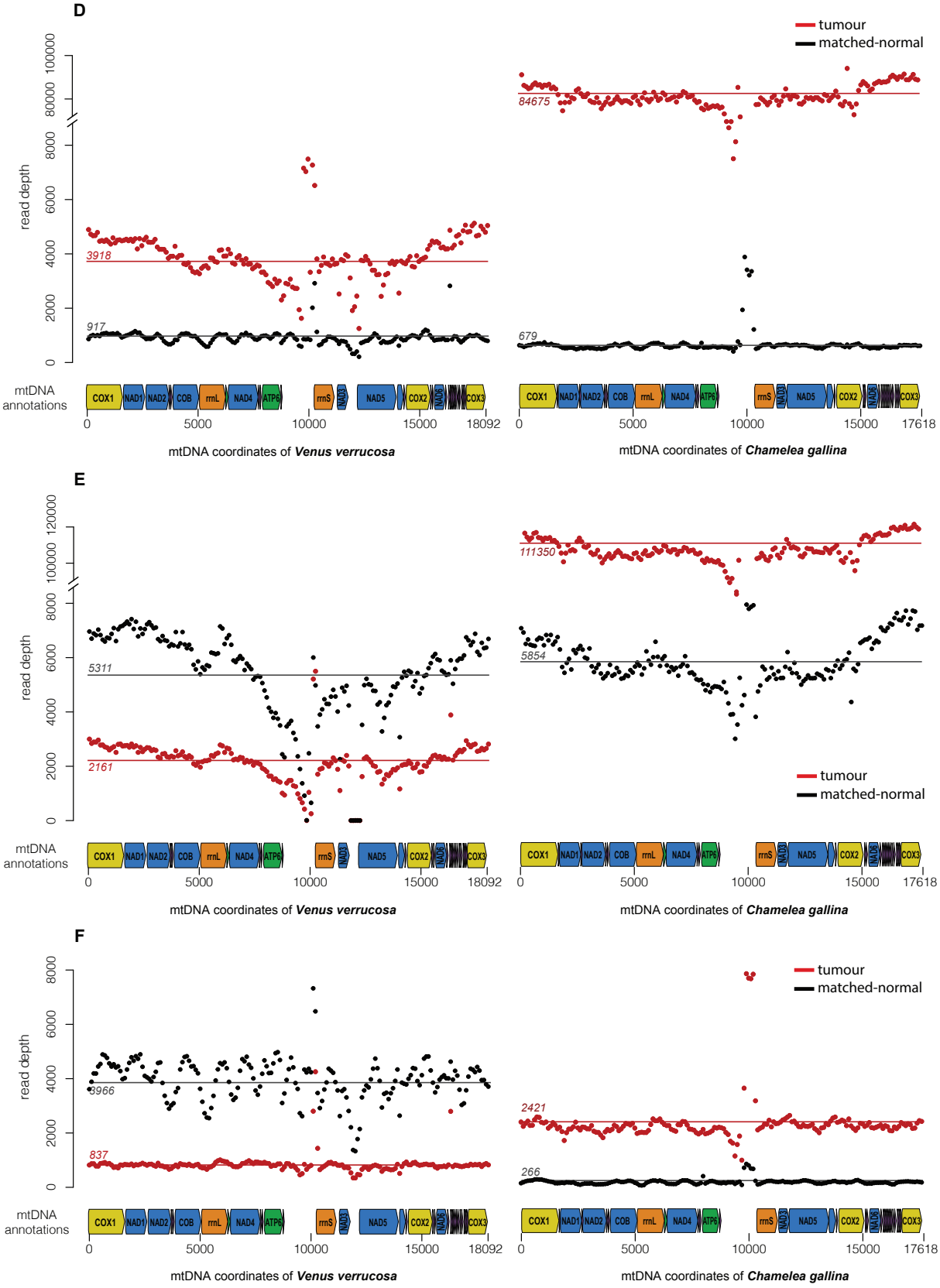


Figure 57. Comparison of read coverage in two mitochondrial genomes of two tissues from warty venus neoplastic specimens (Source: García-Souto et al. 2022, reprinted with permission from eLife, CC-BY 4.0, see Appendix H). **(A)** Old warty venus neoplastic specimen EVVV11-02, only one tissue available. **(B)** Warty venus neoplastic (N3-stage) specimen ERVV17-2995, tumour tissue (red) shows a higher representation of coverage in the mtDNA of *C.*

gallina which is the cancer founder of tumor cells. **(C)** Warty venus neoplastic (N3-stage) specimen ERVV17-3193, tissues show an opposite pattern: tumor tissue (red) has more coverage in *C. gallina* mtDNA which is the cancer founder of tumor cells while matched-normal tissue (black) shows a higher coverage in the mtDNA of *V. verrucosa* which is the host species of this cancer. **(D)** Warty venus neoplastic (N3-stage) specimen EMVV18-391, tumour tissue (red) shows a desproportional higher representation of coverage in the mtDNA of *C. gallina*. **(E)** Warty venus neoplastic (N3-stage) specimen EMVV18-395, tumour tissue (red) shows a desproportional higher representation of coverage in the mtDNA of *C. gallina*. **(F)** Warty venus neoplastic (N2-stage) specimen EMVV18-400.

To further investigate the evolutionary origins and geographic spread of this cancer, we sequenced with Illumina paired-end reads an additional set of eight healthy (*i.e.* non-neoplastic) clams from three different Veneridae species, including five more warty venus specimens (EMVV18-385, IGVV19-666, FGVV18-183, CSVV18-1052, and PLVV18-2249) from five different countries, two striped venus specimens (IMCG15-69 and ECCG15-201) from two countries, and one specimen (EVCS14-09) from its sibling species *Chamelea striatula*, a type of striped venus clam that inhabits the Atlantic Ocean from Norway to the Gulf of Cadiz in Spain. This made a total of 16 Veneridae specimens sequenced, all listed in Table 18.

The complete mitochondrial genomes from all tumoural and healthy *V. verrucosa* specimens (13 individuals), 2 *C. gallina*, and 1 from its sibling species *C. striatula*, were individually de novo assembled from the sequencing reads. As expected, this approach reconstructed two different haplotypes in six out eight sequenced neoplastic animals, supporting the presence of mtDNA from two different species. Despite the high sequencing coverage obtained for these individuals (Table 18), we did not find foreign reads in the N1 tumours (ERVV17-2997 and EMVV18-373), most likely due to a low proportion of neoplastic cells in the haemolymph and the matched-normal tissue. Then, we performed a phylogenetic analysis based on the alignment of these mitochondrial genomes (13 coding and 2 RNA gene sequences, altogether encompassing ~14 kb). The results (Figure 58) show that tumour and non-tumour sequences from neoplastic warty venus specimens define two well-differentiated clades, and that tumoral warty venus sequences are all identical and closer to striped venus mtDNA than to its own (warty venus).

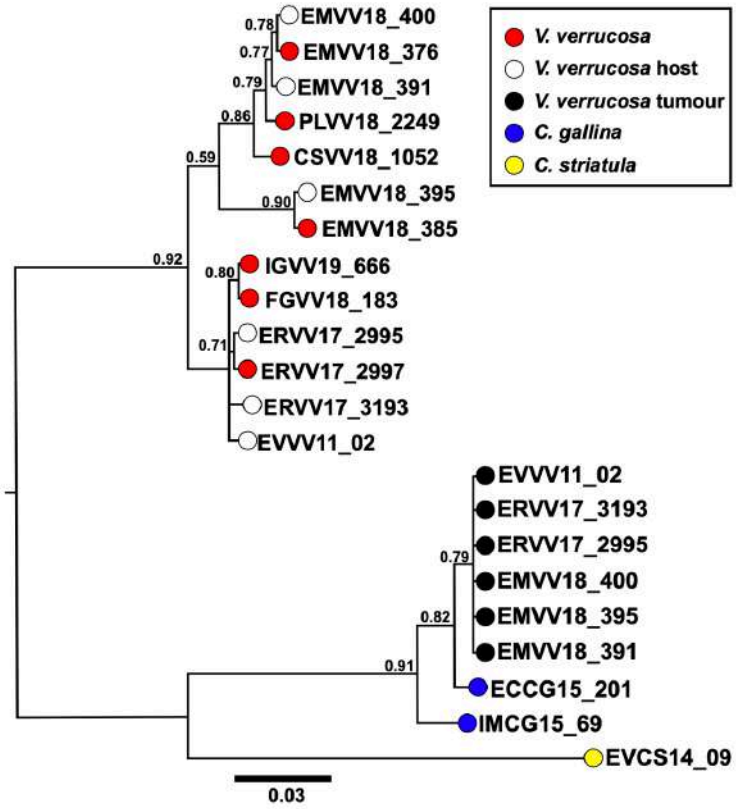


Figure 58. Molecular phylogeny using Bayesian inference inferred on the alignment of all mitochondria coding genes and rRNA gene sequences (15 loci) that includes six neoplastic *V. verrucosa* specimens with evidence of cancer contagion from *C. gallina*. Bootstrap values are shown above the branches (Source: García-Souto et al. 2022, reprinted with permission from eLife, CC-BY 4.0, see Appendix H, generated by D. García-Souto).

Overall, these data support the existence of a single cancer clone originated in the striped venus clam *C. gallina* that was transmitted to *V. verrucosa*.

4.3.2. NUCLEAR EVIDENCE OF CANCER CONTAGION

Transmissible cancers are known to occasionally acquire mitochondria from transient hosts (Strakova *et al*., 2016, 2020), which can lead to misinterpretation of their evolutionary history. Thus, we looked for nuclear markers to confirm the striped venus origin of this cancer lineage. We performed a preliminary draft assembly of the warty venus and the striped venus nuclear 'reference' genomes, using the paired-end sequencing data from two non-neoplastic animals. Then, we used bioinformatic approaches to find single copy nuclear genes that were homologous between the two species, identifying two confident candidate genomic regions:

- A 2.9-kb long region from *DEAH12*, a gene that encodes for an ATP-dependent RNA helicase.
- A 2.2-kb long fragment from the Transcription Factor II Human-like gene, *TFIIH*.

With the idea of finding differentially fixed single-nucleotide variants (SNVs) between both species, we performed PCR amplification and capillary sequencing on a 441 bp fragment from the *DEAH12*, and a 559 bp fragment from *TFIIH*, in 2 cohorts of non-neoplastic warty venus specimens (12 for *DEAH12* and 15 for *TFIIH*), 2 cohorts of non-neoplastic striped venus (9 for *DEAH12* and 12 for *TFIIH*), and 1 specimen of its sister species *C. striatula*. This analysis provided 14 and 15 sites, respectively, for the *DEAH12* and the *TFIIH* loci, with fixed SNVs (allele frequency >95%) that allowed to discriminate between the 3 relevant species and the tumour (Figure 59).

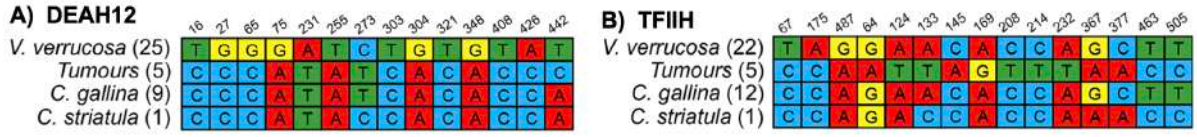


Figure 59. Single-nucleotide variants discriminating between *V. verrucosa* tumours and the three canonical species (*V. verrucosa*, *C. gallina*, and *C. striatula*) along a 441- and a 559-bp long fragments of nuclear genes *DEAH12* and *TFIIH*, respectively (Source: García-Souto et al. 2022, reprinted with permission from eLife, CC-BY 4.0, see Appendix H, generated by D. García-Souto).

These variants were employed to identify the Illumina reads from each sequenced warty venus neoplastic specimens that were specific for either warty venus or striped venus, which allowed to obtain the consensus sequences that corresponded to the tumour tissue and the non-affected tissue from each neoplastic individual. At the end of this process, we performed Maximum Likelihood phylogenetic reconstructions from these individual nuclear consensus sequences. On the one hand, the phylogeny for the *DEAH12* locus confirmed both the monophyly of the tumoral sequences and their closer relationship to *C. gallina* than to the host species (Figure 60A), which were also observed in the mtDNA analysis. However, the phylogeny derived from the *TFIIH* locus showed that, although the tumours remained monophyletic, they were positioned in a basal branch relative to *C. gallina* and *V. verrucosa* (Figure 60B).

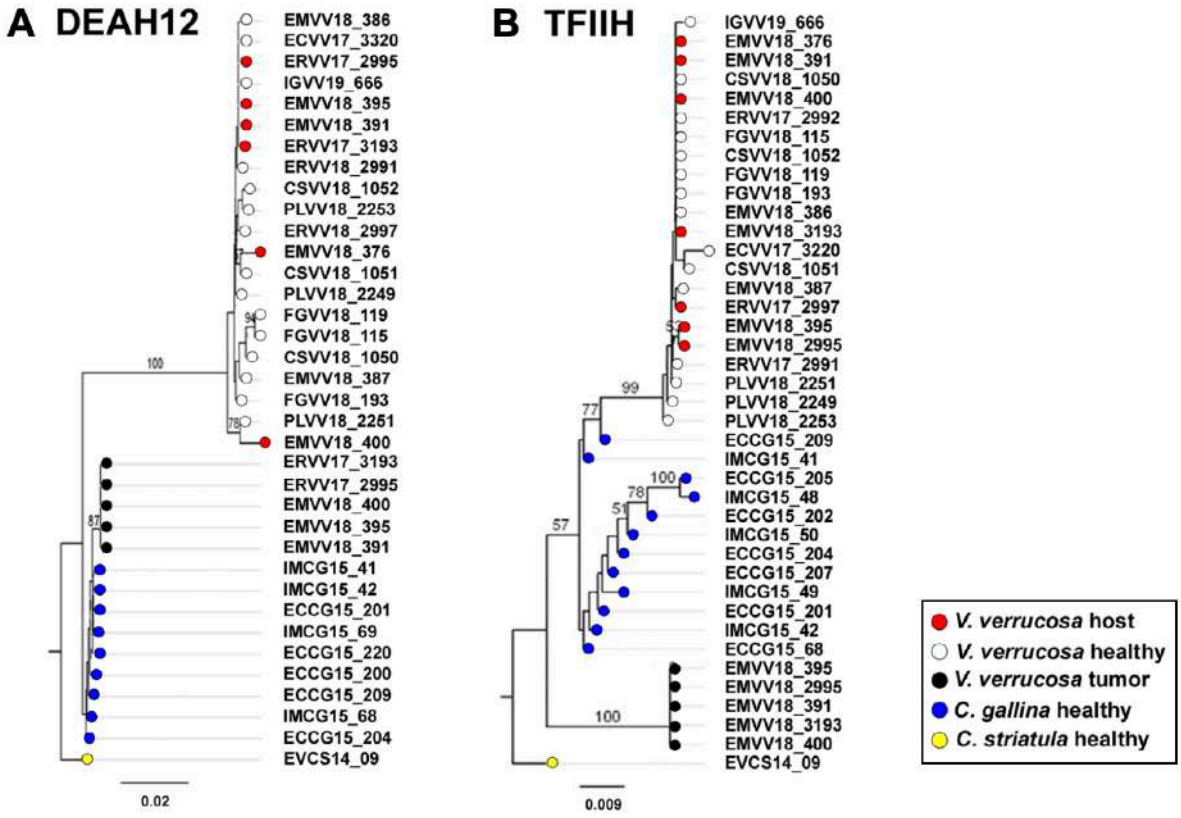


Figure 60. Maximum Likelihood molecular phylogenies based on the two fragments of the nuclear DNA markers DEAH12 and TFIIH. Bootstrap support values (500 replicates) from Maximum Likelihood analyses above 50 are shown on the corresponding branches (Source: García-Souto et al. 2022, reprinted with permission from eLife, CC-BY 4.0, see Appendix H, generated by D. García-Souto).

Hence, to resolve these differences we also obtained a multilocus species tree based on the alignment of both the mtDNA and the two nuclear genes. This new phylogeny confirmed that warty venus tumours are closer to striped venus specimens than to non-neoplastic warty venus sequences from the same diseased specimens, while the non-neoplastic sequences conformed a more distant warty venus lineage (Figure 61).

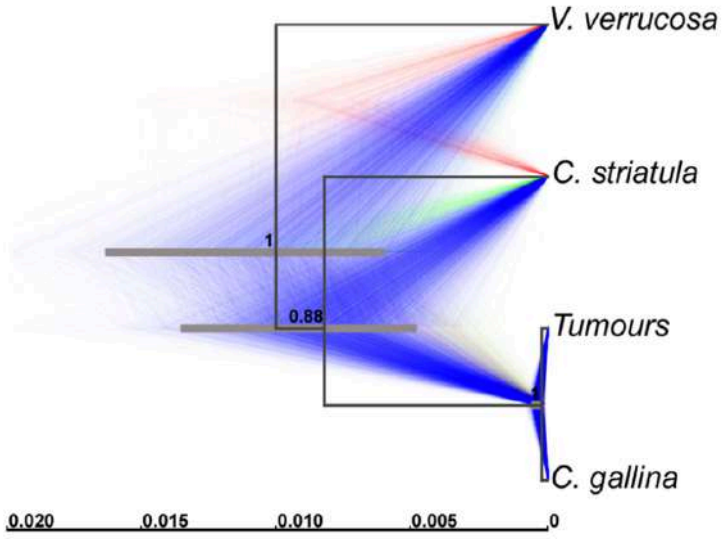


Figure 61. Multispecies coalescent (MSC) tree of *V. verrucosa*, their tumours and *Chamelea* sp. based on the entire mitochondrial DNA (mtDNA) and the two nuclear markers, *DEAH12* and *TFIIH*. A maximum clade credibility (MCC) tree is shown, with posterior probabilities below the branches, and 95% highest probability density (HPD) intervals of node heights as grey bars. The trees distribution shown includes 1000 trees and represents the range of alternative topologies, in which blue is the most common set of topologies, red the second most common one, and green the remaining (Source: García-Souto et al. 2022, reprinted with permission from eLife, CC-BY 4.0, see Appendix H, generated by D. García-Souto).

To obtain further evidence on the striped venus origin of this clam's neoplasia, we performed a comparative screening of tandem repeats in the genomes of *C. gallina* and *V. verrucosa* using fluorescence in situ hybridization (FISH). We focused on two satellite DNA repeats, namely CL4 and CL17. The satellites represent repeats of 332- and 429-bp long monomers, respectively, and were identified in a preliminary bioinformatics screening of the striped venus reference genome. This FISH approach revealed that the mentioned repeats are very abundant in heterochromatic regions from the genomes of the canonical striped venus and the neoplastic warty venus specimens tested (Figure 62). However, the repeats were absent in the metaphases from all the healthy warty venus individuals.

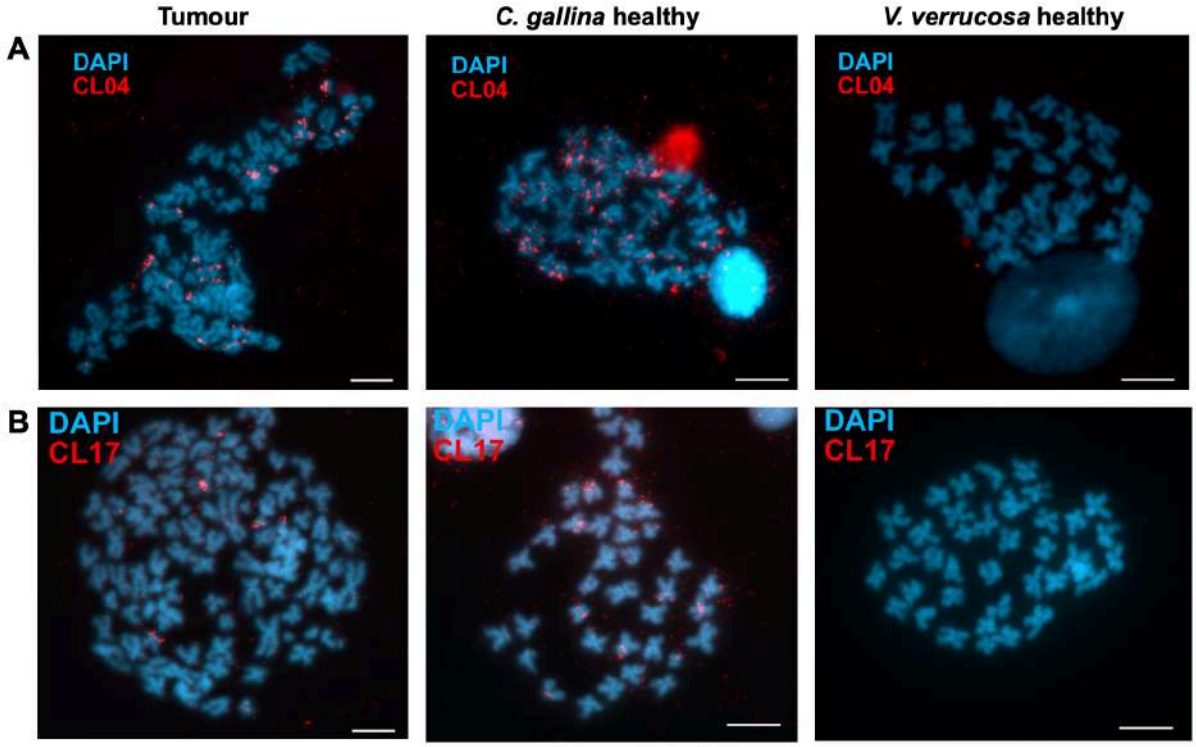


Figure 62. Fluorescence in situ hybridization (FISH) to specifically detect the satellite DNA **(A)** CL4 and **(B)** CL17 in one *V. verrucosa* tumour and healthy specimens from the species *C. gallina* and *V. verrucosa* shows probes accumulate in heterochromatic regions, mainly in subcentromeric and subtelomeric positions, from the chromosomes of the tumour and the healthy *C. gallina* tested but not in healthy *V. verrucose* (Source: García-Souto et al. 2022, reprinted with permission from eLife, CC-BY 4.0, see Appendix H, generated by D. García-Souto).

These results suggest that the relevant chromosomes with CL4 and CL17 satellites found in neoplastic warty venus specimens derive from *C. gallina*, supporting that a tumour originated in *C. gallina* was transmitted to *V. verrucosa*.

3.3.2. CANCER INSPECTION IN THE ORIGIN SPECIES (*C. gallina*)

Both mitochondrial and nuclear DNA suggest that this cancer originated in *C. gallina*. To find out whether this cancer is present in the clam species where it first arose, we performed a screening for its presence in natural populations of striped venus clams from the species *C. gallina* (n = 213) and *C. striatula* (n = 9) at five additional sampling points across two countries (Table 18), including Spain (n = 115) and Italy (n = 107). Histological analyses did not show any traces of HN in these specimens.

The virtual absence of this tumour in natural populations of striped venus clams may suggest that today this leukaemia is being mainly, if not exclusively, transmitted between specimens of the recipient species, warty venus. However, further sampling in other regions across the striped venus area of distribution may be necessary to confirm these findings.

4.3.4. A DECADE SPREADING ON SOUTHERN EUROPEAN SEAS

Overall, the results provided here reveal the existence of a transmissible leukaemia originated in a striped venus clam, most likely *C. gallina*, which was transmitted to a second species, the warty venus clam (*V. verrucosa*), and among whose specimens it currently propagates.

We identified this parasitic cancer in warty venus clams from two sampling points that are more than 1000 nautical miles away in the coasts of Spain, bathed by two different seas, the Atlantic Ocean and the Mediterranean Sea. The analysis of mitochondrial and nuclear gene sequences revealed no nucleotide diversity within the seven tumours sequenced, which supports that all belong to the same neoplastic lineage that spreads between Veneridae clams in the Seas of Southern Europe. Although we ignore the age of this cancer clone, we can confirm it arose before 2011, when the neoplastic warty venus specimen EVVV11-02 was collected. The apparent lack of genetic variation between all tumours, even from distant sampling points, suggests either that this cancer is very recent, or that it may have been unintentionally scattered by the action of man, a way of transmission that has been proposed for other bivalve transmissible cancers (Yonemitsu *et al*., 2019).

Chapter cover shows the illustration created by Sofía Venzel for the initiative *Scientists Meet Artists* of Campus do Mar from Universidade de Vigo (Spain). Campus do Mar and the artist have granted written permission to reproduce the drawing in this thesis.

Chapter 5.
General discussion on the evolution of bivalve transmissible cancers

"A ship in port is safe, but that's not what ships are built for." Grace Hopper

"Research is to see what everybody else has seen, and to think what nobody else has thought." Albert Szent-Gyorgyi

Evolution comes from the Latin word that refers to "unrolling a papyrus scroll" but the modern and biological sense of evolution refers to the changes in the gene pool of a population from generation to generation. The first-time evolution with this modern sense appeared was in 1832 in the works of the British geologist Charles Lyell in a discussion of some invertebrate sea creatures (Dictionary.com, 2022). Almost 200 years later, we will continue a discussion about the evolution of cancer cells that infect invertebrate sea creatures known as bivalves.

In this doctoral thesis, we have aimed to study the evolution of bivalve transmissible cancers. From the cell-of-origin of two independent contagious cancer lineages to uncovering nine independent mitochondrial captures, to the characterization of the current clonal structure of them by means of histo-cytological and genomic approaches to end with the discovery of a novel contagious cancer that jumps from one species to another in clams inhabiting the seas of southern Europe.

5.1.AN OVERVIEW OF FINDINGS ON BIVALVE TRANSMISSIBLE CANCERS

The first report of cancer contagion affecting a bivalve species is from 2015 and seven years later, the list of bivalve transmissible cancer lineages has notoriously increased to eight spreading among nine different bivalve species (Table 1, Metzger *et al*., 2015, 2016; Yonemitsu *et al*., 2019; Garcia-Souto *et al*., 2021; Hammel *et al*., 2021; M. Skazina *et al*., 2021; Michnowska *et al*., 2022), one of them described for the first time in the article included within the pages of this thesis. However, the main focus was on the study of the evolutionary history of cockle contagious cancers to shed light to the genetic causes of cancer transmissibility. Cockles were chosen as a model because back in 2016 – when this thesis was started – cockles were the only species with more than one cancer lineage spreading among its populations (Metzger *et al*., 2016) which opened the window to investigate the genetic causes of transmissibility by getting the common genetic alterations of those two cancer lineages.

5.1.1. INSIGHTS INTO THE GENETIC HISTORY OF COCKLE TRANSMISSIBLE CANCERS

As an analogy to archaeology –the study of human activity through the recovery and analysis of death and material culture–, genomics allows us to do the molecular archaeology of cancer. The cancer genome contains an archaeological record of its past and a cancer's life history can be extracted from sequencing data (Alexandrov *et al.*, 2013). Therefore, we sequenced 595 tumours and healthy cockles to build a large-scale dataset that would give profound insights into cancer evolution and transmissibility. In general, studies using high-throughput sequencing in basic biology have tended to use smaller sample sizes due to its cost. However, limited sampling can greatly impact population genetic inferences (Meirmans, 2015) which is why we have sequenced an extensive sampling collection of healthy cockles to better capture the variability of the population.

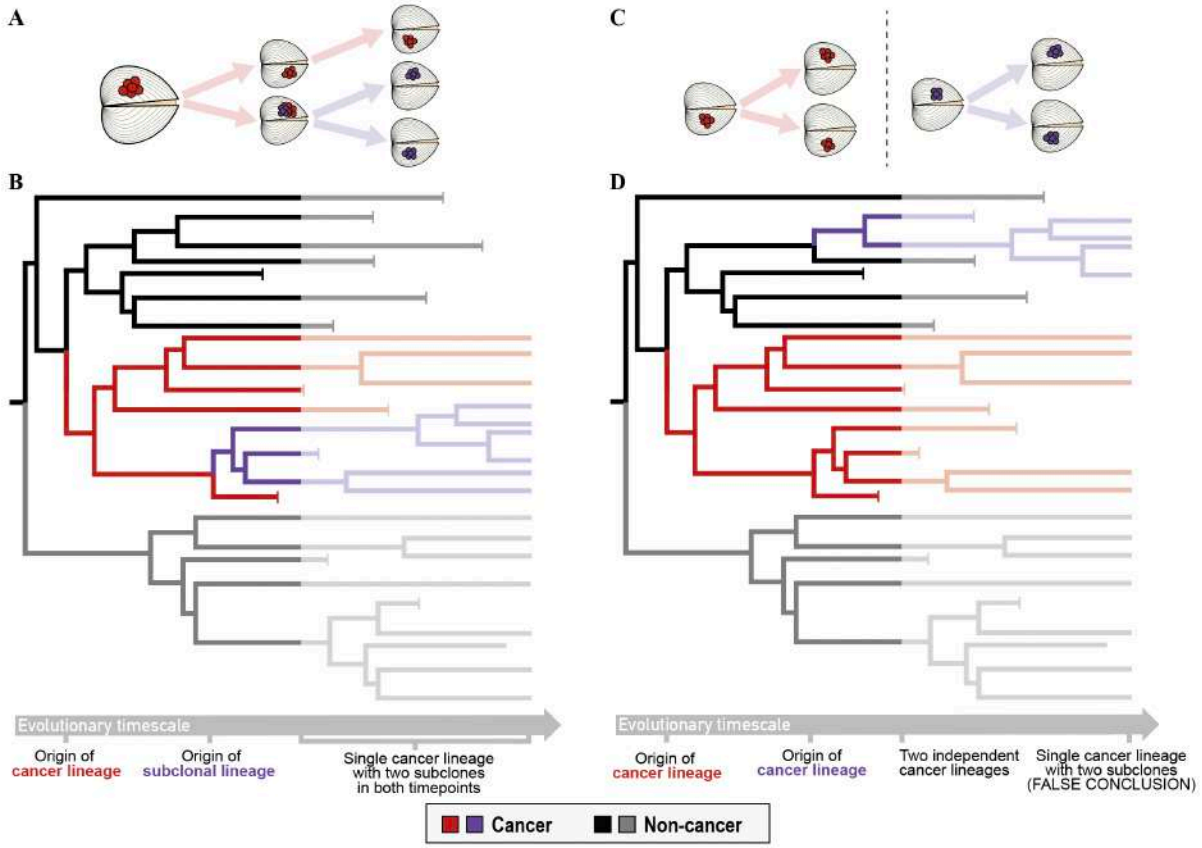


Figure 63. Scenarios of two independent cancer lineages or two subclones of a cancer lineage. **(A)** Diagram of a single cancer lineage with two subclones. **(B)** Phylogeny with a timescale at the bottom that indicates the origin of each subclone, and two sampling points separated on time that show the same phylogenetic results. **(C)** Diagram of two independent cancer lineages. **(D)** Phylogeny with a timescale at the bottom that indicates the origin of each independent lineage and two sampling points separated on time that do now show the same results on the most recent sampling due to the extinction of non-cancer lineages.

The life of a conventional non-contagious cancer lineage is restricted to the lifespan of the host (Pearse and Swift, 2006), in other words, the cancer lineage dies with the host. The founder of the cancer is the host itself and it can be used to filter the germline variation of cancer cells to obtain the somatic alterations (Figure 9). Nevertheless, deciphering the genomic landscape of contagious cancers represents a big challenge given by the fact that it is difficult to reconstruct the genetic background of the populations whence the founder of the lineage or clone derive. Given a set of contagious cancer samples, their evolutionary history can be reconstructed with two scenarios: a single cancer lineage with several subclones arising due to

the clonal evolution (Figure 63A) which represents a monophyletic relationship of the cancer samples (Figure 63B) or (2) the origin of two independent cancer lineages (Figure 63C) that represents a polyphyletic relationship (Figure 63D). In the case of obtaining a monophyletic phylogeny does not necessarily mean monophyly as branches of healthy individuals separating both cancer lineages could have been extinct. Therefore, the oldest are the cancer lineages, the highest probability of obtaining monophyletic results; though both scenarios could be possible.

The most accepted hypothesis of cockle transmissible cancers is two independent cancer lineages based on evidence from the EF1α gene, nine microsatellite loci and two different morphological phenotypes (Figure 25A, Metzger *et al.*, 2016). With the extensive histopathological study and whole-genome sequencing dataset produced in this thesis and within the Scuba Cancers project (unpublished data), more support is added to this hypothesis.

We started investigating the mitochondrial DNA (mtDNA) because in common cockles it is a haploid and short (~15kb) chromosome. This research shed some light into an unknown although quite frequent process happening in cockle contagious cancers: mitochondrial captures (see *Chapter 2*).

As expected, cancer mtDNA genotypes did not group with the host mtDNA genotypes (Figure 9) and instead clustered into nine distinct branches, which added to the inability of finding nuclear DNA support for these nine lineages suggested the potential capture of mitochondria by the two cancer lineages already described.

Cancer cells require functional mitochondria regardless of being contagious or not, but contagious cancer lineages will acquire mutations and copy number alterations of mtDNA that should extinguish the lineage (Wallace, 2012). However, this limitation has been circumvented by the periodic acquisition by cancer cells of normal mtDNA from host cells (Rebbeck, Leroi and Burt, 2011).

Because there is no known mechanism for intercellular transfer of mtDNA across both mitochondrial inner and outer membranes and the plasma membrane (Figure 64), transfer of whole mitochondria is most likely which has previously been shown to occur in vitro and murine models (Tan *et al.*, 2015).

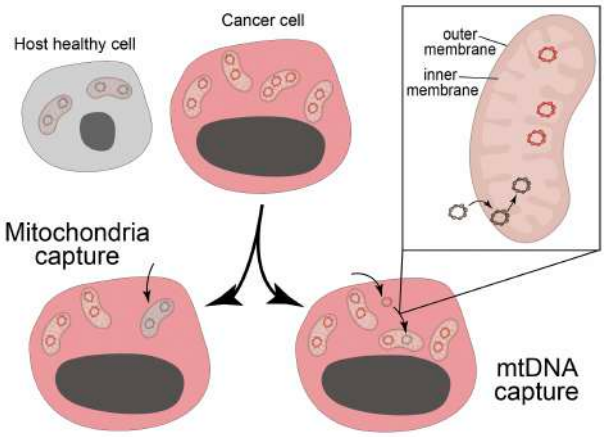


Figure 64. Mitochondria capture versus mtDNA capture.

Mutations in mtDNA have largely been described in cancer, but their contribution to tumour initiation, progression, and metastasis is less clear yet, for instance, mtDNA mutations and low mtDNA copy number are associated with increased metastasis and poor prognosis in breast cancer (Tan *et al.*, 2015). Therefore, we have characterised the mutational landscape of mtDNA on cockle transmissible cancers finding copy number amplifications (*i.e.,* duplications and triplications) on the regulatory region of mtDNA in three out of nine cancer lineages. They seem to be independent events as they do not share the end coordinate but two of them do share the start coordinate suggesting certain susceptibility for the amplification of this region. Not a single mtDNA of the 481 healthy cockles sequenced have shown any evidence of copy number amplifications in this region or any other mitochondrial region. Recently, a

tandem repeat in the mitochondrial genome of mussels with multiple copies of the control region was also reported in a contagious cancer lineage (Yonemitsu *et al.*, 2019).

5.1.2. TRACING THE ORIGIN OF CONTAGIOUS CANCERS IN COCKLES

As stated in *Section 1.3.2,* despite the fact that several HN have been reported in bivalves and corroborated to be contagious, the tissue from which the cell-of-origin of these cancer cells remains unknown in all of them. However, we have been discussing how the cancer genome contains an archaeological record of its past and within the pages of this thesis we have used a transcriptomic approach of different tissues to obtain the histogenesis of two BTN lineages affecting the same species, that is cockles, to give insights into the carcinogenesis of contagious cancers in bivalves.

Bivalve transmissible neoplasia usually consists of the proliferation of abnormal circulating cells with unknown origin disseminating through the circulatory system and infecting other individuals; it was generally considered to be a sarcoma (neoplasia of mesoderm-derived tissues) with a haematopoietic and a gonadal origin proposed (Alderman, Green and Balouet, 2017). Years ago, cells could only be defined by simple characteristics: spatial position, morphology, histochemical staining, or basic biochemical or biophysical properties, such as cell density or dye uptake (Wagner and Klein, 2020) and these cancer cells were first called "haemic neoplasia" suggesting the origin in the haemolymph (Elston *et al.*, 1988). However, with histopathological studies the possibility of a non-haemocytic cell line being the ancestry of HN cancer cells could not be ruled out and the term "haemic" was deprecated in favour of the term "disseminated" that did not imply the histogenesis of the neoplasia. Interestingly, gene-expression profiles of tumours often remain relatively stable during progression from primary tumour to metastasis and even end-stage disease (Visvader, 2011) providing a good scenario to investigate the origin of cancer cells.

We investigated the origin of two transmissible cancer lineages currently spreading among cockles by gene expression profiling of cancer samples, larval stages and seven healthy organs and tissues of cockles. Our analyses suggest a haemocytic origin for both cancer lineages suggesting that haemolymph cells might be prone to serve as the seed for a malignant cell to be able to colonize other individuals and avoid any immunological response.

The leukaemia-like cancer described across several bivalve species share morphological features that makes them fall under the same cancer label, however, we should not extrapolate the haematopoietic cell origin of cockle transmissible cancers to all HN as it might not be the case. Cell and tissue types show profound differences in their response to cancer driver mutations (Visvader, 2011) and it seems reasonable that, if it has happened twice the origin of a contagious cancer lineage in cockles, it might have happened as well in other bivalve transmissible neoplasia. In bivalves, circulating haemocytes leave the haemolymph to gain access to the intervalvar fluid (Caza *et al.*, 2020) and the environmental factors acting in the cells might be a risk to predispose these tissue to develop cancers.

In comparison with mammal contagious cancers (Table 4), we add a new potential cell line of transmissibility origin. For DFTD the histogenesis has been proposed to be a Schwann cell (Murchison *et al.*, 2010) while for CTVT it is thought to have a histiocytic origin (Ajayi *et al.*, 2018). The latest case a histiocyte is a normal immune cell that is found in many parts of the body, especially in the blood stream, and the lymph glands which is a similar location to the haemocyte cells of cockles' HN. Whether the haemolymph cell that gave rise to these two

contagious cancer lineages in cockles was involved in the immune response of these animals remains unclear.

In a nutshell, these results provide fundamental insights into the histogenesis of transmissible cancers in bivalves and opens a framework to investigate the role of mutational processes acting on cockles' haemolymph that allows cells to transform to cancer and become contagious twice.

5.1.3. BEYOND THE LIMITS OF CANCER CONTAGIONS

In marine bivalves, more than 15 species have been diagnosed with leukaemia-like cancers (Alderman *et al.*, 2017) and, here, we have reported a novel contagious cancer lineage affecting, in this case, the warty venus clam that inhabits the seas of Southern Europe. More pathological studies of bivalves could potentially report more species affected by leukaemia-like cancers.

Interestingly, in this thesis we have reported coinfections of two cancer lineages infecting a single individual which is remarkable given the low overall prevalence of the disease. The relative frequency of coinfection by various cancer lineages suggests susceptibility of contagion when a cancer lineage is already developing in an individual. The dynamics of coinfections in cockle transmissible cancers might be further investigated to shed light into the contagion patterns, cancer competitiveness and cockle's health effects.

On top of that, despite cancer cells being typically transmitted between individuals from the same species, on occasion they infect and propagate across populations from a second, different bivalve species (Metzger *et al.*, 2016; Yonemitsu *et al.*, 2019; Garcia-Souto *et al.*, 2022). Like a historical travel journal inscribed in DNA, the mutations in these tumours are a record of the past and allow us to inquire the origin, not the histogenesis in this case, but the species where the cancer cell was originated. We found out that the contagious cancer spreading among warty venus clams was originated in the striped venus clam that cohabits in the same areas (Figure 65). However, we were not able to sample a single striped venus clam affected with this cancer which might be pointing to the fact that the striped venus clams have acquired the defences needed to avoid contagion. Nonetheless, until recently the same was thought about MtrBTN2 in *M. trossulus* and it has recently refuted (Skazina *et al.*, 2021).

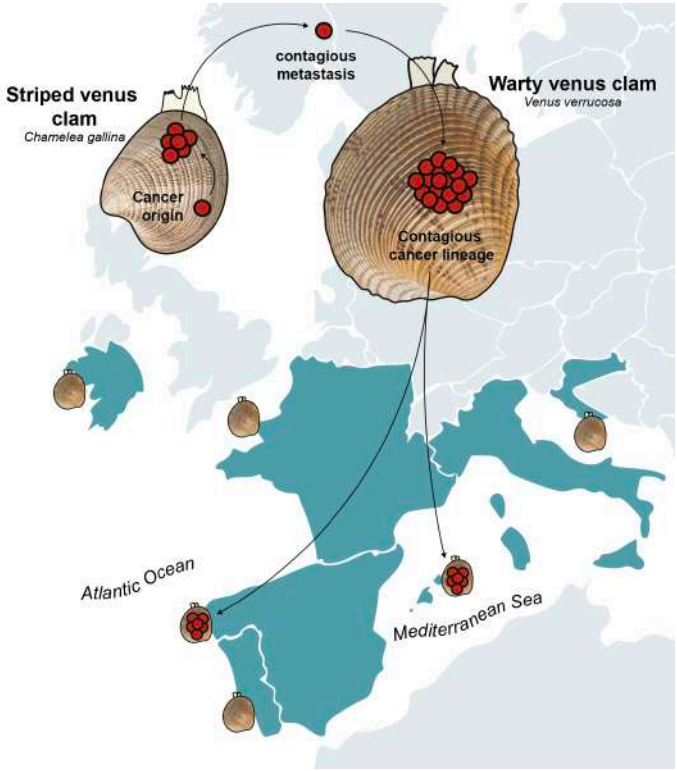


Figure 65. Schematic diagram summarising the interspecies contagion of a cancer lineage found in warty venus clams described in *Chapter 4.*

At least three cancers with interspecies metastases have been described (Table 1), one of them (MtrBTN2) is also found to be spreading among the species of cancer origin. Marine

contagious cancers in clams are able to jump from one genus to another (Figure 66A-B) while in mussels are widely spread and have jump to several species of the same genus (Figure 66C). It must be pointed out that hybridization is prevalent among mussels of three taxons *M. edulis*, *M. galloprovincialis*, and *M. trossulus* (Koehn, 1991), and in all localities where two mussel species are sympatric, hybridization has been detected (Gosling, 1992). However, to the best of my knowledge, no hybridization has been reported between *V. corrugata* and *P. aureus* or *C. gallina* and *V. verrucosa*.

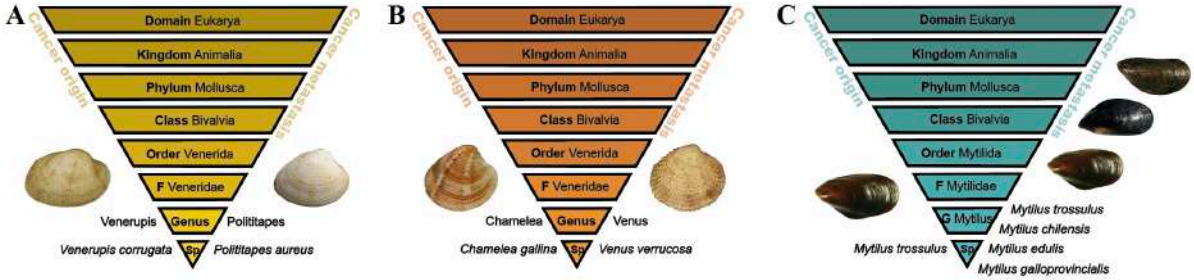


Figure 66. Interspecies metastases. Classification of species involved in **(A)** VcoBTN, **(B)** CgaBTN and **(C)** MtrBTN.

In definitive, cancer is generally addressed as a genetic disease of our own cells, but these findings should make us realise that this view is obsolete and that we should apply parasite knowledge to understand these contagious cancers.

5.2. LESSONS LEARNED: FROM BIVALVES TO CANCER THROUGH CONTAGION

5.2.1. IMPLICATIONS AND LIMITATIONS OF THIS RESEARCH

Every creative and systematic work undertaken to increase the stock of knowledge is positive for our society, today we know much more about marine contagious cancers than a decade ago when we did not even know of their existence. However, it is important to be critical with the results as many laboratory experiments are not reproducible or replicable and some are not even real evidence (Peng and Hicks, 2020).

I believe that the works included in this thesis show evidence of unknown facts of marine contagious cancers such as mitochondrial captures, histogenesis or another case of interspecific contagion in clams. In addition, it adds epidemiological data of these cancers and describes some genetic alterations acquired through the evolutionary history of cockle transmissible cancers. Nonetheless, limitations -some unexpected- have arisen during the research developed in this thesis. Below, I detail some of the lessons learned that can help future researchers in this field.

Cockle and clam bivalves usually live buried in the sand but depending on the region up to 3000 meters of depth and, in general, with increase in depth, bivalve richness decreases (Kamenev, 2013). However, the effort to collect samples just by walking in the beach at the time of the lowest tides against the extraction from the sand bed with a boat is very different, skewing our sample collection to walk-in sample collection or regions that economically exploit these resources. In addition, different collection methods added to different transportation times can stress the individuals prior to the process of samples and, for analysis such as transcriptomics, results may vary.

Bivalve transmissible neoplasia is a leukaemia-like cancer, therefore, as the disease progress, the infiltration of cancer cells among other tissues increases. A common approach to study diseases is to use the tissue composed by malignant cells and compare them to a healthy tissue. To obtain a good number of malignant cells, it is usually needed to be in the latest stages of the disease what, in our case, means infiltration to most of the tissues. The contamination of tumoral cells in the matched-normal cells and the low purity of the haemolymph in not severe cases of cancer has been a big challenge in this project. Cytometry and sorting or laser microdissection could be alternatives to overcome these issues.

Somatic alterations of a cancer reveal a lot about the disease. However, filtering out germline variation in contagious cancers is a big challenge as they are not present in the non-cancer cells of the individual with cancer (Figure 67A-B). Methods established to filter out germline variation of tumour samples from non-contagious individual (Figure 67C) do not work for contagious cancers (Figure 67D) making very tricky the identification of somatic variants, particularly those arising in the initial stages of cancer development. To overcome this issue, we used a panel of normal composed by 481 healthy cockles (Figure 67E). The filtration was not as good as expected and this might probably be related with the fact that the healthy cockles used to filter germline variation are from contemporary populations and not from the population in which the cancer has arisen, therefore we still have many ancestral polymorphisms that we are unable to filter out.

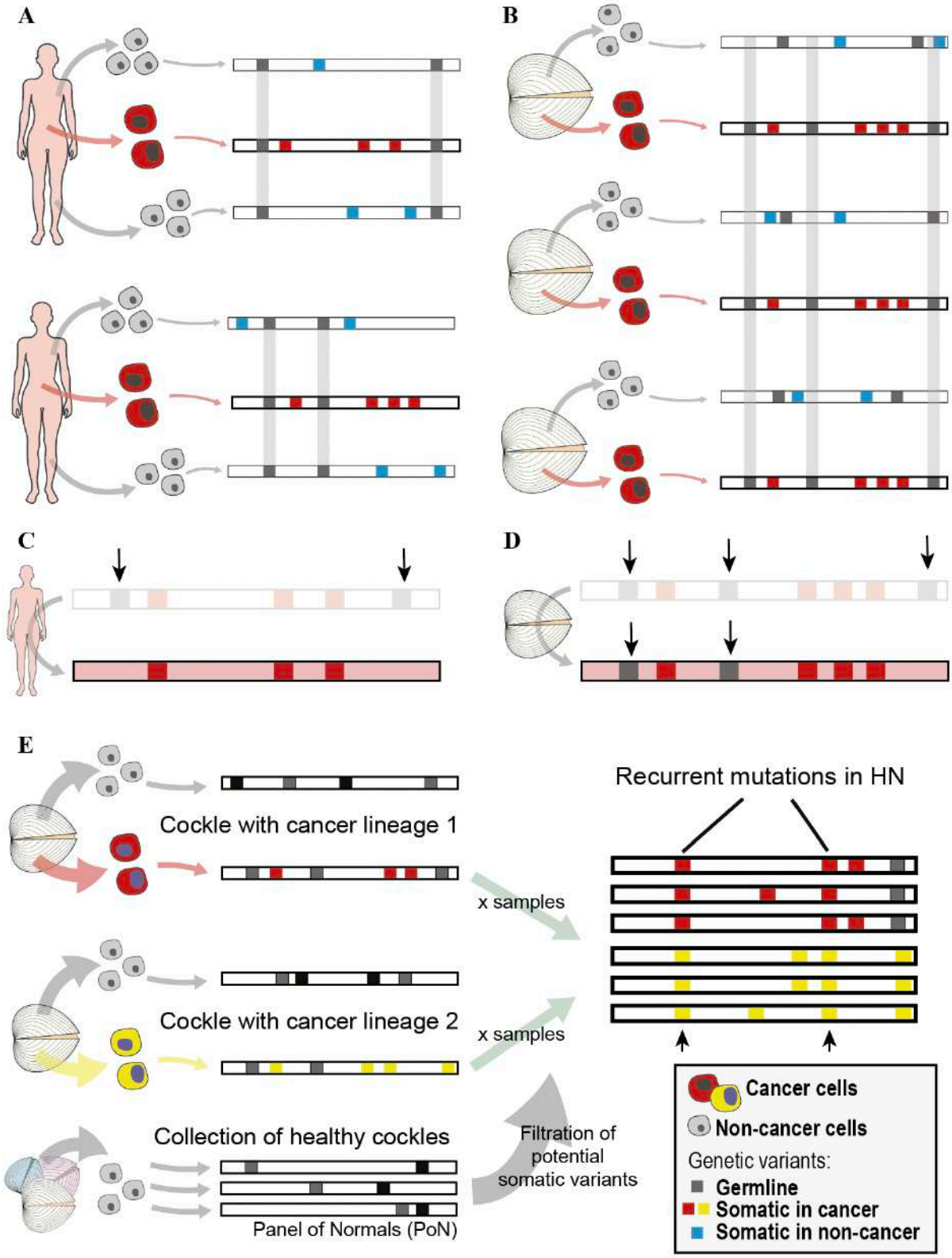


Figure 67. Somatic mutations filtration. **(A)** Landscape of variants and comparison of germline variants (grey) in the cancer cells and non-cancer cells of two patients with a non-contagious cancer. **(B)** Landscape of variants and comparison of germline variants (grey) in the cancer cells and non-cancer cells of three animals with a contagious cancer. **(C)** Filtration of germline variants in the tumoral genome of a non-contagious cancer. **(D)** Unsuccessful filtration of germline variants in the tumoral genome of a non-contagious cancer as they are not present in the matched-normal. **(E)** Strategy to obtain recurrent mutations in contagious cancers to unravel the genetic causes.

In addition to the fact that the recurrent mutations identified in this way (Figure 67E) are going to be enriched in ancestral germline variation common to both cancer lineages, the recurrent somatic mutations that would be expected to be found would probably be different mutations but with the same effect.

As previously discussed mitochondria is a haploid chromosome in cockles which made easier the analysis; however, two limitations difficulted the obtention of results. First, the closest-related species *Cerastoderma glaucum* is very divergent to be used as a root in these phylogenetic trees and, secondly, for time estimations we used the overall mutation rate of mtDNA in invertebrates (Allio *et al.*, 2017) that might not be extrapolated for cancer genomes that usually accumulate more mutations (Larman *et al.*, 2012).

5.2.2. AN ECOLOGICAL WARNING

Contagious cancer cells have acquired the ability to spread naturally to other individuals acting like a parasite, thus, gaining independence from its original host (Murgia, 2006). We should then change the way of thinking about these marine contagious cancers and consider them parasitic diseases with the ecological concerns that come along.

Firstly, the distribution and prevalence study of cockle transmissible cancers reported in *Chapter 2* showed that these two cancer lineages are restricted to Southern-Central European countries. Why are they not present on Northern populations and in the African coast? Four hypothesis or a combination of them could explain their distribution: (1) genetic diversity of cockles' populations that make the northern population resistant or less susceptible to the cancer infection, (2) restrictions in the environmental conditions necessary for cancer cells transmissibility, (3) inability of cancer cells to travel that far with the ocean currents present in the seas or (4) low prevalence that it was not detected in our study. In the latest case, it is of foremost importance to avoid human activity to introduce the cancer into disease-free areas. Unlike in mussels and oysters, transfers for culturing purposes between coastal regions have not been practiced with cockles (Krakau *et al.*, 2012) but human activity does moves seawater and cockles for its purchases (normally kept in sewage treatment plants before they are placed on the market) that could be the via of introduction of cancer in distant regions and we also move cockles for the market.

Secondly, the genetic analysis of eight warty venus clams infected with a contagious cancer described in *Chapter 4* revealed the same cancer lineage in two locations more than 1000 nautical miles away in the Atlantic Ocean and the Mediterranean Sea Coasts of Spain. Was it due to human activity? More research should be done to discard this hypothesis.

Finally, tens of leukaemia-like cancers affecting bivalves have been reported since 1969 and already nine of them have been corroborated to be contagious (Table 1). Are all of them contagious? Are there more HN cancers affecting other bivalve species? More research should be done to clarify these aspects and have an overview of what is happening among bivalves in the seas.

5.3. FUTURE DIRECTIONS

The database of contagious cancers genomes described in the pages of this thesis might allow us to answer other questions in the near future. In the following sections, I describe the three major topics for which our knowledge could be expanded.

5.3.1. TIME TO CHANGE THE CANCER PARADIGM

The power of comparative oncology is usually neglected. We need to change the idea of cancer as a genetic disease of our own cells and start to explore the cancers that do not follow this rule. So far, we know cases in two terrestrial and nine marine species which is not negligible. Despite this, little is known about cancer transmissibility that is not more than a large-scale metastasis. Nowadays, it is estimated that 90% of cancer deaths are due to metastasis, therefore, we need to accept that our knowledge of cancer is not enough and to start exploring this mechanism from as many points of view as we can think of. The new paradigm should reflect that cancer can be contagious, a cancer which invades more than one individual living in a particular time, rather than time-limited disease framed by an organism or mere self-destructive entities that make the host survival incompatible the cancer survival. This could ultimately lead to an improvement in treatments if we find the underlying genetic causes of metastases.

It is not difficult to find analogies between contagious cancers and human cancers. As it happens with many conventional human cancers (Figure 1), cockle transmissible cancers also show numerous chromosomal aberrations when comparing a cockle healthy cell (Figure 68A) against cancer cells (Figure 68B). Investigating the structural variation of contagious cancers might shed light into the driver and passenger mutations of cancer lineages and the essential genes needed for cancer lineage survival and transmissibility.

Cancer progression and metastasis can be explored by investigating somatic mutations accumulated in marine contagious cancers, the differences and similarities of cancer lineages, by timing the driver mutations and estimating mutation rates, by identifying the needs cancer cells need to survive in the marine environment before reaching a new host, by studying the limits of interspecies contagions, by evaluating cancer lineage competition in coinfections, by identifying pathogenicity of certain mutations or by characterizing the role of the cell-of-origin in cancer dispersal.

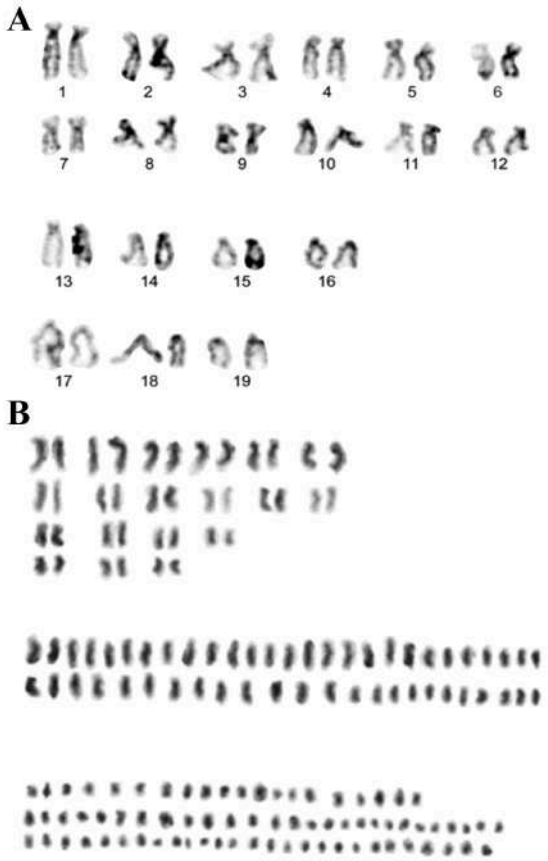


Figure 68. Chromosome comparison of healthy and cancerous cockles. **(A)** Karyotypes showing the set of metaphase chromosomes sorted by length and centromere location (adapted from Leitão *et al.*, 2008; reprinted with permission from Elsevier Ltd., see Appendix H). **(B)** Chromosomes extracted from cancer cells of cockle transmissible cancers (adapted from Matias *et al.*, 2014; reprinted with permission from Elsevier Ltd., see Appendix H).

5.3.2. INFECTION AND IMMUNE RESPONSES

Cancer is thus usually a self-limiting disease—it either regresses or it kills its host, and the death of the host marks the death of the cancer lineage (Metzger and Goff, 2016). However, if cancer cells travel from one individual to another, a normal immune system would be able to recognize them as foreign and reject them. Why does this not happen in the case of contagious cancers? How does it work the immune system of these animals? How contagious cancers evade the host defences?

Bivalves are a special case in transmissible cancer due to the enormous genetic diversity when comparing them to CTVT or DFTD in which it has been argued that one of the causes on cancer contagion is the low genetic diversity. On the other hand, in bivalves, genetic diversity may not play such an important role against transmission, due to the absence of an adaptive immune system (see *Section 1.4.2*) which could explain why there are so many BTNs. However, future studies of BTNs would probably bring a better understanding of the immune system in invertebrates.

5.3.3. STRATEGIES FOR DIAGNOSIS

Histocytological examination of the haemolymph has been the preferred diagnostic method for many species. However, in some species the morphological differences of cancer cells and haemocytes are not that obvious, as it happens in the warty venus clam. In this thesis, flow cytometry was used to characterize cancer cells from warty venus clams as it was previously used in other BTN lineages affecting *C. edule* (Le Grand *et al.*, 2010), *P. aereus* (Carballal *et al.*, 2014) or *M. trossulus* (Skazina *et al.,* 2021, 2022). In warty venus clam, we saw ploidy differences between normal and cancer haemolymph preparations pointing to the potential use of this technique as a diagnostic method in this species.

Furthermore, qPCR of environmental DNA (eDNA) to detect cancer cells has already been developed for the soft-shell clam (Giersch *et al.*, 2022). The development of methodologies to detect the cancer cells by sampling water of the region where a bivalve population is located could be very useful for the management of this infectious cancers. In this thesis, we describe CN amplifications and SNVs that could be used as markers of CedBTN lineages to screen and manage this disease through molecular diagnosis.

5.3.4. TOWARD A COMPREHENSIVE DISEASE MODEL

Last but not least, improving our knowledge on haemic neoplasia could reveal some genomic insights of their bivalve hosts that could be useful for our understanding of wild species as well as the management of them in aquaculture, an economic activity of many families.

A new animal model of disease has tremendous advantages for science but as well for the model species. Apart from being well characterized, their contagious cancers could potentially be monitored and even treated by using molecular markers or target genes that could be obtained while researching the goals of the two previous sections.

Bivalves are good animal models for cancer transmissibility research. Here, I break down some of the advantages of bivalves over mammal species that also suffer from natural-ocurring contagious cancers:

1. Several bivalve species have been reported to be affected by contagious cancers, therefore, more cancer lineages can be used to unravel the similarities and differences between them.
2. Cancer contagion can be achieved in a laboratory mimicking what happens in the seabed and the animals can be easily handled by most researchers. Moreover, the availability of multiple reports of prevalence from different regions might help to reduce the number of animals needed to collect to get a cancerous individual.
3. Bivalves have a short reproductive cycle that can be induced in the laboratory, and they produce a larger number of offspring compared to dogs or Tasmanian devils. Moreover, in less than a year they are reproductively active and external fecundity facilitates to perform desired crosses.
4. HN can be diagnosed, and cancer cells can be extracted from the individual alive with little impact.
5. Cancer progression is shorter than that of dogs or Tasmanian devils, within 2-4 months the animal can achieve the latest stages of the disease (personal communication of Álex Viña).
6. Bivalves are worldwide exploited for food and ornamentation or pearls, therefore, extensive knowledge on the biology of these species has been published.
7. Genome availability of two bivalve species with contagious cancers facilitate the genetic study of this disease and the search for molecular markers.

Some disadvantages of bivalves over dogs of Tasmanian devils for the study of cancer transmissibility are the low prevalence of HN, the small size of the animals hindering the number of cells that can be used and the absence of established lines for cell culture.

5.4. CONCLUSIONS

In this doctoral thesis, *Evolution of Bivalve Transmissible Cancers*, I have come to conclusions that might shed some light into the evolution of marine transmissible cancers and further our understanding of cancer contagion.

The main findings arising from this thesis are listed below:

(1) **Cockle transmissible cancers** are a ~5% prevalence leukaemia-like cancer only found in Southern and Central European countries within the distribution range of the species *Cerastoderma edule* between 2016 and 2021.

(2) Two **histological** phenotypes corresponding to the microsatellites **nuclear profile** were identified among cockle cancer samples.

(3) Mitochondrial DNA analysis revealed **nine captures** corresponding to cockle mitochondrial cancer lineages that happened in different regions and timepoints.

(4) Several mitochondrial cancer lineages were found in the same region while some other regions only have a single mitochondrial cancer lineage.

(5) Mitochondrial cancer lineages did not correspond to A/B histological phenotypes although no mixture of phenotypes was found in any mitochondrial cancer lineage.

(6) **Coinfection** of two histological phenotypes and/or two mitochondrial cancer lineages affecting the same cockle are reported.

(7) Independent **mtDNA CN amplifications** have been reported on three out of nine cancer lineages but not in a single healthy cockle.

(8) A **haematopoietic** origin of cockle transmissible cancers was revealed by a gene expression atlas of cockles.

(9) The histogenesis of two phenotypically different cockle transmissible cancer lineages (A & B) is the **same**.

(10) *Venus verrucosa* from the Atlantic coast and from the Mediterranean Sea are affected by a **leukaemia-like cancer**.

(11) Cancer cells morphology and ploidy of *Venus verrucosa* match the **general features** identified in other bivalve transmissible neoplasia (round shape, pleomorphic nucleus, higher ploidy, chromosome instability).

(12) Atlantic and Mediterranean *Venus verrucosa* cancer cells had likely originated in the same animal, indicating that the cancer is **contagious** and had spread through different populations.

(13) Cancer DNA of *Venus verrucosa* matched the *Chamelea gallina*, a species that cohabits with the cancer host clams suggesting that the cancer started in a *Chamelea gallina* and then spread to a *Venus verrucose*, in other words, an **interspecific transmission**.No leukaemia-like cancer was found in a collection of 207 *Chamelea gallina* from the Mediterranean Sea.

In essence, this thesis combined sample collection, molecular biology experiments and bioinformatic analysis from several lineages of two species affected by bivalve transmissible cancers to understand the genetic diversity and evolution of marine contagious cancers.

Finally, we initiated this thesis defining the etymology of cancer and connecting it to our protagonists because crabs can be find inside a cockle (Longshaw and Malham, 2013) something that made me scream the first time that I opened a Portuguese cockle and found a crab back in 2017. I would like to close the thesis again with a crab anecdote: in the seventeenth century, a cheap paste of crab's eyes was popular to treat cancer with no success (Mukherjee, 2010). Research has remarkably improved our cancer treatments and we are halfway in the journey of eradicating the emperor of all maladies. This is my modest step towards that and it warms the cockles of my heart.

Bibliography

"Life on earth is more like a verb. It repairs, maintains, re-creates, and outdoes itself." Lynn Margulis

Bibliography

Science knowledge has been built over millennia, the doctoral candidate regrets the inability to cite all studies that have shaped the understanding of cancer metastasis and contagious cancers. Below you will find the list of references used on this thesis.

AboElkhair, M., Siah, A., et al. (2009) 'Reverse transcriptase activity associated with haemic neoplasia in the soft-shell clam *Mya arenaria'*, *Diseases of Aquatic Organisms*, 84(1), pp. 57–63. doi: 10.3354/dao02038.

AboElkhair, M., Synard, S., et al. (2009) 'Reverse transcriptase activity in tissues of the soft-shell clam *Mya arenaria* affected with haemic neoplasia', *Journal of Invertebrate Pathology*, 102(2), pp. 133–140. doi: 10.1016/j.jip.2009.06.009.

Adema, C. M. (2021) 'Sticky problems: Extraction of nucleic acids from molluscs', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1825). doi: 10.1098/rstb.2020.0162.

Ades, F., Tryfonidis, K. and Zardavas, D. (2017) 'The past and future of breast cancer treatment - From the papyrus to individualised treatment approaches', *Ecancermedicalscience*, 11, pp. 1–11. doi: 10.3332/ecancer.2017.746.

Ajayi, O. L. et al. (2018) 'Cytomorphological, histopathological and immunohistochemical observations on the histiocytic origin of canine transmissible venereal tumour', *Sokoto Journal of Veterinary Sciences*, 16(2), p. 10. doi: 10.4314/sokjvs.v16i2.2.

Aktipis, C. A. et al. (2015) 'Cancer across the tree of life: cooperation and cheating in multicellularity', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, p. 20140219. doi: 10.1098/rstb.2014.0219.

Al-Khalaifah, H. and Al-Nasser, A. (2019) 'Immune Response of Molluscs', *Molluscs*. doi: 10.5772/intechopen.81778.

Alderman, D., Green, M. and Balouet, G. (2017) 'Disseminated neoplasms in bivalves', *ICES Identification Leaflets for Diseases and Parasites of Fish and Shellfish*, 67, pp. 1–6. doi: http://doi.org/10.17895/ices.pub.2098.

Alexandrov, L. B. et al. (2013) 'Deciphering Signatures of Mutational Processes Operative in Human Cancer', *Cell Reports*, 3(1), pp. 246–259. doi: 10.1016/j.celrep.2012.12.008.

Allam, B. and Raftos, D. (2015) 'Immune responses to infectious diseases in bivalves', *Journal of Invertebrate Pathology*, 131, pp. 121–136. doi: 10.1016/j.jip.2015.05.005.

Allio, R. et al. (2017) 'Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker', *Molecular Biology and Evolution*, 34(11), pp. 2762–2772. doi: 10.1093/molbev/msx197.

Álvarez, E. G. et al. (2021) 'Aberrant integration of Hepatitis B virus DNA promotes major restructuring of human hepatocellular carcinoma genome architecture', *Nature communications*, 12, p. 6910. doi: 10.1038/s41467-021-26805-8.

Aparicio, S. et al. (2002) 'Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes', *Science*, 297(5585), pp. 1301–1310. doi: 10.1126/science.1072104.

Appeldoorn, R., Brown, C. and Chang, P. (1984) 'Field and laboratory studies to define the occurrence of neoplasia in the soft-shell clam, *Mya arenaria'*, *American Petroleum Institute Publication*, 4345.

Arakawa, A. et al. (2021) 'Vaginal Transmission of Cancer from Mothers with Cervical Cancer to Infants', *The New England Journal of Medicine*, 384(1), pp. 42–50. doi: 10.1056/NEJMoa2030391.

Arriagada, G. et al. (2014) 'Activation of transcription and retrotransposition of a novel retroelement, Steamer, in neoplastic hemocytes of the mollusk *Mya arenaria*', *Proceedings of the National Academy of Sciences of the United States of America*, 111(39), pp. 14175–14180. doi: 10.1073/pnas.1409945111.

Van der Auwera, G. and O'Connor, B. (2020) 'Genomics in the Cloud: Using Docker, GATK, and WDL in Terra'. 1st ed. O'Reilly Media, release April 2020. ISBN : 1-4919-7518-0. OCLC : (OCoLC)1151768020

Bachère, E., Chagot, D. and Grizel, H. (1988) 'Separation of *Crassostrea gigas* hemocytes by density gradient centrifugation and counterflow centrifugal elutriation', *Developmental and Comparative Immunology*, 12(3), pp. 549–559. doi: 10.1016/0145-305X(88)90071-7.

Báez, A. et al. (2019) 'Somatic evolution and global expansion of an ancient transmissible cancer lineage', *Science*, 365(August), pp. 1–7. doi: 10.1126/science.aau9923.

Báez, A. (2019) 'The somatic evolution of the Canine Transmissible Venereal Tumour'. Department of Veterinary Medicine. University of Cambridge. *Thesis* supervised by E. Murchinson.

Báez, A. (2021b) 'As cancer grows old', *Science*, 374(6571), p. 1066. doi: 10.1126/science.abm8137.

Bailey, M. H. et al. (2018) 'Comprehensive Characterization of Cancer Driver Genes and Mutations', *Cell*, 173(2), pp. 371-385.e18. doi: 10.1016/j.cell.2018.02.060.

Balouet, G. et al. (1986) 'Proliferative hemocytic condition in European flat oysters (Ostrea edulis) from Breton coasts: A 6-year survey', *Journal of Invertebrate Pathology*, 48(2), pp. 208–215. doi: 10.1016/0022-2011(86)90125-4.

Banfield, W. G. et al. (1965) 'Mosquito Transmission of a Reticulum Cell Sarcoma of Hamsters', *Science*, 148(3674), p. 1240. doi: 10.1126/science.148.3674.1239.

Barber, B. J. (2004) 'Neoplastic diseases of commercially important marine bivalves', *Aquat. Living Resources EDP Sciences*, 17, pp. 449–466. doi: 10.1051/alr:2004052.

Baruzzo, G. et al. (2017) 'Simulation-based comprehensive benchmarking of RNA-seq aligners', *Nature Methods*, 14(2), pp. 135–139. doi: 10.1038/nmeth.4106.

Benjamin, D. et al. (2019) 'Calling Somatic SNVs and Indels with Mutect2', *bioRxiv*. doi: 10.1101/861054.

Bergers, G. and Fendt, S.-M. (2021) 'The metabolism of cancer cells during metastasis', *Nature Reviews Cancer*, 21, pp. 162–180. doi: 10.1038/s41568-020-00320-2.

Bernt, M. et al. (2013) 'MITOS: Improved de novo metazoan mitochondrial genome annotation', *Molecular Phylogenetics and Evolution*, 69(2), pp. 313–319. doi: 10.1016/j.ympev.2012.08.023.

Bettisworth, B. and Stamatakis, A. (2021) 'Root Digger: a root placement program for phylogenetic trees', *BMC Bioinformatics*, 22(1):225. doi: 10.1186/s12859-021-03956-5.

Blaine, D. (1810) 'A Domestic Treatise on the Diseases of Dogs', Oxford University Press. ISBN: 9781374582613

Bouckaert, R. et al. (2019) 'BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis', *PLoS Computational Biology*, 15(4), pp. 1–28. doi: 10.1371/journal.pcbi.1006650.

Bouckaert, R. R. (2010) 'DensiTree: Making sense of sets of phylogenetic trees', *Bioinformatics*, 26(10), pp. 1372–1373. doi: 10.1093/bioinformatics/btq110.

Boveri, T. (1914) Zur Frage der Entstehung maligner Tumoren. Fischer, Jena. OCoLC: 10215868

Boyden, C. R. (1973) 'Observations on the shell morphology of two species of cockle *Cerastoderma edule* and *C. glaucum*', *Zoological Journal of the Linnean Society*, 52(4), pp. 269–292. doi: 10.1111/j.1096-3642.1973.tb01885.x.

Boyko, A. R. (2011) 'The domestic dog: man's best friend in the genomic era', *Genome Biology*, 12(2). doi: 10.1186/gb-2011-12-2-216.

Brindley, D. C. and Banfield, W. G. (1961) 'A contagious tumor of the hamster', *Journal of the National Cancer Institute*, 26(4), pp. 949–957. doi: 10.1093/jnci/26.4.949.

Brousseau, D. J. (1987) 'Seasonal aspects of sarcomatous neoplasia in *Mya arenaria* (soft-shell clam) from Long Island Sound.', *Journal of invertebrate pathology*, 50(3), pp. 269–276. doi: 10.1016/0022-2011(87)90092-9.

Brown, R. S. et al. (1985) 'Prevalence of neoplasia in 10 New England populations of the soft-shell clam (*Mya arenaria*)', *Annals New York Academy of Sciences*, 112, pp. 522–534.

Burioli, E. A. V. et al. (2019) 'Implementation of various approaches to study the prevalence, incidence, and progression of disseminated neoplasia in mussel stocks', *Journal of Invertebrate Pathology*, 168(October), p. 107271. doi: 10.1016/j.jip.2019.107271.

Canapa, A. et al. (1996) 'Phylogenetic analysis of veneridae (Bivalvia): Comparison of molecular and palaeontological data', *Journal of Molecular Evolution*, 43(5), pp. 517–522. doi: 10.1007/BF02337522.

Carballal, M. J. et al. (2001) 'Parasites and pathologic conditions of the cockle *Cerastoderma edule* populations of the coast of Galicia (NW Spain)', *Journal of Invertebrate Pathology*, 78(2), pp. 87–97. doi: 10.1006/jipa.2001.5049.

Carballal, M. J. et al. (2015) 'Neoplastic diseases of marine bivalves', *Journal of Invertebrate Pathology*, 131, pp. 83–106. doi: 10.1016/j.jip.2015.06.004.

Carballal, M. J. et al. (2016) 'Parasites, pathological conditions and resistance to *Marteilia cochillia* in lagoon cockle *Cerastoderma glaucum* from Galicia (NW Spain)', *Diseases of Aquatic Organisms*, 122(2), pp. 137–152. doi: 10.3354/dao03070.

Carrilho Rodrigues da Silva, J. (2012) 'Caracterização e mapeamento de marcadores moleculares em espécies da família veneridae de interesse comercial em Portugal e Espanha. Estudo da hibridação entre *Ruditapes decussatus* e *Ruditapes philippinarum*', Instituto de Ciencias Biomedicas Abel Salazar. Universidade do Porto. Available at: http://cyber.usask.ca/login?url=https://search.proquest.com/docview/1926728557?accountid=14739&bdid=6493&_bd=ujrer7bJfB8ehBHROApq3xgUED8%3D.

Carss, D. N. et al. (2020) 'Ecosystem services provided by a non-cultured shellfish species: The common cockle Cerastoderma edule', *Marine Environmental Research*, 158(September 2019). doi: 10.1016/j.marenvres.2020.104931.

Caulin, A. F. and Maley, C. C. (2011) 'Peto's Paradox: Evolution's prescription for cancer prevention', *Trends in Ecology and Evolution*, 26(4), pp. 175–182. doi: 10.1016/j.tree.2011.01.002.

Caza, F. et al. (2020) 'Hemocytes released in seawater act as Trojan horses for spreading of bacterial infections in mussels', *Scientific Reports*, 10(1), pp. 1–12. doi: 10.1038/s41598-020-76677-z.

Chakraborty, S., Ray, M. and Ray, S. (2021) 'Bivalve haemocyte adhesion, aggregation and phagocytosis: A tool to reckon arsenic induced threats to freshwater ecosystem', *Fish and Shellfish Immunology*, 114(March), pp. 229–237. doi: 10.1016/j.fsi.2021.05.008.

Christensen, D. J., Farley, C. A. and Kern, F. G. (1974) 'Epizootic neoplasms in the clam *Macoma balthica* (L.) from chesapeake bay', *Journal of the National Cancer Institute*, 52(6), pp. 1739–1749. doi: 10.1093/jnci/52.6.1739.

Chu, F.-L. (1988) 'Humoral defense factors in marine bivalves', *Disease Processes in Marine Bivalve Molluscs*, p. 121. Available at: https://scholarworks.wm.edu/vimsbooks/121.

Cieślik, M. and Chinnaiyan, A. M. (2018) 'Cancer transcriptome profiling at the juncture of clinical translation', *Nature Reviews Genetics*, 19(2), pp. 93–109. doi: 10.1038/nrg.2017.96.

Cohen, D. (1985) 'The Canine Transmissible Venereal Tumor: A Unique Result of Tumor Progression', *Advances in Cancer Research*. doi: 10.1016/S0065-230X(08)60943-4.

Collins, C. M. and Mulcahy, M. F. (2003) 'Cell-free transmission of a haemic neoplasm in the cockle *Cerastoderma edule*', *Diseases of Aquatic Organisms*, 54(1), pp. 61–67. doi: 10.3354/dao054061.

Cooper, H. L., Mackay, C. M. and Banfield, W. G. (1964) 'Chromosome studies of a contagious reticulum cell sarcoma of the syrian hamster', *Journal of the National Cancer Institute*, 33(4), pp. 691–706. doi: 10.1093/jnci/33.4.691.

Cooper, K. R., Brown, R. S. and Chang, P. W. (1982) 'Accuracy of blood cytological screening techniques for the diagnosis of a possible hematopoietic neoplasm in the bivalve mollusc, *Mya arenaria*', *Journal of Invertebrate Pathology*, 39, pp. 281–289.

Coorens, T. H. H. and Behjati, S. (2022) 'Tracing and Targeting the Origins of Childhood Cancer', *Annual Review of Cancer Biology*, 6(1), pp. 35–47. doi: 10.1146/annurev-cancerbio-070620-091632.

Craig Venter, J. et al. (2001) 'The sequence of the human genome', *Science*, 291(5507), pp. 1304–1351. doi: 10.1126/science.1058040.

Dairain, A. et al. (2020) 'High prevalences of disseminated neoplasia in the Baltic tellin *Limecola balthica* in the Wadden Sea', *Diseases of Aquatic Organisms*, 138, pp. 89–96. Available at: https://www.int-res.com/abstracts/dao/v138/p89-96/.

Darriba, D. et al. (2012) 'JModelTest 2: More models, new heuristics and parallel computing', *Nature Methods*, 9(8), p. 772. doi: 10.1038/nmeth.2109.

Darriba, Di. et al. (2020) 'ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models', *Molecular Biology and Evolution*, 37(1), pp. 291–294. doi: 10.1093/molbev/msz189.

Das, U. and Das, A. K. (2000) 'Review of canine transmissible venereal sarcoma', *Veterinary Research Communications*, 24(8), pp. 545–556. doi: 10.1023/A:1006491918910.

Decker, B. et al. (2015) 'Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumor', *Genome Research*, 25(11), pp. 1646–1655. doi: 10.1101/gr.190314.115.

Díaz, S. (2005) 'Estudio de Neoplasia Diseminada y otras alteraciones patológicas y parásitos que afectan a las poblaciones de berberecho, *Cerastoderma edule*, de Galicia'. Universidade de Santiago de Compostela. Supervised by A. Villalba and M. J. Carballal.

Díaz, S. et al. (2013) 'Disseminated neoplasia causes changes in ploidy and apoptosis frequency in cockles *Cerastoderma edule*', *Journal of Invertebrate Pathology*, 113(3), pp. 214–219. doi: 10.1016/j.jip.2013.03.010.

Díaz, S. et al. (2010) 'Expression of mutant protein p53 and Hsp70 and Hsp90 chaperones in cockles *Cerastoderma edule* affected by neoplasia', *Diseases of Aquatic Organisms*, 90(3), pp. 215–222. doi: 10.3354/dao02231.

Díaz, S. et al. (2011) 'Disseminated neoplasia in cockles *Cerastoderma edule*: Ultrastructural characterisation and effects on haemolymph cell parameters', *Diseases of Aquatic Organisms*, 96(2), pp. 157–167. doi: 10.3354/dao02384.

Díaz, S. (2015) 'Neoplasia diseminada en poblaciones del berberecho *Cerastoderma edule* en Galicia'. Universidade de Santiago de Compostela. *Thesis* supervised by A.Villalba and M. J. Carballal.

Díaz, S. et al. (2016) 'Long-term epidemiological study of disseminated neoplasia of cockles in Galicia (NW Spain): Temporal patterns at individual and population levels, influence of environmental and cockle-based factors and lethality', *Journal of Fish Diseases*, pp. 1–16. doi: 10.1111/jfd.12436.

Díaz, S., Villalba, A. and Carballal, M. J. (2017) 'Assessment of transmissibility of the disseminated neoplasia affecting cockles *Cerastoderma edule* in Galicia (NW Spain)', *Bulletin of the European Association of Fish Pathologists*, 37(1), pp. 5–15.

Dictionary.com (2022) 'Evolution', Retrieved from URL: https://www.dictionary.com/browse/evolution.

Dobin, A. et al. (2013) 'STAR: Ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1), pp. 15–21. doi: 10.1093/bioinformatics/bts635.

Drummond, A. J. et al. (2006) 'Relaxed phylogenetics and dating with confidence', *PLoS Biology*, 4(5), pp. 699–710. doi: 10.1371/journal.pbio.0040088.

Edgar, R. C. (2004) 'MUSCLE: Multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*, 32(5), pp. 1792–1797. doi: 10.1093/nar/gkh340.

Effron, M., Griner, L. and Benirschke, K. (1977) 'Nature and rate of neoplasia found in captive wild mammals, birds, and reptiles at necropsy', *Journal of the National Cancer Institute*, 59(1), pp. 185–198. doi: 10.1093/jnci/59.1.185.

Ekblom, R. and Wolf, J. B. W. (2014) 'A field guide to whole-genome sequencing, assembly and annotation', *Evolutionary Applications*, 7(9), pp. 1026–1042. doi: 10.1111/eva.12178.

Ellegren, H. (2014) 'Genome sequencing and population genomics in non-model organisms', *Trends in Ecology and Evolution*, 29(1), pp. 51–63. doi: 10.1016/j.tree.2013.09.008.

Elliott, T. A. and Gregory, T. R. (2015) 'What's in a genome? The C-value enigma and the evolution of eukaryotic genome content', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, pp. 1–10. doi: 10.1098/rstb.2014.0331.

Elston, R. A. et al. (1988) 'Using Whole Cells and Cell H', *Comparative and General Pharmacology*, 12(c), pp. 719–727.

Elston, R. A., Kent, M. L. and Drum (1988) 'Progression, remission and transplantation of hemic neoplasia in the bay mussel, *Mytilus edulis'*, *Diseases of Aquatic Organisms*, 4, pp. 135–142. doi: 10.1016/0145-305X(89)90177-8.

Elston, R., Moore, D. and Brooks, K. (1992) 'Disseminated neoplasia Bivalve Molluscs', Reviews in Aquatic Sciences, 6(5,6), pp. 405–466.

Enriquez-Navas, P. M., Wojtkowiak, J. W. and Gatenby, R. A. (2015) 'Application of evolutionary principles to cancer therapy', *Cancer Research*, 75(22), pp. 4675–4680. doi: 10.1158/0008-5472.CAN-15-1337.

Escoubas, J. M. et al. (2016) 'Immunity in Molluscs', *Encyclopedia of Immunobiology*, 1, pp. 417–436. doi: 10.1016/B978-0-12-374279-7.12004-1.

Fares, J. et al. (2020) 'Molecular principles of metastasis: a hallmark of cancer revisited', *Signal Transduction and Targeted Therapy*, 5(28), pp. 1–17. doi: 10.1038/s41392-020-0134-x.

Farhat, S. et al. (2022) 'Comparative analysis of the *Mercenaria mercenaria* genome provides insights into the diversity of transposable elements and immune molecules in bivalve mollusks', *BMC Genomics*, 23(1), pp. 1–23. doi: 10.1186/s12864-021-08262-1.

Farley, C. A. (1969a) 'Probable neoplastic disease of the hematopoietic system in oysters, *Crassostrea virginica* and *Crassostrea gigas*.', *National Cancer Institute monograph*, 31, pp. 541–555.

Farley, C. A. (1969b) 'Sarcomatoid proliferative disease in a wild population of blue mussels (*Mytilus edulis*).', *Journal of the National Cancer Institute*, 43(2), pp. 509–516.

Farley, C. A., Plutschak, D. L. and Scott, R. F. (1991) 'Epizootiology and distribution of transmissible sarcoma in Maryland softshell clams, Mya arenaria, 1984-1988', *Environmental Health Perspectives*, 90, pp. 35–41. doi: 10.2307/3430843.

Farley, C. a, Otto, S. V and Reinisch, C. L. (1986) 'New occurrence of epizootic sarcoma in Chesapeake Bay soft shell clams, *Mya arenaria'*, *Fishery Bulletin*, 84(4), pp. 851–857.

Ferlay, J. et al. (2020) 'Global Cancer Observatory: Cancer Today', International Agency for Research on Cancer.

Freire, R. et al. (2011) 'Identification of European commercial cockles (*Cerastoderma edule* and *C. glaucum*) by species-specific PCR amplification of the ribosomal DNA ITS region', *European Food Research and Technology*, 232(1), pp. 83–86. doi: 10.1007/s00217-010-1369-5.

Fu, L. et al. (2012) 'CD-HIT: Accelerated for clustering the next-generation sequencing data', *Bioinformatics*, 28(23), pp. 3150–3152. doi: 10.1093/bioinformatics/bts565.

Galimany, E. and Sunila, I. (2008) 'Several cases of disseminated neoplasia in mussels *Mytilus edulis* (L.) in western Long Island Sound', *Journal of Shellfish Research*, 27(5), pp. 1201–1207. doi: 10.2983/0730-8000-27.5.1201.

Gandhi, M. J. and Strong, D. M. (2007) 'Donor derived malignancy following transplantation: a review', *Cell Tissue Banking*, 8, pp. 267–286. doi: 10.1007/s10561-007-9036-1.

Garcia-Souto, D. et al. (2022) 'Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe', *eLife*, 11, pp. 1–20. doi: 10.7554/eLife.66946.

García-Souto, D. et al. (2015) 'Divergent evolutionary behavior of H3 histone gene and rDNA clusters in venerid clams', *Molecular Cytogenetics*, 8(1), pp. 1–10. doi: 10.1186/s13039-015-0150-7.

García-Souto, D., Qarkaxhija, V. and Pasantes, J. J. (2017) 'Resolving the Taxonomic Status of *Chamelea gallina* and *C. striatula* (Veneridae, Bivalvia): A Combined Molecular Cytogenetic and Phylogenetic Approach', *BioMed Research International*, 2017. doi: 10.1155/2017/7638790.

Gartner, H. V et al. (1996) 'Genetic Analysis of a Sarcoma Accidentally Transplanted from a Patient to a Surgeon', *The New England Journal of Medicine*, 335(20), pp. 1494–1496.

Geoghegan, P. et al. (2021) 'Long-Term Trends in the Occurrence of Disseminated Neoplasia in a Population of *Mya arenaria* (Softshell Clam) from a New Hampshire Estuary', *Northeastern Naturalist*, 28(2), pp. 137–148. doi: 10.1656/045.028.0203.

Gerdol, M. et al. (2018) 'Immunity in Molluscs: Recognition and Effector Mechanisms, with a Focus on Bivalvia', *Advances in Comparative Immunology*. doi: 10.1007/978-3-319-76768-0.

Giersch, R. M. et al. (2022) 'Survival and Detection of Bivalve Transmissible Neoplasia from the Soft-Shell Clam *Mya arenaria* (MarBTN) in Seawater', Pathogens, 11(3), pp. 1–14. doi: 10.3390/pathogens11030283.

Gillies, R. J., Verduzco, D. and Gatenby, R. A. (2012) 'Evolutionary dynamics of carcinogenesis and why targeted therapy does not work', *Nature Reviews Cancer*, 12(7), pp. 487–493. doi: 10.1038/nrc3298.

Goffeau, A. et al. (1996) 'Life with 6000 Genes', *Science*, 274(October), pp. 546–567.

Gomes-dos-Santos, A. et al. (2020) 'Molluscan genomics: the road so far and the way forward', *Hydrobiologia*, 847(7), pp. 1705–1726. doi: 10.1007/s10750-019-04111-1.

Gosling, E. (1992) 'Systematics and geographic distribution of *Mytilus* spp.'. In: Gosling EM (ed) 'The mussel Mytilus: ecology, physiology, genetics and culture'. Elsevier, Amsterdam.

Gosling, E. (2015) 'Marine Bivalve Molluscs'. Wiley, 2nd edition. ISBN:9780470674949 doi: 10.1002/9781119045212.

Le Grand, F. et al. (2010) 'Prevalence, intensity, and aneuploidy patterns of disseminated neoplasia in cockles (*Cerastoderma edule*) from Arcachon Bay: Seasonal variation and position in sediment', *Journal of Invertebrate Pathology*, 104(2), pp. 110–118. doi: 10.1016/j.jip.2010.01.011.

Greaves, M. F. et al. (2003) 'Leukemia in twins: Lessons in natural history', *Blood*, 102(7), pp. 2321–2333. doi: 10.1182/blood-2002-12-3817.

Greaves, M. and Hughes, W. (2018) 'Cancer cell transmission via the placenta', *Evolution, Medicine and Public Health*, 2018(1), pp. 106–115. doi: 10.1093/emph/eoy011.

Grossmann, I. and Klotzbach, P. J. (2009) 'A review of North Atlantic modes of natural variability and their driving mechanisms', *Journal of Geophysical Research Atmospheres*, 114(24), pp. 1–14. doi: 10.1029/2009JD012728.

Guan, X. (2015) 'Cancer metastases: Challenges and opportunities', *Acta Pharmaceutica Sinica B*, 5(5), pp. 402–418. doi: 10.1016/j.apsb.2015.07.005.

Gugel, E. A. and Sanders, M. E. (1986) 'Needle-stick transmission of human colonic adenocarcinoma', *The New England Journal of Medicine*, 315(23), p. 1487.

Guo, L. T. et al. (2015) 'Flow cytometry and K-mer analysis estimates of the genome sizes of *Bemisia tabaci* B and Q (Hemiptera: Aleyrodidae)', *Frontiers in Physiology*, 6(MAY), pp. 1–7. doi: 10.3389/fphys.2015.00144.

Haber, D. A. and Fearon, E. R. (1998) 'The promise of cancer genetics', *Lancet*, 351(SUPPL.2). doi: 10.1016/s0140-6736(98)90326-9.

Hahn, C., Bachmann, L. and Chevreux, B. (2013) 'Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - A baiting and iterative mapping approach', *Nucleic Acids Research*, 41(13). doi: 10.1093/nar/gkt371.

Hammel, M. et al. (2021) 'Prevalence and polymorphism of a mussel transmissible cancer in Europe', *Molecular Ecology*, pp. 1–16. doi: 10.1111/mec.16052.

Hammel, M. (2022) 'Biologie, écologie et évolution du cancer transmissible MtrBTN2 des moules *Mytilus* sp'. *Thesis* in French supervised by G. Charrière and N. Bierne. Available at: https://tel.archives-ouvertes.fr/tel-03604679.

Hanahan, D. and Weinberg, R. A. (2000) 'The Hallmarks of Cancer', *Cell*, 100, pp. 57–70.

Hansemann, D. Von (1890) 'Ueber asymmetrische Zelltheilung in Epithelkrebsen und deren biolo- gische Bedeutung.', *Virchows Arch A Pathol Anat Histol*, 19, pp. 299–326.

Hawkins, C. E. et al. (2006) 'Emerging disease and population decline of an island endemic, the Tasmanian devil *Sarcophilus harrisii*', *Biological Conservation*, 131(2), pp. 307–324. doi: 10.1016/j.biocon.2006.04.010.

Hawkins, C. E. et al. (2008) 'Tasmanian Devil: *Sarcophilus harrisii*.', *The IUCN Red List of Threatened Species*, e.T40540A1. doi: https://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T40540A10331066.en.

He, K. et al. (2016) 'Genome sizes of nine insect species determined by flow cytometry and k-mer analysis', *Frontiers in Physiology*, 7(NOV), pp. 1–7. doi: 10.3389/fphys.2016.00569.

Helm, M. M. and Bourne, N. (2004) 'Basic bivalve biology: taxonomy, anatomy and life history', in 'Hatchery culture of bivalves: a practical manual'. *Food and Agriculture Organization of the United Nations*, p. 471. doi: https://www.fao.org/3/y5720e/y5720e00.htm.

Hendrick, M. J. (2017) 'Mesenchymal Tumors of the Skin and Soft Tissues', *Tumors in Domestic Animals*, pp. 142–175. doi: 10.1002/9781119181200.ch5.

Heng, H. H. Q. et al. (2011) 'Evolutionary Mechanisms and Diversity in Cancer', *Advances in Cancer Research*. Elsevier Inc. doi: 10.1016/B978-0-12-387688-1.00008-9.

Hong, D. et al. (2008) 'Initiating and Cancer-Propagating Cells in TEL-AML1–Associated Childhood Leukemia', *Science*, 319.

House, M. and Elston, R. (2006) 'Disseminated Neoplasia of Bivalve Molluscs', *Aqua technics*, pp. 1–12.

House, M., Kim, C. and Reno, P. (1998) 'Soft shell clams *Mya arenaria* with disseminated neoplasia demonstrate reverse transcriptase activity', *Diseases of Aquatic Organisms*, 34(3), pp. 187–192.

Howe, K. et al. (2013) 'The zebrafish reference genome sequence and its relationship to the human genome', *Nature*, 496(7446), pp. 498–503. doi: 10.1038/nature12111.

Huelsenbeck, J. P. and Ronquist, F. (2001) 'MRBAYES: Bayesian inference of phylogenetic trees', *Bioinformatics*, 17(8), pp. 754–755. doi: 10.1093/bioinformatics/17.8.754.

Hummel, H., Wolowicz, M. and Bogaards, R. H. (1994) 'Genetic variability and relationships for populations of *Cerastoderma edule* and of the *C. glaucum* complex', *Netherlands Journal of Sea Research*, 33(1), pp. 81–89. doi: 10.1016/0077-7579(94)90053-1.

Hunter, K. W. et al. (2018) 'Genetic insights into the morass of metastatic heterogeneity', *Nature reviews Cancer*, 18(4), pp. 211–223. doi: 10.1038/nrc.2017.126.

Jin, Y. and Brown, R. P. (2018) 'Partition number, rate priors and unreliable divergence times in Bayesian phylogenetic dating', *Cladistics*, 34(5), pp. 568–573. doi: 10.1111/cla.12223.

Jombart, T. and Ahmed, I. (2011) 'adegenet 1.3-1: New tools for the analysis of genome-wide SNP data', *Bioinformatics*, 27(21), pp. 3070–3071. doi: 10.1093/bioinformatics/btr521.

Jones, M. E. et al. (2019) 'Emergence, transmission and evolution of an uncommon enemy: Tasmanian devil facial tumour disease', in 'Wildlife Disease Ecology: Linking Theory to Data and Application', pp. 321–341. doi: 10.1017/9781316479964.011.

Ju, Y. S. (2021) 'A large-scale snapshot of intratumor heterogeneity in human cancer', *Cancer Cell*, 39(4), pp. 463–465. doi: 10.1016/j.ccell.2021.03.005.

Kamenev, G. M. (2013) 'Species composition and distribution of bivalves in bathyal and abyssal depths of the Sea of Japan', Deep-Sea Research Part II: Topical Studies in Oceanography, 86–87, pp. 124–139. doi: 10.1016/j.dsr2.2012.08.004.

Kater, B. J., Geurts Van Kessel, A. J. M. and Baars, J. J. M. D. (2006) 'Distribution of cockles *Cerastoderma edule* in the Eastern Scheldt: Habitat mapping with abiotic variables', *Marine Ecology Progress Series*, 318, pp. 221–227. doi: 10.3354/meps318221.

Katzir, N. et al. (1987) 'Common origin of transmissible venereal tumors (TVT) in dogs.', *Oncogene*, 1(4), pp. 445–448.

Kelley, J. L. et al. (2016) 'The life aquatic: advances in marine vertebrate genomics', *Nature Reviews Genetics*, 17(9), pp. 523–534. doi: 10.1038/nrg.2016.66.

Koehn, R. (1991) 'The genetics and taxonomy of species in the genus Mytilus.', *Aquaculture*, 94, pp. 125–145.

Kofler, R., Schlötterer, C. and Lelley, T. (2007) 'SciRoKo: A new tool for whole genome microsatellite search and investigation', *Bioinformatics*, 23(13), pp. 1683–1685. doi: 10.1093/bioinformatics/btm157.

Kõressaar, T. et al. (2018) 'Primer3-masker: Integrating masking of template sequence with primer design software', *Bioinformatics*, 34(11), pp. 1937–1938. doi: 10.1093/bioinformatics/bty036.

Kosyan, A. R. and Divinsky, B. V (2019) '*Chamelea gallina* in the coastal waters of the Anapa Bay bar (the Black Sea) as a carbonate sediment producer', *Oceanologia*, 61(4), pp. 471–483. doi: https://doi.org/10.1016/j.oceano.2019.04.003.

Kozlov, A. M. et al. (2019) 'RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference', *Bioinformatics*, 35(21), pp. 4453–4455. doi: 10.1093/bioinformatics/btz305.

Krakau, M. et al. (2012) 'The cockle *Cerastoderma edule* at Northeast Atlantic shores: Genetic signatures of glacial refugia', *Marine Biology*, 159(1), pp. 221–230. doi: 10.1007/s00227-011-1802-8.

Kreiss, A. et al. (2011) 'Allorecognition in the tasmanian devil (*Sarcophilus harrisii*), an endangered marsupial species with limited genetic diversity', *PLoS ONE*, 6(7), pp. 1–8. doi: 10.1371/journal.pone.0022402.

Kuo, C. J. and Curtis, H. (2018) 'Organoids reveal cancer dynamics Transition states that allow cancer to spread', *Nature*, 556, pp. 441–442.

Lander, E. S. et al. (2001) 'Initial sequencing and analysis of the human genome', *Nature*, 409(6822), pp. 860–921. doi: 10.1038/35057062.

Landsberg, J. H. (1996) 'Neoplasia and biotoxins in bivalves: Is there a connection?', *Journal of Shellfish Research*, 15(2), pp. 203–230.

Larman, T. C. et al. (2012) 'Spectrum of somatic mitochondrial mutations in five cancers', *Proceedings of the National Academy of Sciences of the United States of America*, 109(35), pp. 14087–14091. doi: 10.1073/pnas.1211502109.

Lazenby, B. T. et al. (2018) 'Density trends and demographic signals uncover the long-term impact of transmissible cancer in Tasmanian devils', *Journal of Applied Ecology*, 55(3), pp. 1368–1379. doi: 10.1111/1365-2664.13088.

Leitão, A. et al. (2008) 'Supernumerary chromosomes on Southern European populations of the cockle *Cerastoderma edule*: Consequence of environmental pollution?', Estuarine, Coastal and Shelf Science, 79(1), pp. 152–156. doi: 10.1016/j.ecss.2008.03.015.

Li, D. et al. (2015) 'MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph', *Bioinformatics*, 31(10), pp. 1674–1676. doi: 10.1093/bioinformatics/btv033.

Li, H. et al. (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Li, H. (2013) 'Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM', *Arxiv*, pp. 1–3. Available at: http://arxiv.org/abs/1303.3997.

Li, H. (2016) 'Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences', *Bioinformatics*, 32(14), pp. 2103–2110. doi: 10.1093/bioinformatics/btw152.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics*, 25(14), pp. 1754–1760. doi: 10.1093/bioinformatics/btp324.

Lobo, I. (2008) 'Chromosome abnormalities and cancer cytogenetics', *Nature Education*, 1(1), p. 68.

Loeb, L. A. and Harris, C. C. (2008) 'Advances in chemical carcinogenesis: A historical review and prospective', *Cancer Research*, 68(17), pp. 6863–6872. doi: 10.1158/0008-5472.CAN-08-2852.

Longshaw, M. and Malham, S. K. (2013) 'A review of the infectious agents, parasites, pathogens and commensals of European cockles (*Cerastoderma edule* and *C. glaucum*)', *Journal of the Marine Biological Association of the United Kingdom*, 93(1), pp. 227–247. doi: 10.1017/s0025315412000537.

Lopez, J. V. et al. (2019) 'Multiple Facets of Marine Invertebrate Conservation Genomics', *Annual Review of Animal Biosciences*, 7, pp. 473–497. doi: 10.1146/annurev-animal-020518-115034.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), pp. 1–21. doi: 10.1186/s13059-014-0550-8.

Lowe, D. M. and Moore, M. N. (1978) 'Cytology and quantitative cytochemistry of a poliferative atypical hemocytic condition in *Mytilus edulis* (Bivalvia, mollusca).', *Journal of the National Cancer Institute*, 60(6), pp. 1455–1459. doi: 10.1093/jnci/60.6.1455.

Lucentini, L. et al. (2020) 'Additional taxonomic coverage of the doubly uniparental inheritance in bivalves: Evidence of sex-linked heteroplasmy in the razor clam *Solen marginatus* Pulteney, 1799, but not in the lagoon cockle *Cerastoderma glaucum* (Bruguière, 1789)', *Journal of Zoological Systematics and Evolutionary Research*, 58(2), pp. 561–570. doi: 10.1111/jzs.12386.

Lugassy, C. et al. (2004) 'Pericyte-Like Location of GFP-Tagged Melanoma Cells: Ex Vivo and in Vivo Studies of Extravascular Migratory Metastasis', *American Journal of Pathology*, 164(4), pp. 1191–1198. doi: 10.1016/S0002-9440(10)63207-5.

Maia, F., Barroso, C. M. and Gaspar, M. B. (2021) 'Biology of the common cockle *Cerastoderma edule* (Linnaeus, 1758) in Ria de Aveiro (NW Portugal): Implications for fisheries management', *Journal of Sea Research*, 171, p. 102024. doi: 10.1016/j.seares.2021.102024.

Manso, C. F. et al. (2012) 'Detection of reverse transcriptase activity in golden carpet shell clams (*Venerupis aurea*) with disseminated neoplasia', *Bulletin of the European Association of Fish Pathologists*, 32(2), pp. 56–63.

Marchesi, F. et al. (2010) 'Molecular mechanisms of perineural invasion, a forgotten pathway of dissemination and metastasis', *Cytokine and Growth Factor Reviews*, 21(1), pp. 77–82. doi: 10.1016/j.cytogfr.2009.11.001.

Margres, M. J. et al. (2018) 'Large-effect loci affect survival in Tasmanian devils (*Sarcophilus harrisii*) infected with a transmissible cancer', *Molecular Ecology*, 27(21), pp. 4189–4199. doi: 10.1111/mec.14853.

Margres, M. J. et al. (2020) 'Spontaneous tumor regression in tasmanian devils associated with RASL11A activation', *Genetics*, 215(4), pp. 1143–1152. doi: 10.1534/genetics.120.303428.

Martcheva, M. and Pilyugin, S. S. (2006) 'The role of coinfection in multidisease dynamics', SIAM Journal on Applied Mathematics, 66(3), pp. 843–872. doi: 10.1137/040619272.

Martínez-Castro, C. and Vázquez, E. (2012) 'Reproductive cycle of the cockle *Cerastoderma edule* (Linnaeus 1758) in the Ría de Vigo (Galicia, Northwest Spain)', *Journal of Shellfish Research*, 31(3), pp. 757–767. doi: 10.2983/035.031.0320.

Martínez, L. et al. (2009) 'Development of twelve polymorphic microsatellite markers in the edible cockle cerastoderma edule (Bivalvia: Cardiidae)', Conservation Genetics Resources, 1(1), pp. 107–109. doi: 10.1007/s12686-009-9026-7.

Martínez, L. et al. (2013) 'Genetic diversity and population differentiation in the cockle Cerastoderma edule estimated by microsatellite markers', Helgoland Marine Research, 67(1), pp. 179–189. doi: 10.1007/s10152-012-0314-3.

Martínez, L. et al. (2015) 'Patterns of genetic variation across the distribution range of the cockle *Cerastoderma edule* inferred from microsatellites and mitochondrial DNA', *Marine Biology*. doi: 10.1007/s00227-015-2676-y.

Mateo, D. R., MacCallum, G. S. and Davidson, J. (2016) 'Field and laboratory transmission studies of haemic neoplasia in the soft-shell clam, *Mya arenaria*, from Atlantic Canada', *Journal of Fish Diseases*, 39(8), pp. 913–927. doi: 10.1111/jfd.12426.

Matser, Y. A. H. et al. (2018) 'Transmission of breast cancer by a single multiorgan donor to 4 transplant recipients', *American Journal of Transplant*, 18, pp. 1810–1814. doi: 10.1111/ajt.14766.

Mayrand, E., St-Jean, S. D. and Courtenay, S. C. (2005) 'Haemocyte responses of blue mussels (*Mytilus edulis* L.) transferred from a contaminated site to a reference site: Can the immune system recuperate?', *Aquaculture Research*, 36(10), pp. 962–971. doi: 10.1111/j.1365-2109.2005.01302.x.

McGranahan, N. and Swanton, C. (2017) 'Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future', Cell, 168(4), pp. 613–628. doi: 10.1016/j.cell.2017.01.018.

Meirmans, P. G. (2015) 'Seven common mistakes in population genetics and how to avoid them', *Molecular Ecology*, 24(13), pp. 3223–3231. doi: 10.1111/mec.13243.

Metzger, M. J. et al. (2015) 'Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams', *Cell*, 161(2), pp. 255–263. doi: 10.1016/j.cell.2015.02.042.

Metzger, M. J. et al. (2016) 'Widespread transmission of independent cancer lineages within multiple bivalve species', *Nature*, pp. 1–11. doi: 10.1038/nature18599.

Metzger, M. J. and Goff, S. P. (2016) 'A Sixth Modality of Infectious Disease: Contagious Cancer from Devils to Clams and Beyond', *PLoS Pathogens*, 12(10), pp. 1–7. doi: 10.1371/journal.ppat.1005904.

Miccoli, A. et al. (2021) 'Evolution of immune defence responses as incremental layers among Metazoa', *European Zoological Journal*, 88(1), pp. 44–57. doi: 10.1080/24750263.2020.1849435.

Michnowska, A. et al. (2022) 'Horizontal transmission of disseminated neoplasia in the widespread clam *Macoma balthica* from the Southern Baltic Sea', *Molecular Ecology,* 31(11):3128-3136. doi: 10.1111/mec.16464.

Miller, W. et al. (2011) 'Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil)', Proceed*ings of the National Academy of Sciences of the United States of America*, 108(30), pp. 12348–12353. doi: 10.1073/pnas.1102838108.

Minh, B. et al. (2020) 'IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era', *Mol. Biol. Evol.*, 37(5), pp. 1530–1534. doi: 10.1093/molbev/msaa015.

Mitelman, F., Obe, G. and Natarajan, A. T. (1990) 'Patterns of Chromosome Variation in Neoplasia', in 'Chromosomal Aberrations: Basic and Applied Aspects' 1st edition. Berlin Heidelberg: Springer-Verlag, p. 327. doi: 10.1007/978-3-642-75682-5.

Miura, S. et al. (2018) 'Predicting clone genotypes from tumor bulk sequencing of multiple samples', *Bioinformatics*, 34(23), pp. 4017–4026. doi: 10.1093/bioinformatics/bty469.

Mix, M. C., Hawkes, J. W. and Sparks, A. K. (1979) 'Observations on the ultrastructure of large cells associated with putative neoplastic disorders of mussels, *Mytilus edulis*, from Yaquina Bay, Oregon', *Journal of Invertebrate Pathology*, 34(1), pp. 41–56. doi: 10.1016/0022-2011(79)90052-1.

Montaudouin, X. de et al. (2021) 'Catalogue of parasites and diseases of the common cockle *Cerastoderma edule*'. 1st edition. UA Editora – Universidade de Aveiro. doi: 10.34624/9a9c-9j21.

Moore, J. D. et al. (1991) 'Alternate pathogenesis of systemic neoplasia in the bivalve mollusc Mytilus', *Journal of invertebrate pathology*, 58(2), pp. 231–243. doi: 10.1016/0022-2011(91)90067-z.

Muehlenbachs, A. et al. (2015) 'Malignant Transformation of Hymenolepis nana in a Human Host', *The New England Journal of Medicine*, 373(19), pp. 1845–1852. doi: 10.1056/NEJMoa1505892.

Mukherjee, S. (2010) 'The Emperor Of All Maladies: A Biography Of Cancer'. 1st ed. New York: Simon & Schuster, Inc.

Mun, S. et al. (2017) 'The whole-genome and transcriptome of the Manila clam (*Ruditapes philippinarum*)', *Genome Biology and Evolution*, 9(6), pp. 1487–1498. doi: 10.1093/gbe/evx096.

Murchison, E. P. (2008) 'Clonally transmissible cancers in dogs and Tasmanian devils', *Oncogene*, 27(S2), pp. S19–S30. doi: 10.1038/onc.2009.350.

Murchison, E. P. et al. (2010) 'The Tasmanian Devil Transcriptome Reveals Schwann Cell Origins of a Clonally Transmissible Cancer', *Science*, 327(5961), pp. 84–87. doi: 10.1126/science.1180616.

Murchison, E. P. et al. (2012) 'Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer', *Cell*, 148(4), pp. 780–791. doi: 10.1016/j.cell.2011.11.065.

Murchison, E. P. et al. (2014) 'Transmissible dog cancer genome reveals the origin and history of an ancient cell lineage.', *Science*, 343(6169), pp. 437–40. doi: 10.1126/science.1247167.

Murchison, E. P. (2016) 'Cancer: Transmissible tumours under the sea', *Nature*, 534(7609), pp. 628–629. doi: 10.1038/nature18455.

Murgarella, M. et al. (2016) 'A first insight into the genome of the filter-feeder mussel *Mytilus galloprovincialis'*, *PLoS ONE*, 11(3), pp. 1–22. doi: 10.1371/journal.pone.0151561.

Murgia, C. et al. (2006) 'Clonal Origin and Evolution of a Transmissible Cancer', *Cell*, 126, pp. 477–487. doi: 10.1016/j.cell.2006.05.051.

Murgia, C. (2006) 'Pathogenesis and evolution of Canine Transmissible Venereal Tumour'. *Thesis* at University College of London. Available at: https://discovery.ucl.ac.uk/id/eprint/1446029/

Muttray, A. F. et al. (2010) 'An invertebrate mdm homolog interacts with p53 and is differentially expressed together with p53 and ras in neoplastic *Mytilus trossulus* haemocytes', *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 156(4), pp. 298–308. doi: 10.1016/j.cbpb.2010.04.008.

Muttray, A. F., Schulte, P. M. and Baldwin, S. A. (2008) 'Invertebrate p53-like mRNA isoforms are differentially expressed in mussel haemic neoplasia', *Marine Environmental Research*, 66(4), pp. 412–421. doi: 10.1016/j.marenvres.2008.06.004.

Muttray, A. F. and Vassilenko, K. (2018) 'Mollusca: Disseminated Neoplasia in Bivalves and the p53 Protein Family', *Advances in Comparative Immunology*, pp. 773–791. doi: 10.1007/978-3-319-76768-0.

Nancy, P. et al. (2012) 'Chemokine gene silencing in decidual stromal cells limits T cell access to the maternal-fetal interface', *Science*, 336(6086), pp. 1317–1321. doi: 10.1126/science.1220030.

Nik-Zainal, S. et al. (2012) 'The life history of 21 breast cancers', *Cell*, 149(5), pp. 994–1007. doi: 10.1016/j.cell.2012.04.023.

Van Nimwegen, K. J. M. et al. (2016) 'Is the 1000 genome as near as we think? A cost analysis of next-generation sequencing', *Clinical Chemistry*, 62(11), pp. 1458–1464. doi: 10.1373/clinchem.2016.258632.

Novák, P., Neumann, P. and Macas, J. (2010) 'Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data', *BMC Bioinformatics*, 11, pp. 1–12. doi: 10.1186/1471-2105-11-378.

Novinski, M. A. (1876) 'Zur Frage uber die Impfung der Krebsigen Geschwulste', *Zentralbl. Med. Wissensch.*, 14, pp. 90–791.

Nurk, S. et al. (2022) 'The complete sequence of a human genome', *Science*, 53, pp. 44–53. doi: 10.1126/science.abj6987

Odintsova, N. A. (2020) 'Leukemia-Like Cancer in Bivalves', *Russian Journal of Marine Biology*, 46(2), pp. 59–67. doi: 10.1134/S1063074020020078.

Ogrodowczyk, I. (2017) 'Neoplasia in bivalves from Gulf of Gdansk: comparative study'. Master thesis, Faculty of Oceanography and Geography of University of Gdansk, Gdynia.

Okonechnikov, K., Conesa, A. and García-Alcalde, F. (2016) 'Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data', *Bioinformatics*, 32(2), pp. 292–294. doi: 10.1093/bioinformatics/btv566.

Oprandy, J. J. et al. (1981) 'Isolation of a viral agent causing hematopoietic neoplasia in the soft-shell clam, *Mya arenaria*', *Journal of Invertebrate Pathology*, 38(1), pp. 45–51. doi: 10.1016/0022-2011(81)90033-1.

Oprandy, J. J. and Chang, P. W. (1983) '5-Bromodeoxyuridine induction of hematopoietic neoplasia and retrovirus activation in the soft-shell clam, *Mya arenaria*', *Journal of Invertebrate Pathology*, 42(2), pp. 196–206. doi: 10.1016/0022-2011(83)90062-9.

Ordás, M. C. and Figueras, A. (2005) 'Histopathological study of abnormal mortalities of cockle (*Cerastoderma edule*) in Carril (Galicia, NW Spain)', *Bulletin of the European Association of Fish Pathologists*, 25(3), pp. 124–130.

Oshimura, M., Sasaki, M. and Makino, S. (1973) 'Chromosomal banding patterns in primary and transplanted venereal tumors of the DOG', *Journal of the National Cancer Institute*, 51(4), pp. 1197–1203. doi: 10.1093/jnci/51.4.1197.

Ostrander, E. A., Davis, B. W. and Ostrander, G. K. (2016) 'Transmissible Tumors: Breaking the Cancer Paradigm', *Trends in Genetics*, 32(1), pp. 1–15. doi: 10.1016/j.tig.2015.10.001.

Parrish, N., Hormozdiari, F. and Eskin, E. (2014) 'Assembly of non-unique insertion content using next-generation sequencing', *Bioinformatics: The Impact of Accurate Quantification on Proteomic and Genetic Analysis and Research*, pp. 21–40. doi: 10.1201/b16589.

Pearse, A.-M. and Swift, K. (2006) 'Transmission of devil facial-tumour disease', *Nature*, 439, p. 549. doi: 10.1038/439549a.

Pekkarinen, M. and Lei, H. (1994) 'Comparisons of Gymnophallid Digeneans from North Pacific and Baltic Clams, *Macoma balthica* (Bivalvia)', *The Journal of Parasitology*, 80(4), pp. 630–636.

Pendleton, M. et al. (2015) 'Assembly and diploid architecture of an individual human genome via single-molecule technologies', *Nature Methods*, 12(8), pp. 780–786. doi: 10.1038/nmeth.3454.

Peng, R. D. and Hicks, S. C. (2020) 'Reproducible Research: A Retrospective', *Annual Review of Public Health*, 42, pp. 79–93. doi: 10.1146/annurev-publhealth-012420-105110.

Peteiro, L. G. et al. (2018) 'Responses to salinity stress in bivalves: Evidence of ontogenetic changes in energetic physiology on *Cerastoderma edule*', *Scientific Rep*orts, 8(1), pp. 1–9. doi: 10.1038/s41598-018-26706-9.

Peters, E. (1988) 'Recent investigations on the disseminated sarcomas of marine bivalve molluscs.', *Am. Fish. Soc. Spec. Publ.*, 18, pp. 74–92.

Poder, M. and Auffret, M. (1986) 'Sarcomatous lesion in the cockle *Cerastoderma edule*. I. Morphology and population survey in Brittany, France', *Aquaculture*, 58(1–2), pp. 1-8. doi: 10.1016/0044-8486(86)90151-1.

Polak, P. et al. (2015) 'Cell-of-origin chromatin organization shapes the mutational landscape of cancer', *Nature*, 518(7539), pp. 360–364. doi: 10.1038/nature14221.

Popic, V. et al. (2015) 'Fast and scalable inference of multi-sample cancer lineages', *Genome Biology*, 16(1), pp. 1–17. doi: 10.1186/s13059-015-0647-8.

Poppe, G. T. and Goto, Y. (2000) 'European Seashells, Volume 2: Scaphopoda, Bivalvia, Cephalopoda, European Seashells'. Hackenheim, Germany: ConchBooks.

Pradeu, T. and Du Pasquier, L. (2018) 'Immunological memory: What's in a name?', *Immunological Reviews*, 283(1), pp. 7–20. doi: 10.1111/imr.12652.

Pye, R. J. et al. (2016) 'A second transmissible cancer in Tasmanian devils', *Proceedings of the National Academy of Sciences*, 113(2), pp. 374–379. doi: 10.1073/pnas.1519691113.

Quinteiro, J. and Rey-Mendez, M. (2017) 'Characterization of the comom cockle, *Cerastoderma edule* (Linnaeus, 1758), mitochondrial DNA', National Center for Biotechnology Information, NIH.

Rambaut, A. et al. (2018) 'Posterior summarization in Bayesian phylogenetics using Tracer 1.7', *Systematic Biology*, 67(5), pp. 901–904. doi: 10.1093/sysbio/syy032.

Ran, Z. et al. (2019) 'Chromosome-level genome assembly of the razor clam *Sinonovacula constricta* (Lamarck, 1818)', *Molecular Ecology Resources*, 19(6), pp. 1647–1658. doi: 10.1111/1755-0998.13086.

Real, L. A. and Biek, R. (2007) 'Infectious Disease Modeling and the Dynamics', *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances, and Consequences of Cross-Species Transmissions*, pp. 33–49.

Rebbeck, C. A., Leroi, A. M. and Burt, A. (2011) 'Mitochondrial Capture by a Transmissible Cancer', *Science*, 331, p. 303. doi: 10.5061/dryad.8081.C.

Reinisch, C. L., Charles, A. M. and Troutner, J. (1983) 'Unique antigens on neoplastic cells of the soft-shell clam *Mya arenaria'*, *Developmental and Comparative Immunology*, 7(1), pp. 33–39. doi: 10.1016/0145-305X(83)90052-6.

Robinson, J. T. et al. (2011) 'Integrative Genome Viewer', *Nature Biotechnology*, 29(1), pp. 24–6. doi: 10.1038/nbt.1754.Integrative.

Rodriguez-Juiz, A. M., Torrado, M. and Mendez, J. (1996) 'Genome-size variation in bivalve molluscs determined by flow cytometry', *Marine Biology*, 126(3), pp. 489–497. doi: 10.1007/BF00354631.

Rodriguez, H. et al. (1997) 'A case of disseminated haemic sarcoma in *Cerastoderma* (=Cardium) *glaucum* Poiret, 1789 collected from Ría de Vigo (NW Spain)', *Bulletin of the European Association of Fish Pathologists*, 17(3/4), pp. 88–90.

Romalde, J. L. et al. (2007) 'Evidence of retroviral etiology for disseminated neoplasia in cockles (*Cerastoderma edule*)', *Journal of Invertebrate Pathology*, 94(2), pp. 95–101. doi: 10.1016/j.jip.2006.09.007.

Rombel, I. T. et al. (2002) 'ORF-FINDER: A vector for high-throughput gene identification', *Gene*, 282(1–2), pp. 33–41. doi: 10.1016/S0378-1119(01)00819-8.

Ronquist, F. et al. (2012) 'Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space', *Systematic Biology*, 61(3), pp. 539–542. doi: 10.1093/sysbio/sys029.

Rübben, A. and Araujo, A. (2017) 'Cancer heterogeneity: converting a limitation into a source of biologic information', *Journal of Translational Medicine*, 15(190), pp. 1–10. doi: 10.1186/s12967-017-1290-9.

Rufino, M. et al. (2006) 'Use of Shape to Distinguish *Chamelea gallina* and *Chamelea striatula* (Bivalvia: Veneridae): Linear and Geometric Morphometric Methods', *Journal of Morphology*, 267, pp. 1433–1440. doi: 10.1002/jmor.

Ruiz-Aravena, M. et al. (2018) 'Sex bias in ability to cope with cancer: Tasmanian devils and facial tumour disease', *Proceedings of the Royal Society B: Biological Sciences*, 285(1891). doi: 10.1098/rspb.2018.2239.

Ruiz, P. et al. (2013) 'Biomarkers and transcription levels of cancer-related genes in cockles *Cerastoderma edule* from Galicia (NW Spain) with disseminated neoplasia', *Aquatic Toxicology*, 136–137, pp. 101–111. doi: 10.1016/j.aquatox.2013.03.018.

Ryan Poplin et al. (2018) 'Scaling accurate genetic variant discovery to tens of thousands of samples', *bioRxiv*, pp. 1–22.

Rycaj, K. and Tang, D. G. (2015) 'Cell-of-origin of cancer versus cancer stem cells: Assays and interpretations', *Cancer Research*, 75(19), pp. 4003–4011. doi: 10.1158/0008-5472.CAN-15-0798.

Salvi, D. and Mariottini, P. (2017) 'Molecular taxonomy in 2D: A novel ITS2 rRNA sequencestructure approach guides the description of the oysters' subfamily Saccostreinae and the genus Magallana (Bivalvia: Ostreidae)', Zoological Journal of the Linnean Society, 179(2), pp. 263–276. doi: 10.1111/zoj.12455.

Scanlon, E. F. et al. (1965) 'Fatal homotransplanted melanoma. A case report', *Cancer*, 18(6), pp. 782–789. doi: 10.1002/1097-0142(196506)18

Schneider, V. A. et al. (2017) 'Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly', *Genome Research*, 27(5), pp. 849–864. doi: 10.1101/gr.213611.116.

Seckl, M. J., Sebire, N. J. and Berkowitz, R. S. (2010) 'Gestational trophoblastic disease', *The Lancet*, 376(9742), pp. 717–729. doi: 10.1016/S0140-6736(10)60280-2.

Seppey, M., Manni, M. and Zdobnov, E. M. (2019) 'BUSCO: Assessing genome assembly and annotation completeness', *Methods in Molecular Biology*, 1962, pp. 227–245. doi: 10.1007/978-1-4939-9173-0_14.

Seyfried, T. N. and Huysentruyt, L. C. (2013) 'On the Origin of Cancer Metastasis', *Clinical Reviews in Oncogenesis*, 18(1–2), pp. 43–73. doi: 0893-9675/13/3500.

Siddle, H. V. et al. (2013) 'Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease illustrates immune escape by a contagious cancer', *Proceedings of the National Academy of Sciences of the United States of America*, 110(13), pp. 5103–5108. doi: 10.1073/pnas.1219920110.

Da Silva, P. M. et al. (2005) 'Flow cytometric DNA content analysis of neoplastic cells in haemolymph of the cockle *Cerastoderma edule'*, *Diseases of Aquatic Organisms*, 67(1–2), pp. 133–139. doi: 10.3354/dao067133.

Skazina, M. et al. (2021) 'First description of a widespread *Mytilus trossulus* - derived bivalve transmissible cancer lineage in *M. trossulus* itself', *Scientific Reports*, 11(5809), pp. 1–13. doi: 10.1038/s41598-021-85098-5.

Skazina, M. et al. (2022) 'Two lineages of bivalve transmissible neoplasia affect the blue mussel *Mytilus trossulus* Gould in the subarctic Sea of Okhotsk', *Current Zoology*, pp. 1–23. doi: 10.1093/cz/zoac012

Sleeman, J. P., Nazarenko, I. and Thiele, W. (2011) 'Do all roads lead to Rome? Routes to metastasis development', *International Journal of Cancer*, 128(11), pp. 2511–2526. doi: 10.1002/ijc.26027.

Smolarz, K. et al. (2005) 'Survey for neoplasia in *Macoma balthica* from the Gulf of Gdansk by flow cytometry', *Diseases of Aquatic Organisms*, 66(1), pp. 41–46. doi: 10.3354/dao066041.

Smolowitz, R. M., Miosky, D. and Reinisch, C. L. (1989) 'Ontogeny of leukemic cells of the soft-shell clam', *Journal of Invertebrate Pathology*, 53(1), pp. 41–51. doi: 10.1016/0022-2011(89)90072-4.

Smolowitz, R. M. and Reinisch, C. L. (1986) 'Indirect peroxidase staining using monoclonal antibodies specific for *Mya arenaria* neoplastic cells', *Journal of Invertebrate Pathology*, 48(2), pp. 139–145. doi: 10.1016/0022-2011(86)90115-1.

Soudant, P., E. Chu, F. L. and Volety, A. (2013) 'Host-parasite interactions: Marine bivalve molluscs and protozoan parasites, Perkinsus species', *Journal of Invertebrate Pathology*, 114(2), pp. 196–216. doi: 10.1016/j.jip.2013.06.001.

Sousos, N. et al. (2022) 'In utero origin of myelofibrosis presenting in adult monozygotic twins', *Nature medicine*. doi: 10.1038/s41591-022-01793-4.

Sporn, M. B. (1996) 'The war on cancer', *The Lancet*, 347, pp. 1377–1381.

Stammnitz, M. R. et al. (2018) 'The Origins and Vulnerabilities of Two Transmissible Cancers in Tasmanian Devils', *Cancer Cell*, 33(4), pp. 607-619.e15. doi: 10.1016/j.ccell.2018.03.013.

Star, B. et al. (2011) 'The genome sequence of Atlantic cod reveals a unique immune system', *Nature*, 477(7363), pp. 207–210. doi: 10.1038/nature10342.

Strakova, A. et al. (2016) 'Mitochondrial genetic diversity, selection and recombination in a canine transmissible cancer', *eLife*, 5(MAY2016), pp. 1–25. doi: 10.7554/eLife.14552.

Strakova, A. (2017) 'Genome Diversity and Evolution in Canine Transmissible Venereal Tumour'. Department of Veterinary Medicine. University of Cambridge. *Thesis* supervised by Elizabeth Murchinson.

Strakova, A. et al. (2020) 'Recurrent horizontal transfer identifies mitochondrial positive selection in a transmissible cancer', *Nature Communications*, 11(1), pp. 1–11. doi: 10.1038/s41467-020-16765-w.

Strakova, A. and Murchison, E. P. (2014) 'The changing global distribution and prevalence of canine transmissible venereal tumour', *BMC Veterinary Research*, 10(1), pp. 1–10. doi: 10.1186/s12917-014-0168-9.

Strakova, A. and Murchison, E. P. (2015) 'The cancer which survived: Insights from the genome of an 11,000 year-old cancer', *Current Opinion in Genetics and Development*, 30, pp. 49–55. doi: 10.1016/j.gde.2015.03.005.

Stratton, M. R. (2013) 'Journeys into the genome of cancer cells', *EMBO Mol Med*, 5, pp. 169–172.

Stratton, M. R., Campbell, P. J. and Futreal, P. A. (2009) 'The cancer genome', *Nature*, 458(7239), pp. 719–724. doi: 10.1038/nature07943.

Swofford, D. L. (2002) 'PAUP: phylogenetic analysis using parsimony, version 4.0 b10.' Sinauer Associates. doi: 10.1111/j.0014-3820.2002.tb00191.x

Takeuchi, T. et al. (2012) 'Draft genome of the pearl oyster *Pinctada fucata*: A platform for understanding bivalve biology', *DNA Research*, 19(2), pp. 117–130. doi: 10.1093/dnares/dss005.

Tan, A. S. et al. (2015) 'Mitochondrial genome acquisition restores respiratory function and tumorigenic potential of cancer cells without mitochondrial DNA', *Cell Metabolism*, 21(1), pp. 81–94. doi: 10.1016/j.cmet.2014.12.003.

Tebble, N. (1976) 'British bivalve seashells; a handbook for identification', Royal Scottish Museum, 212 pages.

Thai, B. T. et al. (2019) 'Whole Genome Assembly of the Snout Otter Clam, *Lutraria rhynchaena*, Using Nanopore and Illumina Data, Benchmarked Against Bivalve Genome Assemblies', *Frontiers in Genetics*, 10, pp. 1–8. doi: 10.3389/fgene.2019.01158.

Thiriot-Quiévreux, C. and Wolowicz, M. (1996) 'Karyologic study of gill neoplasia in *Macoma balthica* (Mollusca, Bivalvia).', *Comptes rendus de l'Academie des sciences*. Serie III, Sciences de la vie, 319(10), pp. 887–892.

Tirado, C., Salas, C. and Márquez, I. (2003) 'Reproduction of *Venus verrucosa* L., 1758 (Bivalvia: Veneridae) in the littoral of Málaga (southern Spain)', *Fisheries Research*, 63(3), pp. 437–445. doi: 10.1016/S0165-7836(03)00106-1.

Tischler, G. and Leonard, S. (2014) 'Biobambam: Tools for read pair collation-based algorithms on BAM files', *Source Code for Biology and Medicine*, 9(1), pp. 1–18. doi: 10.1186/1751-0473-9-13.

Tolar, J. and Neglia, J. P. (2003) 'Transplacental and Other Routes of Cancer Transmission Between Individuals', *Journal of Pediatric Hematology/Oncology*, 25(6), pp. 430–434.

Trigo, J. et al. (2018) 'Guía de los Moluscos Marinos de Galicia'. 1st edition. Servizo de Publicacións da Universidade de Vigo, Spain.

Twomey, E. and Mulcahy, M. (1984) 'A proliferative disorder of possible haemic origin in the common cockle, Cerasto- derma edule', *Journal of Invertebrate Pathology*, 44, pp. 109–111.

Twomey, E. and Mulcahy, M. (1988) 'Epizootiological aspects of a sarcoma in the cockle *Cerastoderma edule'*, *Diseases of Aquatic Organisms*, 5, pp. 225–238. doi: 10.3354/dao005225.

Twomey and Mulcahy, M. F. (1988) 'Transmission of a sarcoma in the cockle *Cerastoderma edule* (Bivalvia; Mollusca) using cell transplants', *Developmental and Comparative Immunology*, 12(1), pp. 195–200. doi: 10.1016/0145-305X(88)90037-7.

Tyler-Walters, H. (2007) '*Cerastoderma edule* Common cockle'. In Tyler-Walters H. and Hiscock K. (eds) 'Marine Life Information Network: Biology and Sensitivity Key Information Reviews', *Marine Biological Association of the United Kingdom*, pp. 1–23. Available at: https://www.marlin.ac.uk/species/detail/1384.

Ujvari, B., Gatenby, R. A. and Thomas, F. (2016) 'Transmissible cancers, are they more common than thought?', *Evolutionary Applications*, 9(5), pp. 633–634. doi: 10.1111/eva.12372.

Uliano-Silva, M. et al. (2018) 'A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, *Limnoperna fortunei'*, *GigaScience*, 7(2), pp. 1–10. doi: 10.1093/GIGASCIENCE/GIX128.

Untergasser, A. et al. (2012) 'Primer3-new capabilities and interfaces', *Nucleic Acids Research*, 40(15), pp. 1–12. doi: 10.1093/nar/gks596.

Vaser, R. et al. (2017) 'Fast and accurate de novo genome assembly from long uncorrected reads', *Genome Research*, 27(5), pp. 737–746. doi: 10.1101/gr.214270.116.

Vassilenko, E. and Baldwin, S. A. (2014) 'Using flow cytometry to detect haemic neoplasia in mussels (*Mytilus trossulus*) from the Pacific Coast of Southern British Columbia Canada.', *Journal of Invertebrate Pathology*, 117, pp. 68–72.

Veeneman, B. A. et al. (2016) 'Two-pass alignment improves novel splice junction quantification', *Bioinformatics*, 32(1), pp. 43–49. doi: 10.1093/bioinformatics/btv642.

Vera, M. et al. (2021) 'Genomic survey of edible cockle (*Cerastoderma edule*) in the Northeast Atlantic: A baseline for sustainable management of its wild resources', *Evolutionary Applications*, pp. 262–285. doi: 10.1111/eva.13340.

Villalba, A. et al. (2008) 'Workshop for the analysis of the impact of Perkinsosis to the European Shellfish Insdustry' Centro de Investigacións Mariñas, Consellería de Pesca e Asuntos Marítimos da Xunta de Galicia.

Villalba, A. et al. (2014) 'Cockle *Cerastoderma edule* fishery collapse in the Ría de Arousa (Galicia, NW Spain) associated with the protistan parasite *Marteilia cochilli*', *Diseases of Aquatic Organisms*, 109(1), pp. 55–80. doi: 10.3354/dao02723.

Villalba, A., Carballal, M. J. and López, C. (2001) 'Disseminated neoplasia and large foci indicating heavy haemocytic infiltration in cockles *Cerastoderma edule* from Galicia (NW Spain)', *Diseases of Aquatic Organisms*, 46(3), pp. 213–216. doi: 10.3354/dao046213.

Visvader, J. E. (2011) 'Cells of origin in cancer', *Nature*, 469(7330), pp. 314–322. doi: 10.1038/nature09781.

Vogt, G. (2008) 'How to minimize formation and growth of tumours: Potential benefits of decapod crustaceans for cancer research', *International Journal of Cancer*, 123(12), pp. 2727–2734. doi: 10.1002/ijc.23947.

Wagner, D. E. and Klein, A. M. (2020) 'Lineage tracing meets single-cell omics: opportunities and challenges', *Nature Reviews Genetics*, 21(7), pp. 410–427. doi: 10.1038/s41576-020-0223-2.

Walker, B. J. et al. (2014) 'Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement', *PLoS ONE*, 9(11). doi: 10.1371/journal.pone.0112963.

Walker, C. W. and Böttger, S. A. (2008) 'A naturally occurring cancer with molecular connectivity to human diseases', *Cell Cycle*, 7(15), pp. 2286–2289. doi: 10.4161/cc.6366.

Wallace, D. C. (2012) 'Mitochondria and cancer', *Nature Reviews Cancer*, 12(10), pp. 685–698. doi: 10.1038/nrc3365.

Wang, C. et al. (2021) 'A novel canine reference genome resolves genomic architecture and uncovers transcript complexity', *Communications Biology*, 4(1), pp. 1–11. doi: 10.1038/s42003-021-01698-x.

Wang, S. et al. (2017) 'Scallop genome provides insights into evolution of bilaterian karyotype and development', *Nature Ecology & Evolution*, 1(April), p. 0120. doi: 10.1038/s41559-017-0120.

Warren, D. L. et al. (2017) 'RWTY (R We There Yet): An R package for examining convergence of Bayesian phylogenetic analyses', *Molecular Biology and Evolution*, 34(4), pp. 1016–1020. doi: 10.1093/molbev/msw279.

Weber, W. T., Nowell, P. C. and Hare, W. C. D. (1965) 'Chromosome studies of a transplanted and a primary canine venereal sarcoma', *Journal of the National Cancer Institute*, 35(3), pp. 537–547. doi: 10.1093/jnci/35.3.537.

Wei, M. et al. (2020) 'Chromosome-Level Clam Genome Helps Elucidate the Molecular Basis of Adaptation to a Buried Lifestyle', *iScience*, 23(6), p. 101148. doi: 10.1016/j.isci.2020.101148.

Weinberg, R. A. (1994) 'Oncogenes and Tumor Suppressor Genes', *CA Cancer J Clin*, 44(3), pp. 160–170.

Weinstock, G. et al. (2006) 'Insights into social insects from the genome of the honeybee *Apis mellifera*', *Nature*, 443, pp. 931–949.

Welch, D. R. and Hurst, D. R. (2019) 'Defining the Hallmarks of Metastasis', *Cancer Research AACR*, 79, pp. 3011–3028. doi: 10.1158/0008-5472.CAN-19-0458.

Wertheim, J. O. et al. (2015) 'RELAX: Detecting relaxed selection in a phylogenetic framework', *Molecular Biology and Evolution*, 32(3), pp. 820–832. doi: 10.1093/molbev/msu400.

Williams, D. et al. (2013) 'Rapid quantification of sequence repeats to resolve the size, structure and contents of bacterial genomes.', *BMC genomics*, 14(1), p. 537. doi: 10.1186/1471-2164-14-537.

Wilmes, S. B. and Robins, P. (2020) 'How are cockle beds connected? Using ocean model simulations to predict larval transport', *COCKLES newsletter* 4, pp. 1–15. Available at: https://cockles-project.eu/public/uploads/downloads/8b0aad5a9b7b9a81bba683a8accadfc8.pdf.

Wolstenholme, D. R. (1992) 'Animal Mitochondrial DNA: Structure and Evolution', *International Review of Cytology*, 141(C), pp. 173–216. doi: 10.1016/S0074-7696(08)62066-5.

Wootton, E. C., Dyrynda, E. A. and Ratcliffe, N. A. (2006) 'Interaction between non-specific electrostatic forces and humoral factors in haemocyte attachment and encapsulation in the edible cockle, *Cerastoderma edule*', *Journal of Experimental Biology*, 209(7), pp. 1326–1335. doi: 10.1242/jeb.02118.

Yates, L. R. (2017) 'Intratumoral heterogeneity and subclonal diversification of early breast cancer', *Breast*, 34, pp. S36–S42. doi: 10.1016/j.breast.2017.06.025.

Yevich, P. P. and Barszcz, C. A. (1976) 'Gonadal and hematopoietic neoplasms in *Mya arenaria*', *Marine Fisheries Review*, 38(10), pp. 42–43.

Yonemitsu, M. A. et al. (2019) 'A single clonal lineage of transmissible cancer identified in two marine mussel species in South America and Europe', *eLife*, 8(e47788), pp. 1–21. doi: 10.7554/eLife.47788.

Yuan, Y. et al. (2020) 'Comprehensive molecular characterization of mitochondrial genomes in human cancers', *Nature Genetics*, 52(3), pp. 342–352. doi: 10.1038/s41588-019-0557-x.

Yunis, J. (1976) 'High resolution of Human Chromosomes', *Science*, 191, pp. 1268–1270.

Zare, H. et al. (2014) 'Inferring Clonal Composition from Multiple Sections of a Breast Cancer', *PLoS Computational Biology*, 10(7). doi: 10.1371/journal.pcbi.1003703.

Zhang, G. et al. (2012) 'The oyster genome reveals stress adaptation and complexity of shell formation', *Nature*, 490(7418), pp. 49–54. doi: 10.1038/nature11413.

List of Figures

List of Tables

Appendix A: Supplementary material

Research chapters of this thesis (*i.e.*, 2, 3 and 4) have additional supplementary material such as big size sample tables or summary videos. The following pages contain such information.

Chapter II:

- Sampling summary table (Table 19).
- Schematic workflow of sample processing (Figure 69).
- Schematic compendium of rules for the biobank of scuba cancers (Figure 70).
- VAF plots of tumour mitogenomes (Figure 71).
- Phylogenies:
    - Whole dataset (tumours, matched-normals, healthy):
        - Maximum-likelihood with RaxML (Figure 72).
        - Bayesian inference with BEAST (Figure 73) or MrBayes (Figure 74).
    - Only healthy cockles:
        - Maximum-likelihood with RaxML (Figure 75).
        - Bayesian inference with BEAST (Figure 76).
- Microsatellite amplifications (Figure 77).
- Extension of copy number amplifications on cockle transmissible cancers (Figure 78).
- Nomenclature of mitochondrial horizontal transfers.
- Piecharts of mitochondrial cancer lineages by sampling locations or areas (Figure 71).

Chapter III:

- VAF plots and phylogeny of HN samples sequenced for the RNA analysis (Figure 81).

Chapter IV:

- Video 1.
- eLife digest. Summary cutting jargon and putting research in context.

Table 19. Sampling summary. Locations and year of sampling collection of common cockles (*Cerastoderma edule*) are described in columns 1-4; sample storage codes in columns 5-6; HN reports, and total number of samples screened in columns 7-9; knowledge that *C. glaucum* samples were included in that collection in column 10.

Country	Location	GPS coordinates		Sampling year	Sampling code	Sample number codes	HN found	HN samples*	Processed Samples	*C. glaucum* found
Denmark	Nykobing Mors	56°52'25.9"N	8°58'00.6"E	2017	DNCE	3608 to 3706; 4359 to 4499	no	0	240	no
Denmark	Veno Limfjorden	56°33'17.6"N	8°40'28.1"E	2019	DVCE	2001 to 2030	no	0	30	no
Denmark	Nykobing Mors	56°52'25.9"N	8°58'00.6"E	2019	DNCE	2031 to 2060	no	0	30	no
France	Roscoff	48°43'14.08"N	4°0'17.3"W	2017	FRCE	699 to 938	yes	6	240	no
France	Arcachon	44°39'10.54"N	1°11'50.54"W	2017	FACE	1502 to 1741	no	0	240	no
France	Roscoff	48°44'05.9"N	3°59'00.8"W	2017	FRCE	3011 to 3013	yes	3	144	no
Germany	Sylt	54°48'50"N	8°17'53"E	2017	ASCE	1891 to 2073	no	0	179	no
Ireland	Cork	51°49'16.1"N	8°17'14.2"W	2017	ICCE	3555 to 3567	no	0	13	no
Ireland	Westport	53°47'05.9"N	9°39'09.1"W	2019	IWCE	208 to 357	yes	8	150	no
Ireland	Cork	51°50'43.4"N	8°14'33.6"W	2019	ICCE	358 to 477	yes	9	150	no
Ireland	Inch Beach	52°06'29.1"N	9°57'35.2"W	2019	ITCE	478 to 527	yes	1	63	no
Ireland	Wexford	52°18'25.5"N	6°24'54.7"W	2019	IXCE	528 to 599	yes	12	137	no
Ireland	Dublin	53°21'56.3"N	6°08'18.4"W	2019	IDCE	600 to 649	yes	11	50	no
Morocco	Oualidia	32°44'34.4"N	9°02'31.3"W	2018	MOCE	1061 to 1094; 1250 to 1299; 1767 to 1922	no	0	240	no
Netherlands	Slikken van Viane	51°35'37.9"N	3°57'15.7"E	2017	HSCE	3044 to 3187	no	0	144	no
Norway	Hjeltefjorden	60°24'38.5"N	5°05'31.9"E	2017	NHCE	1762 to 1890	no	0	129	no
Norway	Hjeltefjorden	60°24'38.5"N	5°05'31.9"E	2019	NHCE	3000 to 3014	no	0	15	no
Norway	Bodo	67°N17'	14° 37'E	2019	NBCE	4001 to 4012	no	0	12	no
Norway	Bodo	67°N17'	14° 37'E	2013	NBCE	1 to 5	no	0	5	NA
Portugal	Algarve	36°59'54.5"N	7°58'42.4"W	2017	PACE	381 to 500; 555 to 698; 939 to 986	yes	70	312	no
Portugal	Aveiro	40°37'41.14"N	8°44'32.65"W	2017	PVCE	1247 to 1486	yes	25	240	no
Russia	Murmansk	69° 10'N	36° 05'E	2019	RMCE	5001 to 5020	no	0	20	no
Russia	Dalnye Zelentsy	69°06'38.1"N	36°06'00.0"E	2017	RDCE	8 to 27	no	0	15	NA

Spain	Noia	42°47'30.6"N	8°54'45.6"W	2016	BNg	1 to 20	yes	1	20	no
Spain	Noia	42°47'30.6"N	8°54'45.6"W	2016	ENCE	806 to 831	yes	14	216	no
Spain	Noia	42°47'30.6"N	8°54'45.6"W	2017	ENCE	224 to 343	yes	6	120	no
Spain	Noia	42°47'30.6"N	8°54'45.6"W	2017	ENCE	3568 to 3605; 4500 to 4534	yes	23	240	no
Spain	Noia	42°47'30.6"N	8°54'45.6"W	2019	ENCE	1 to 8	yes	1	8	no
Spain	Noia	42°47'30.6"N	8°54'45.6"W	2019	ENCE	101 to 179	NA	NA	179	no
Spain	País Vasco	43°21'04.1"N	3°01'42.2"W	2017	EBCE	4535 to 4580	no?	0	46	no
Spain	Baiona	42°07'28.0"N	8°51'15.9"W	2018	EYCE	1 to 63	yes	37	240	no
Spain	Carril	42°36'46.2"N	8°46'45.2"W	2018	ECCE	64 to 106	yes	8	240	yes
Spain	Placeres	42°24'35.7"N	8°41'14.4"W	2018	ELCE	207 to 242	no	0	129	yes
Spain	Moaña	42°17'08.9"N	8°43'37.6"W	2018	EMCE	243 to 281; 314 to 324	yes	12	240	no
Spain	Combarro	42°25'41.0"N	8°42'07.7"W	2018	EACE	282 to 313	yes	2	240	yes
Spain	Barallobre	43°28'19.2"N	8°11'59.2"W	2018	EOCE	432 to 484	yes	22	240	no
Spain	Río Anllóns	43°14'07.8"N	8°56'59.2"W	2018	EPCE	485 to 499; 820 to 853	yes	20	240	no
Spain	Camariñas	43°07'43.9"N	9°10'36.3"W	2018	EICE	856 to 859; 865 to 911	yes	23	240	no
Spain	Espasante	43°43'06.9"N	7°48'44.9"W	2018	EECE	854 to 855; 912 to 953	yes	7	240	yes
Spain	Muros	42°46'28.4"N	9°02'58.8"W	2018	EUCE	985 to 1040; 1042	yes	26	240	no
Spain	Grove	42°29'59.2"N	8°52'09.3"W	2018	EGCE	954 to 984; 1041	no	0	240	no
United Kingdom (England)	Plymouth	50°20'5.51"N	4°3'5"W	2017	UDCE	987 to 1226	no	0	240	no
United Kingdom (Scotland)	Loch Gair	56°03'40.1"N	5°19'39.9"W	2020	ULCE	6000 to 6018	no	0	19	NA
United Kingdom (Scotland)	Tràigh Mhòr	57°01'23.4"N	7°26'21.1"W	2020	UTCE	6019 to 6050	no	0	32	NA
United Kingdom (Scotland)	Strollamus	57°16'32.3"N	5°59'31.8"W	2020	USCE	6051 to 6082	no	0	32	NA
United Kingdom (Wales)	Wales	51°22'42.7"N	4°02'21.3"W	2017	UGCE	2098 to 2506	yes	5	240	no
							22	**356**	**6719**	

**Numbers from cytological diagnosis.*

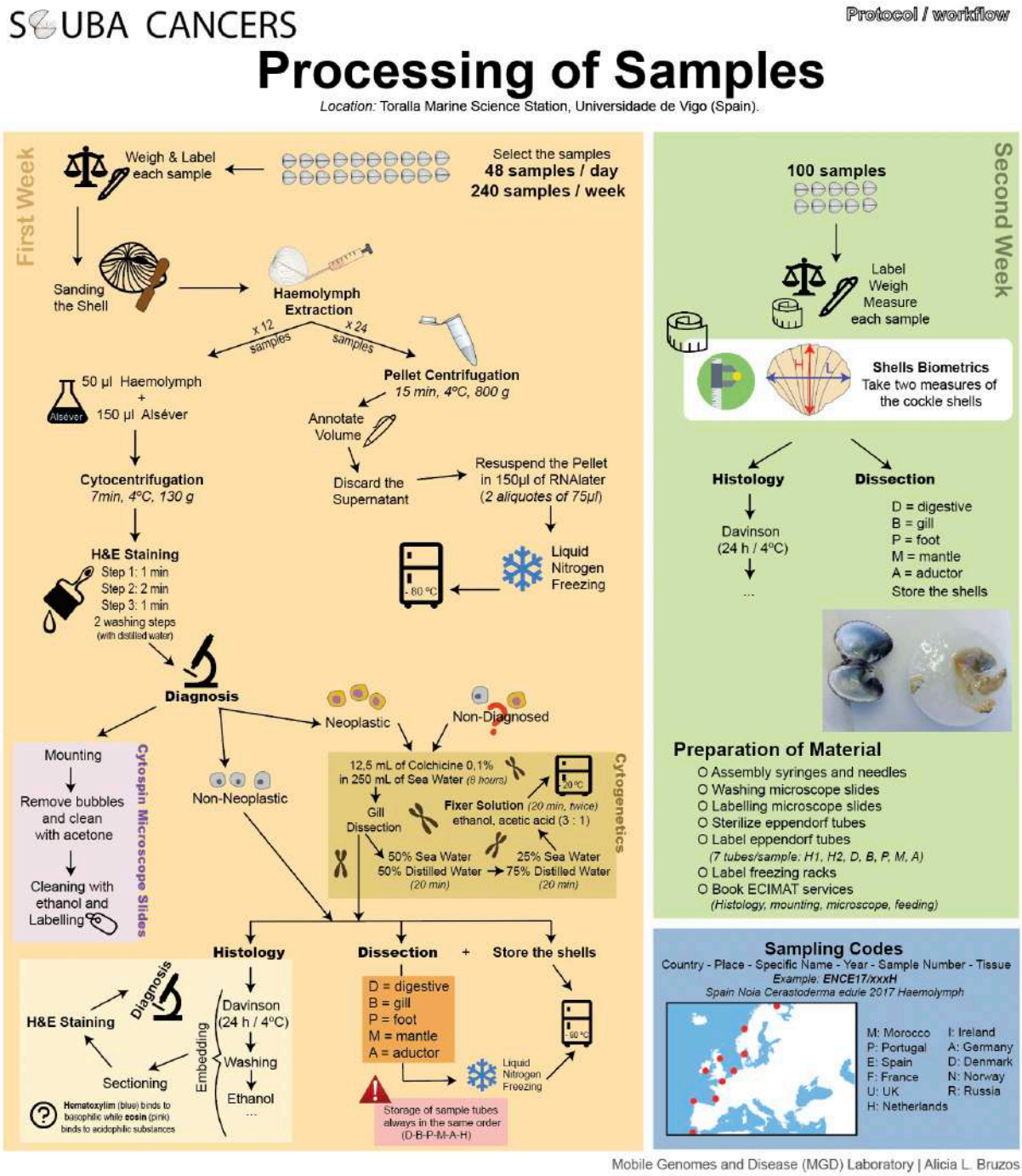


Figure 69. Schematic workflow of sample processing.

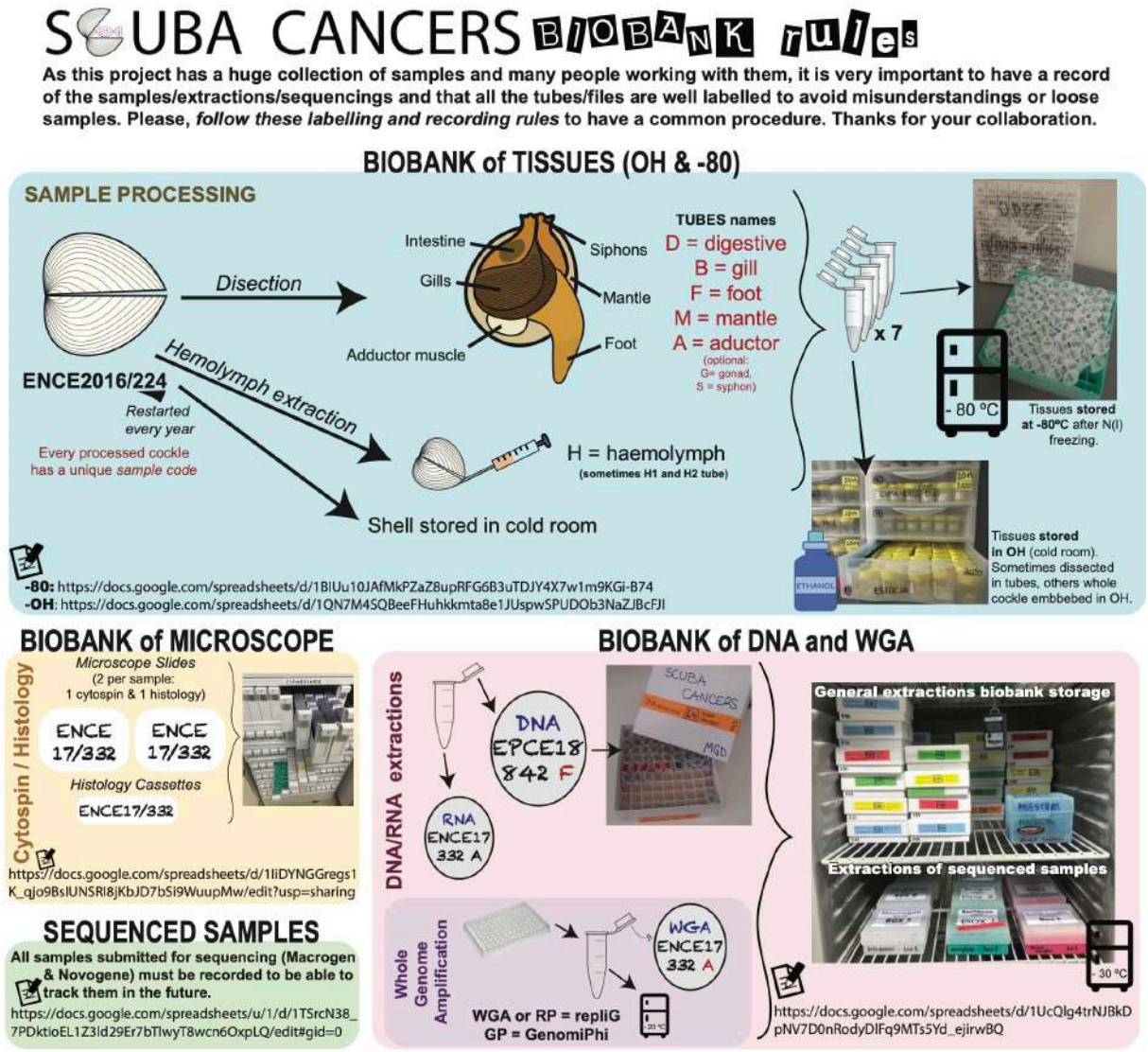


Figure 70. Schematic compendium of rules for the biobank of scuba cancers.

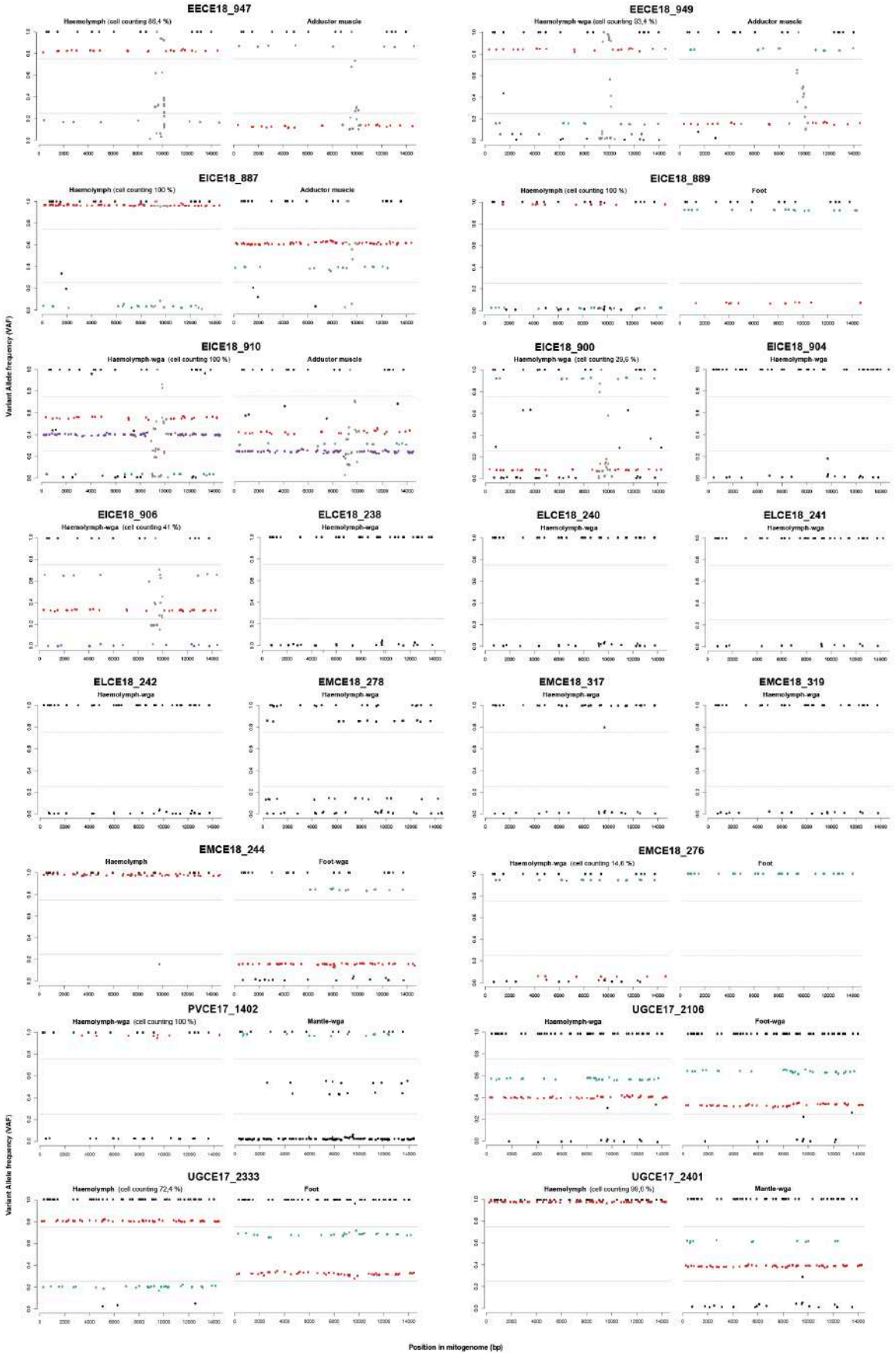


Figure 71. VAF plots of tumour and matched-normal mitogenomes (part 1/4).

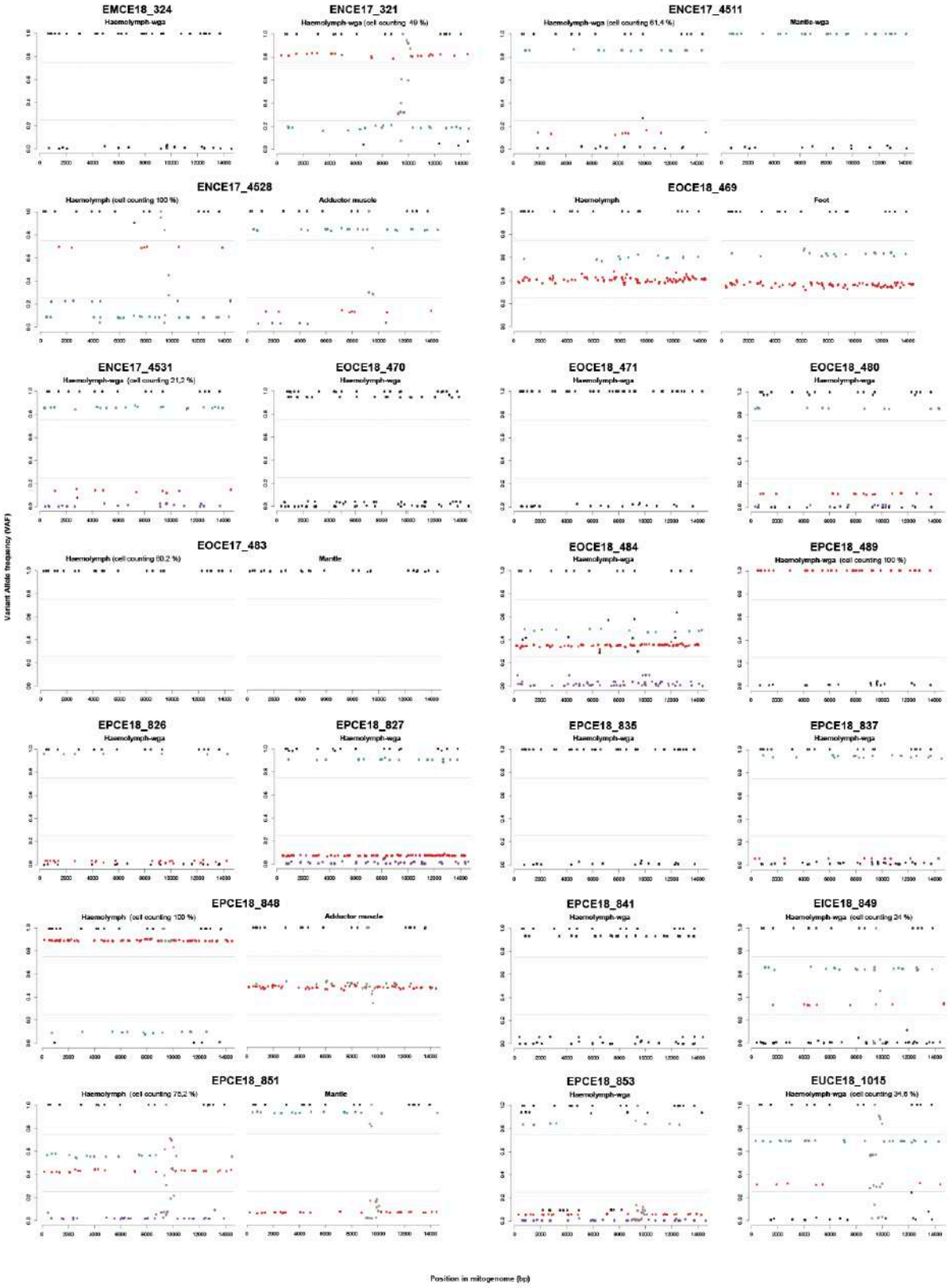


Figure continuation. VAF plots of tumour and matched-normal mitogenomes (part 2/4).

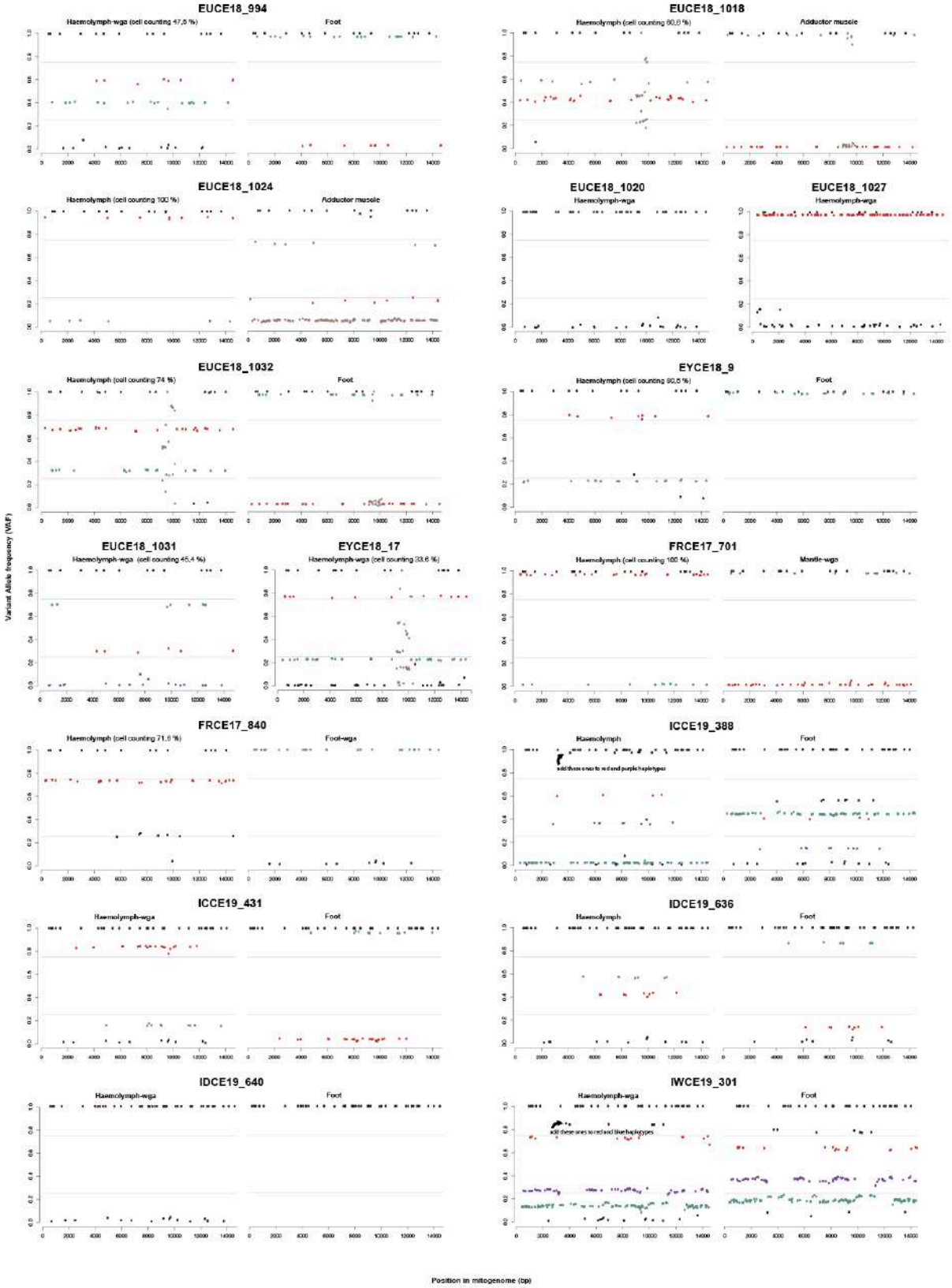


Figure continuation. VAF plots of tumour and matched-normal mitogenomes (part 3/4).

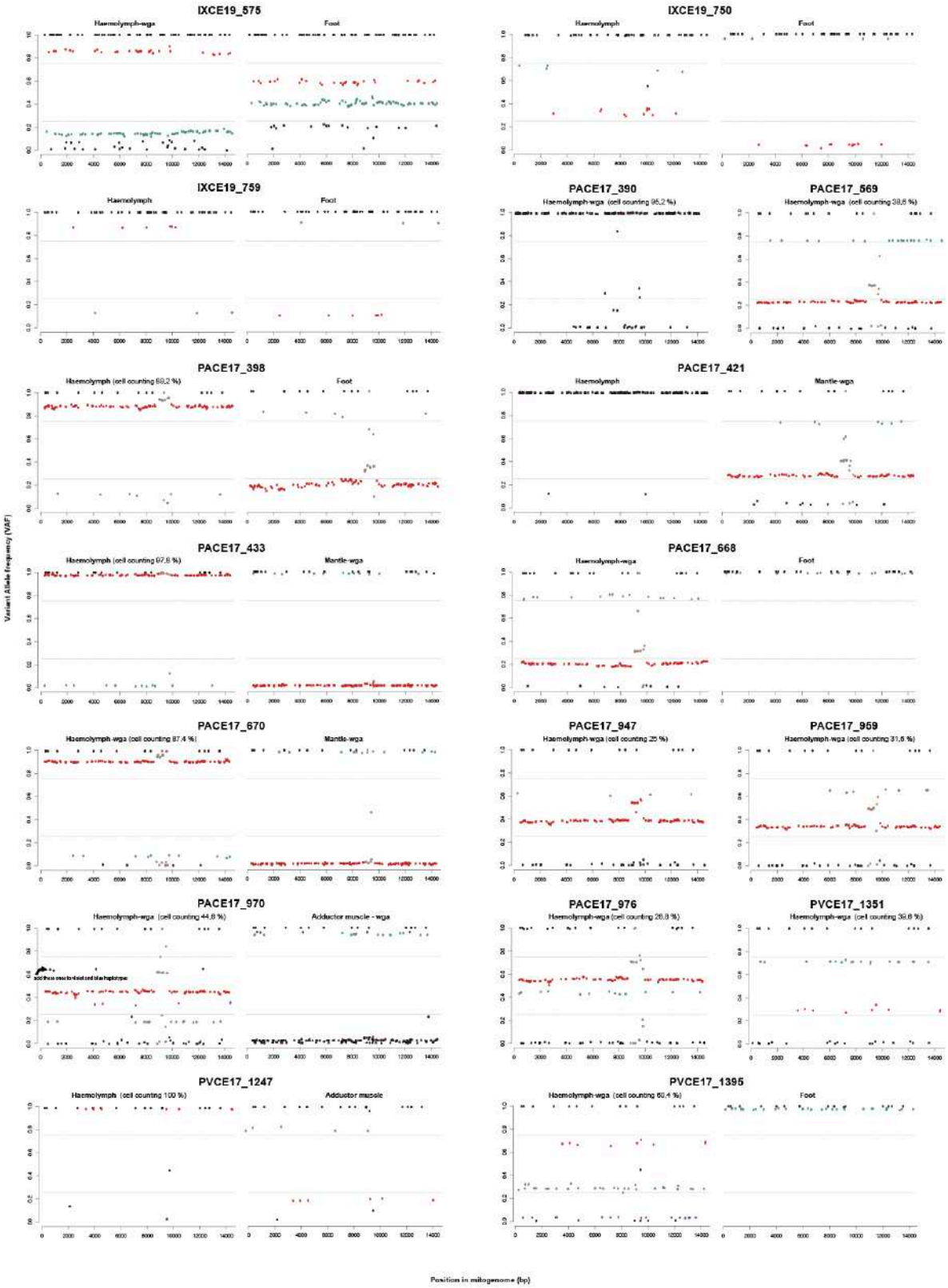


Figure continuation. VAF plots of tumour and matched-normal mitogenomes (part 4/4).

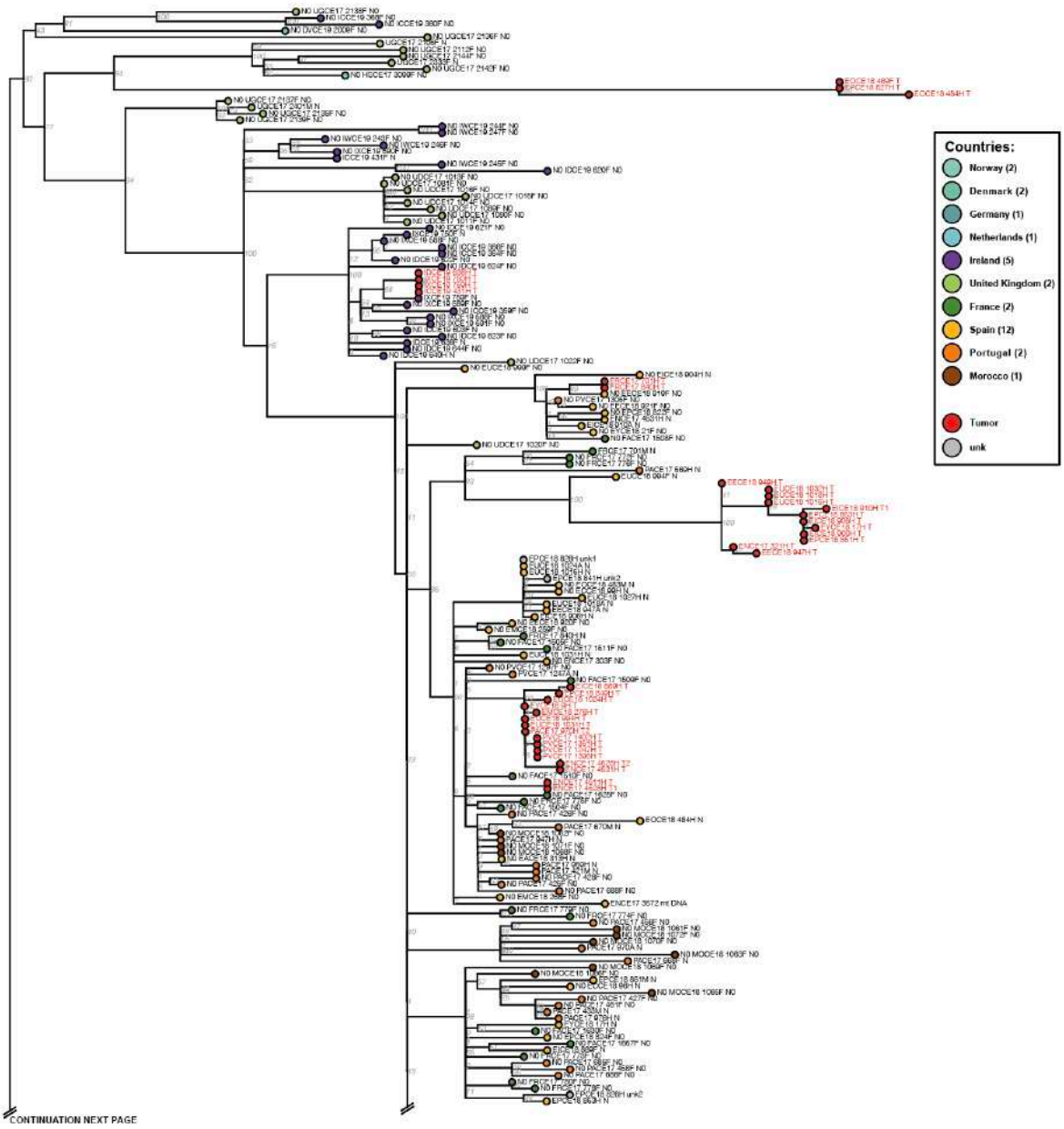


Figure 72. Full mitogenome ML phylogeny of tumours, matched-normal and healthy cockles. Tree is midpoint rooted and 1000 bootstraps are presented in % for relevant nodes.

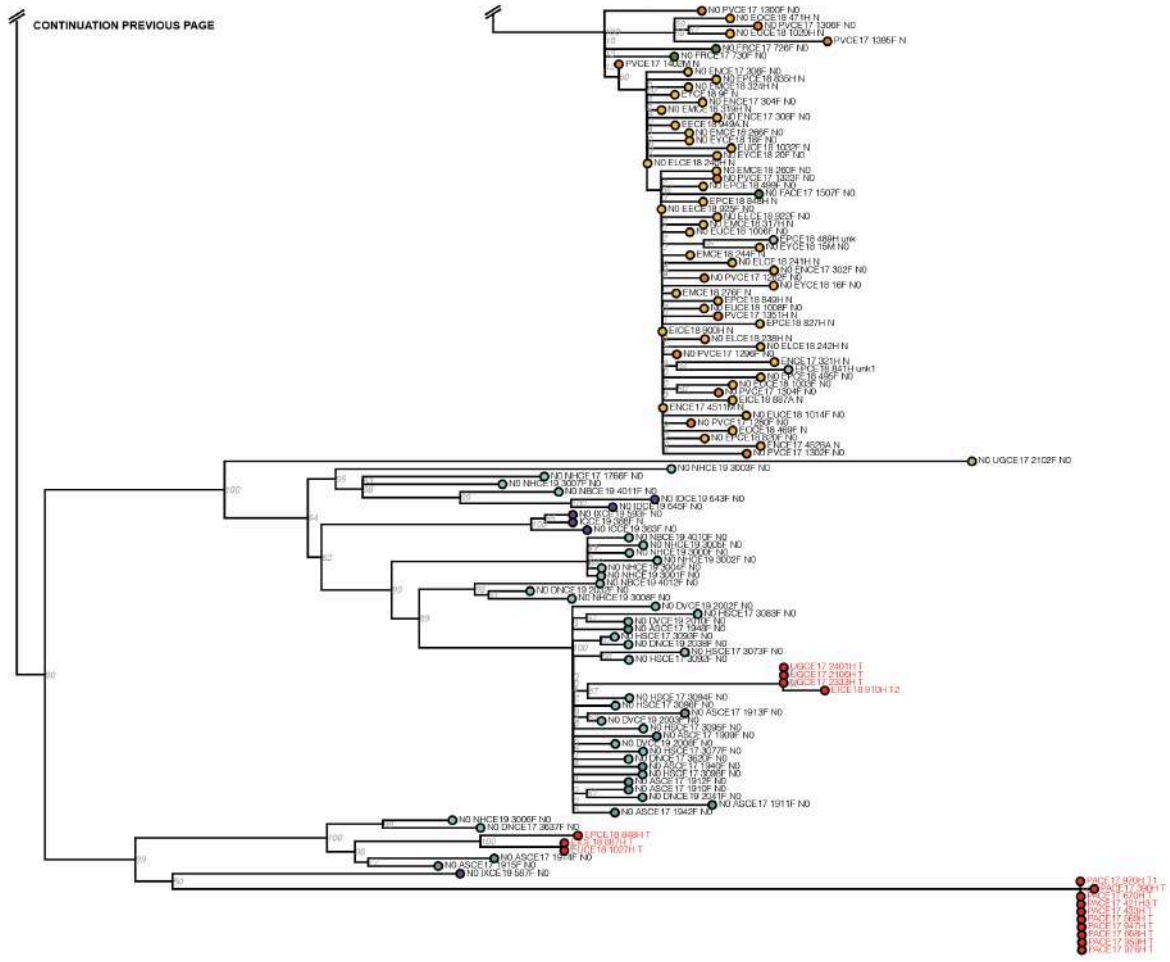


Figure continuation. Full mitogenome ML phylogeny of tumours, matched-normal and healthy cockles. Tree is midpoint rooted and 1000 bootstraps are presented in % for relevant nodes.

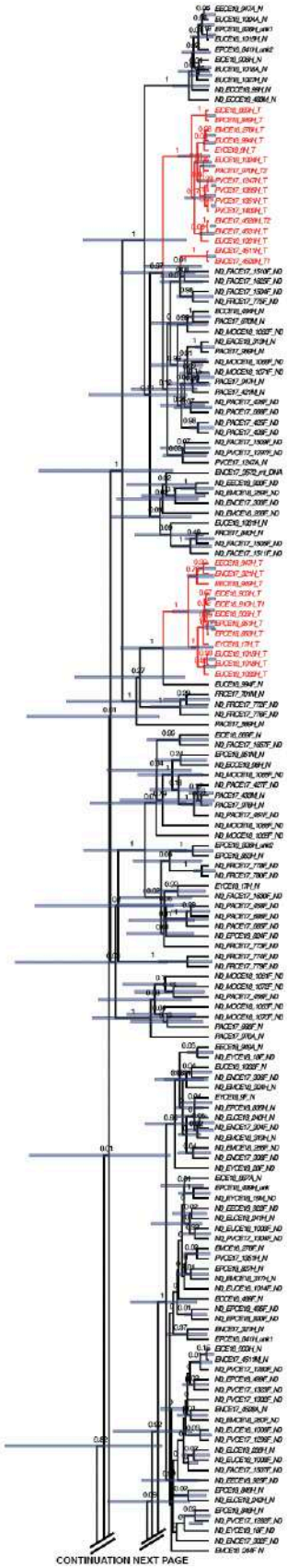


Figure 73. Full mitogenome phylogeny inference of tree and node ages (BEAST) of tumours, matched-normal and healthy cockles; maximum clade credibility (MCC), node bars represent the 95% HDP interval for their age-node high summarized by their common ancestor (20bmedH, 1P, coalescent).

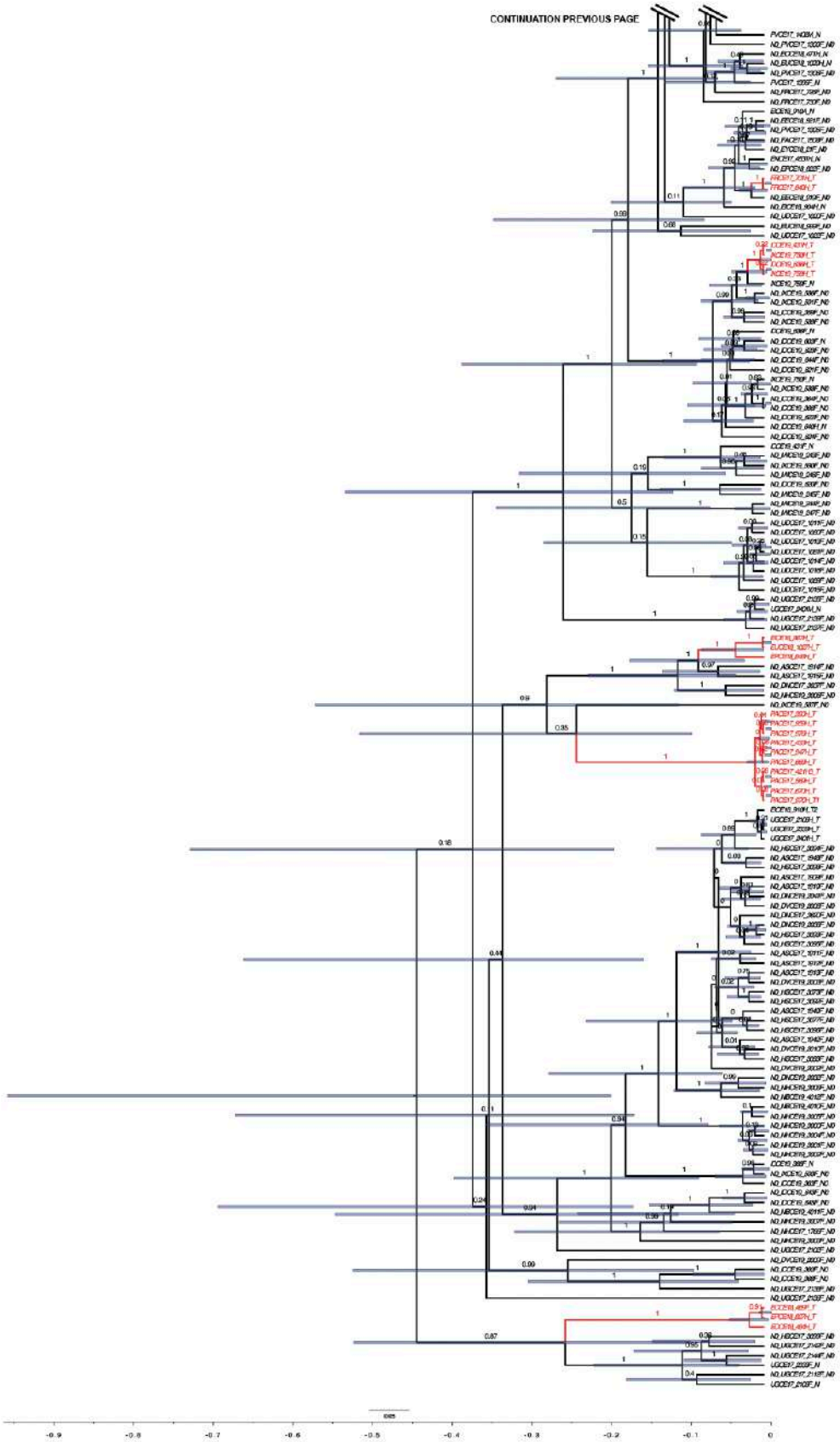


Figure continuation. Full mitogenome phylogeny inference of tree and node ages (BEAST) of tumours, matched-normal and healthy cockles; maximum clade credibility (MCC), node bars represent the 95% HDP interval for their age-node high summarized by their common ancestor (20bmedH, 1P, coalescent).

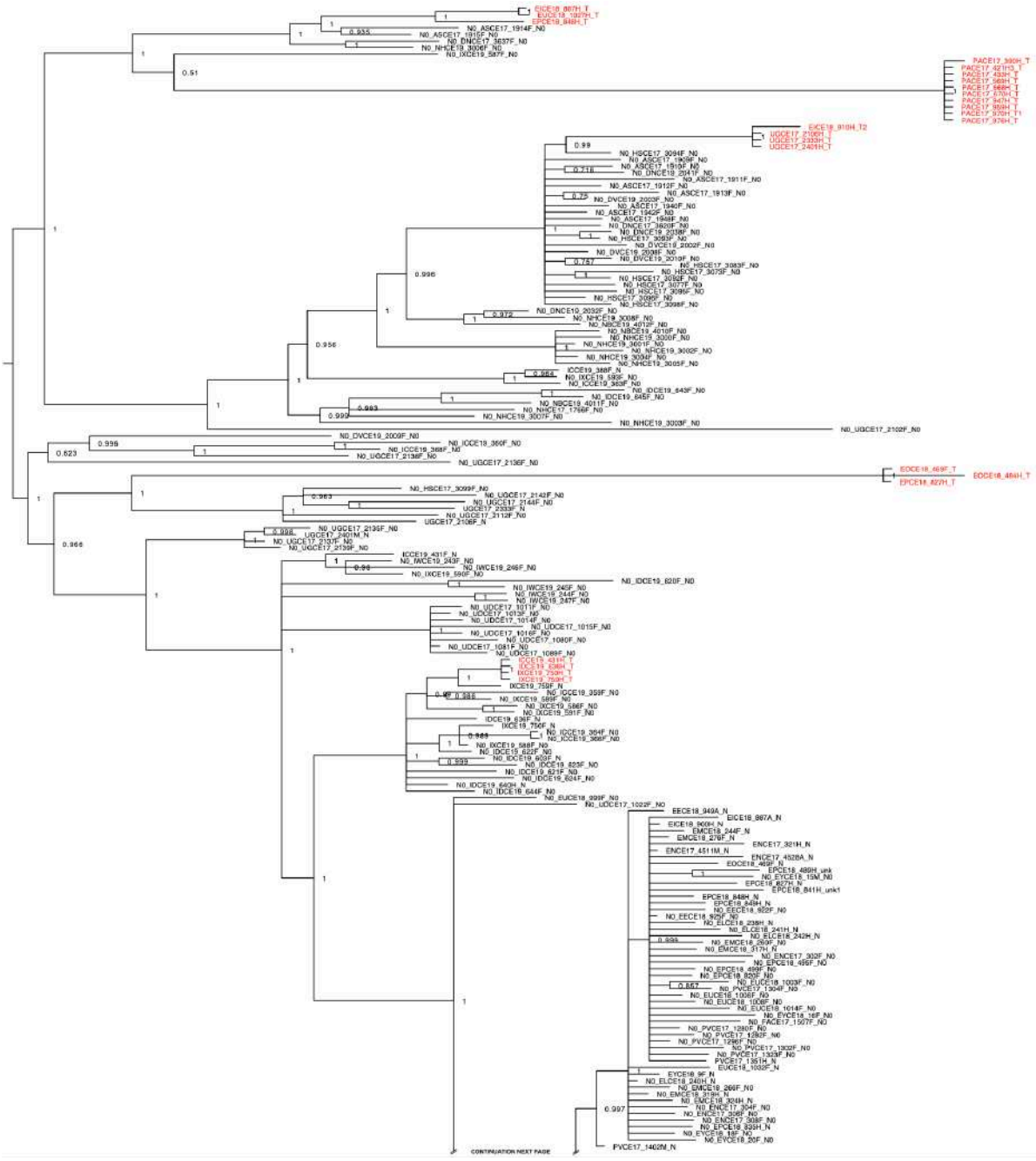


Figure 74. Full mitogenome Bayesian phylogeny (MrBayes inference, 50% majority rule consensus tree of 2 converged and identical runs) of tumours, matched-normal and healthy cockles (posterior probabilities indicated in the nodes, tumour samples highlighted in red).

Figure continuation. Full mitogenome Bayesian phylogeny (MrBayes inference, 50% majority rule consensus tree of 2 converged and identical runs) of tumours, matched-normal and healthy cockles (posterior probabilities indicated in the nodes, tumour samples highlighted in red).

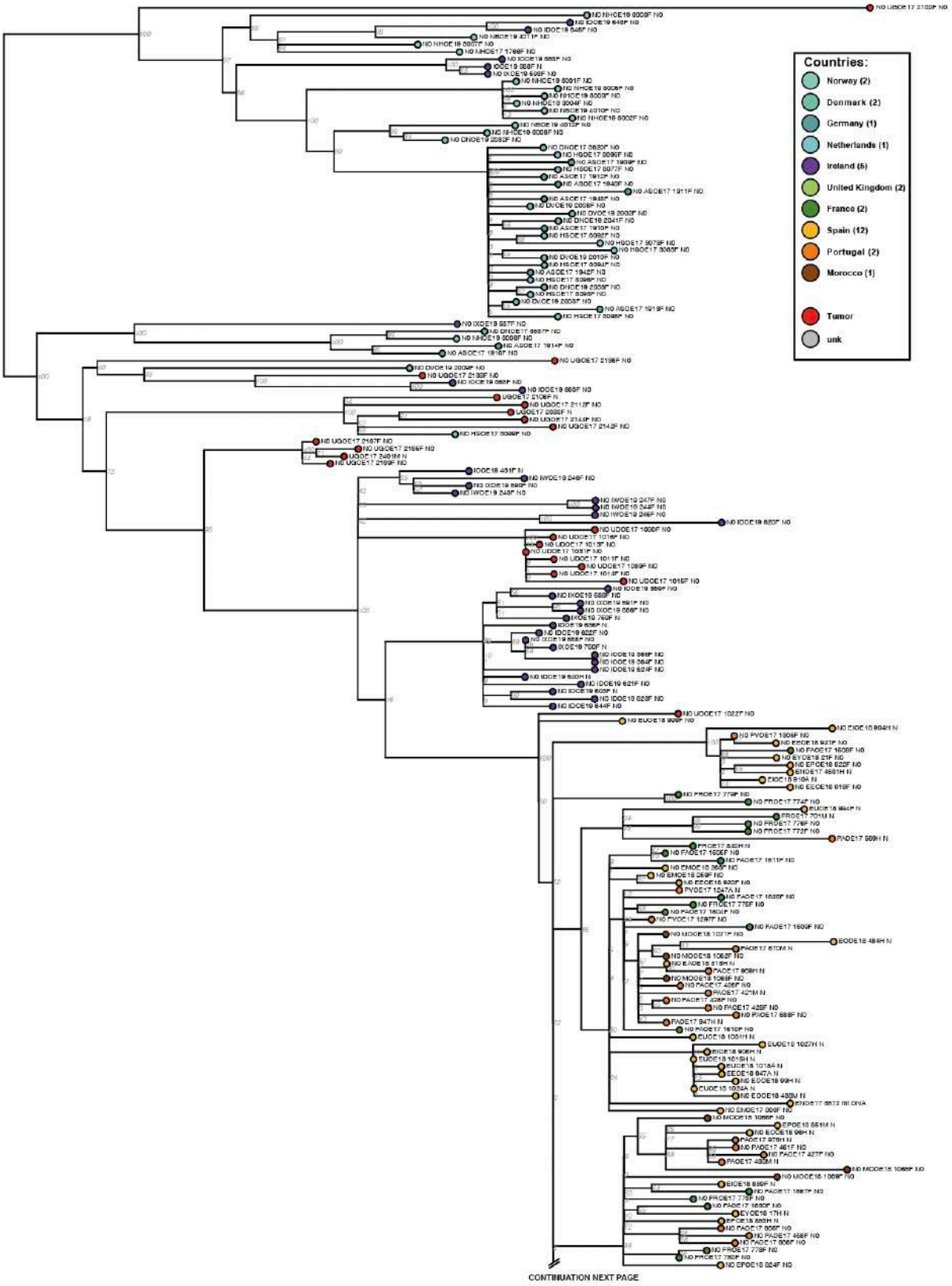


Figure 75. Full mitogenome ML phylogeny of healthy cockles. Tree is midpoint rooted and 1000 bootstraps are presented in % for relevant nodes.

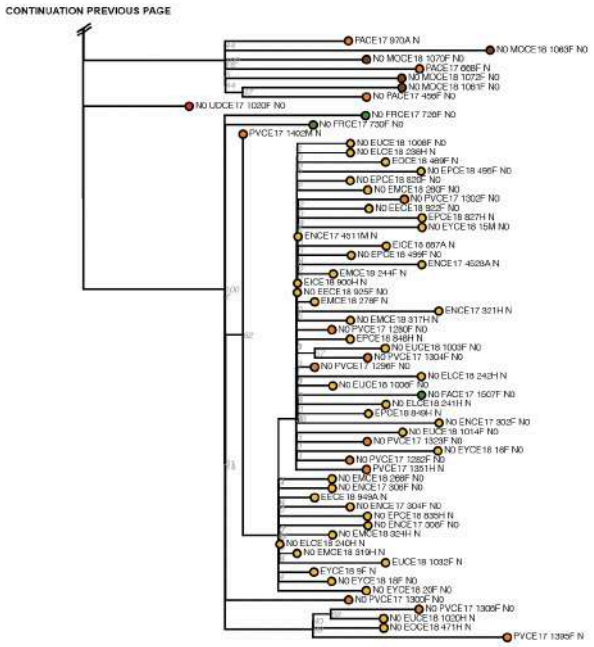


Figure continuation. Full mitogenome ML phylogeny of healthy cockles. Tree is midpoint rooted and 1000 bootstraps are presented in % for relevant nodes.

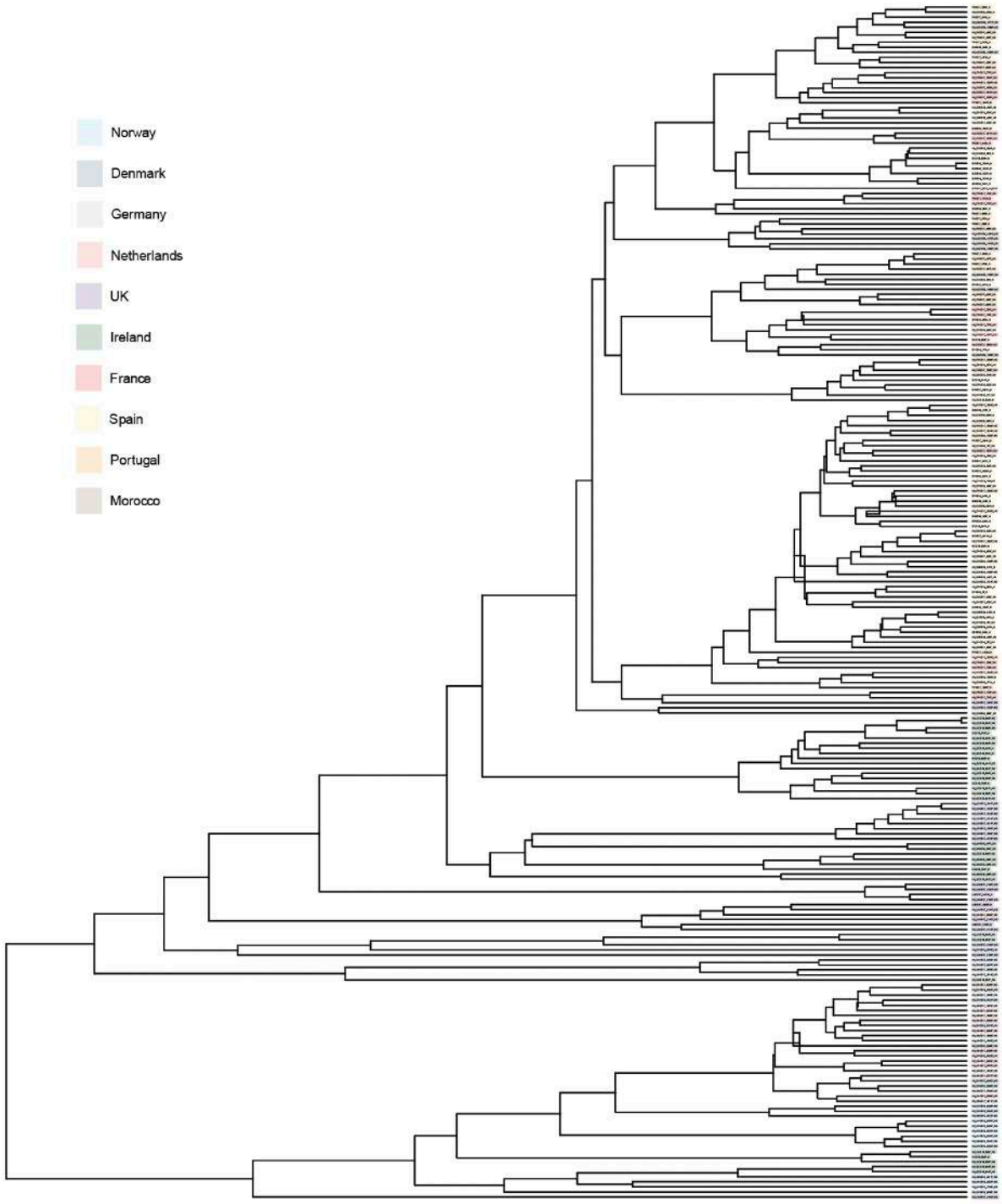


Figure 76. Full mitogenome Bayesian phylogeny (BEAST inference) of healthy cockles.

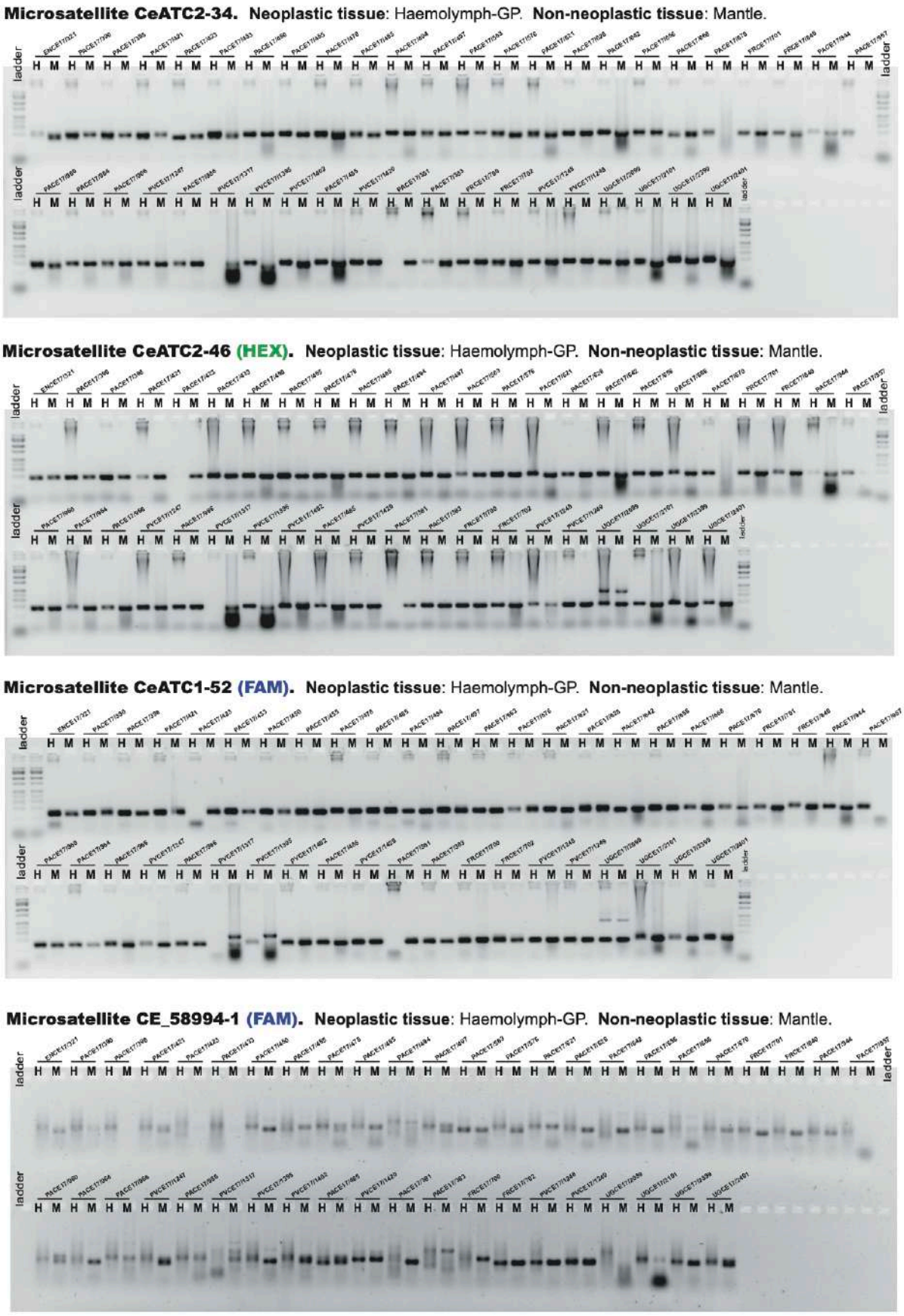
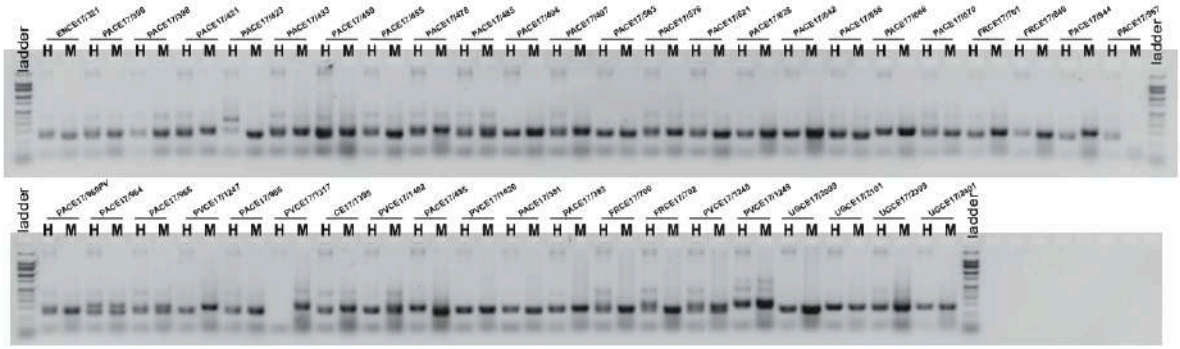


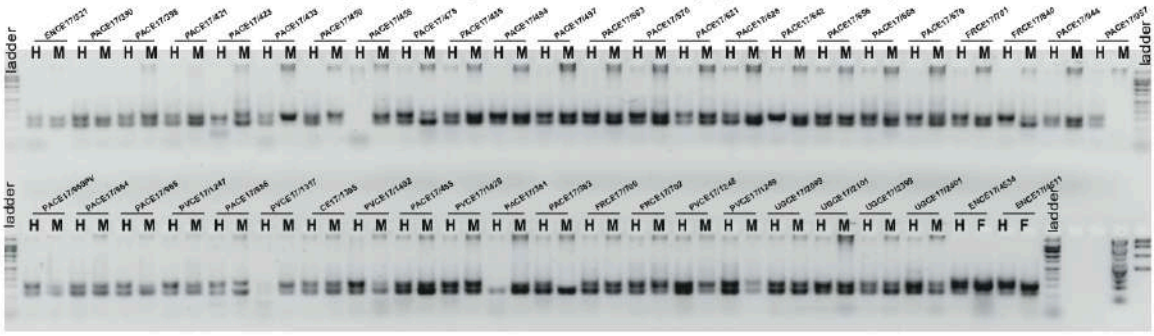
Figure 77. Microsatellite amplifications of tumours and matched-normal tissues (part 1/6).

Figure continuation. Microsatellite amplifications of tumours and matched-normal tissues (part 2/6).

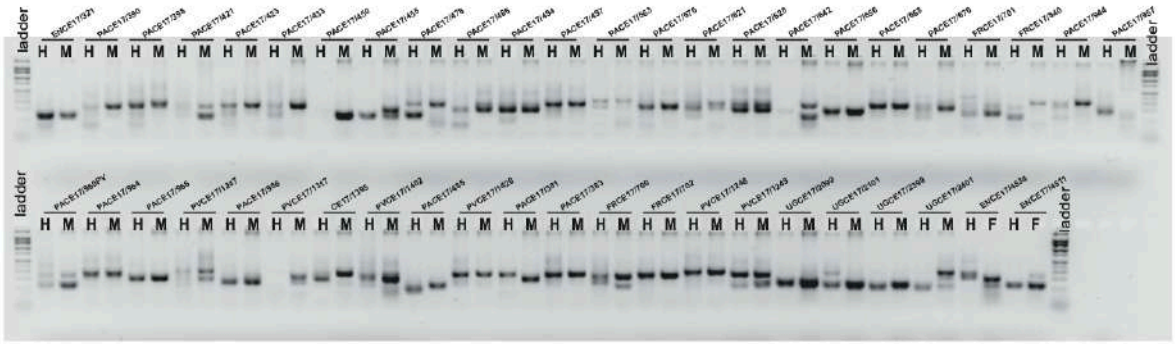
Microsatellite CE_211025_1 (NED). **Neoplastic tissue**: Haemolymph-GP. **Non-neoplastic tissue**: Mantle.



Microsatellite CE_309365_1P (FAM). **Neoplastic tissue**: Haemolymph-GP. **Non-neoplastic tissue**: Mantle-RG/foot.



Microsatellite CE_209688-1 (NED). **Neoplastic tissue**: Haemolymph-GP. **Non-neoplastic tissue**: Mantle-RG/foot.



Microsatellite CE_32645-1 (HEX). **Neoplastic tissue**: Haemolymph-GP. **Non-neoplastic tissue**: Mantle-RG/foot.

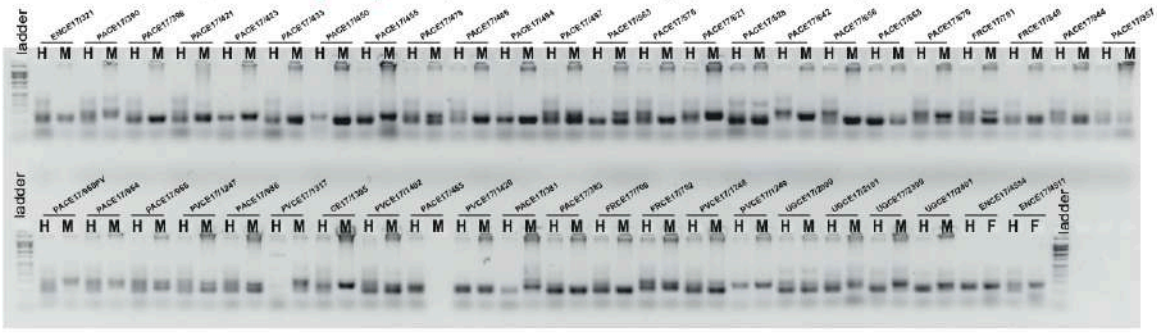


Figure continuation. Microsatellite amplifications of tumours and matched-normal tissues (part 3/6).

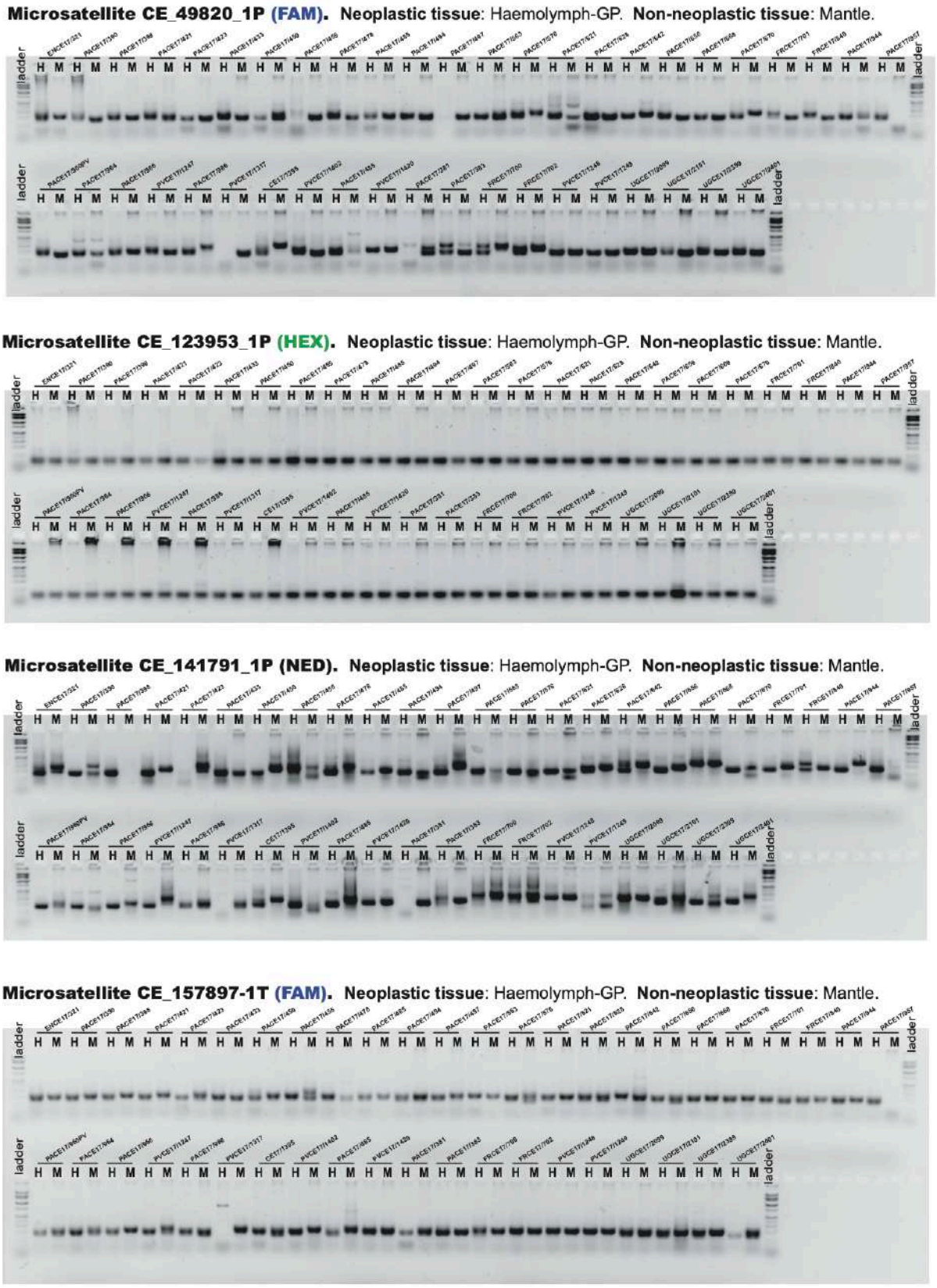
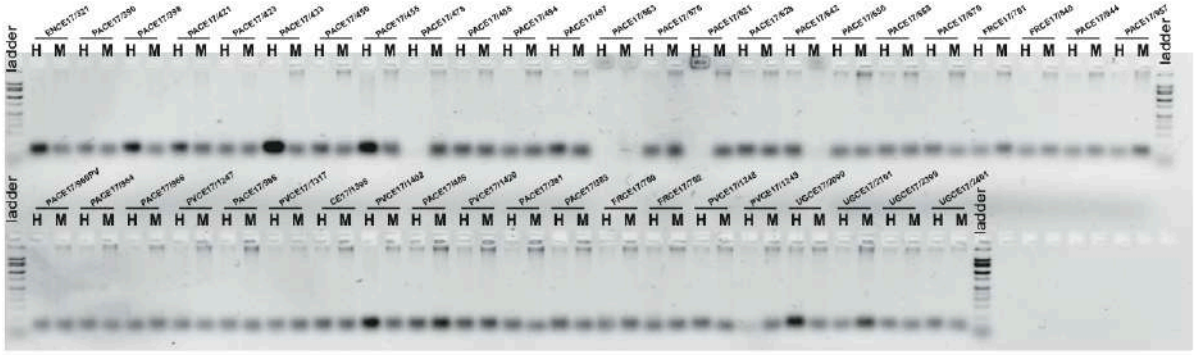
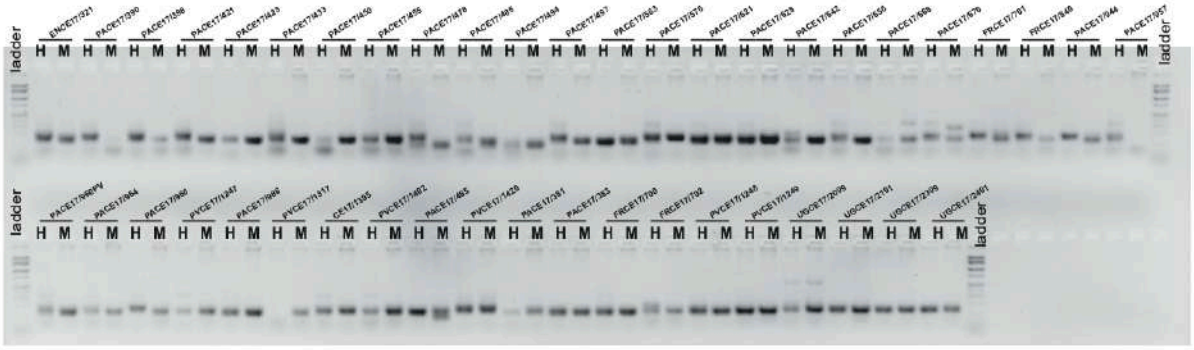


Figure continuation. Microsatellite amplifications of tumours and matched-normal tissues (part 4/6).

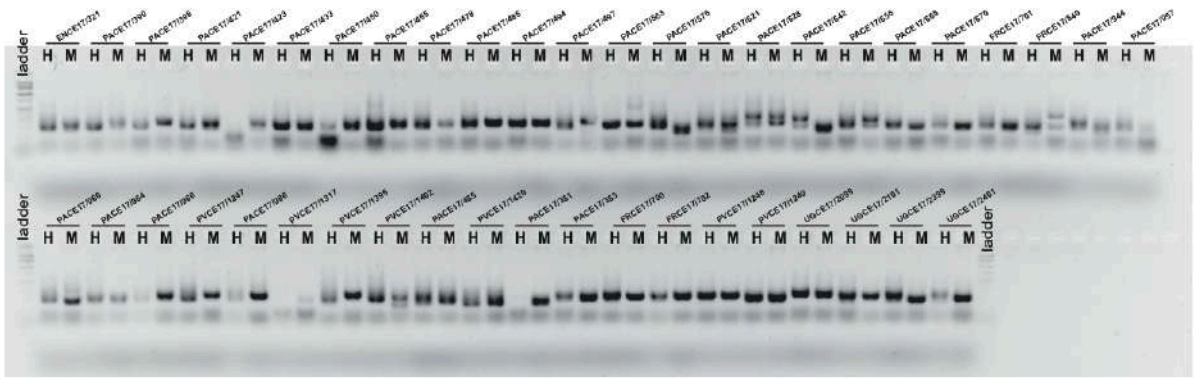
Microsatellite CE_13925_1T (HEX). **Neoplastic tissue**: Haemolymph-GP. **Non-neoplastic tissue**: Mantle.



Microsatellite CeATC2-12 (NED). **Neoplastic tissue**: Haemolymph-GP. **Non-neoplastic tissue**: Mantle.



Microsatellite CeATC2-11 (FAM). **Neoplastic tissue**: Haemolymph-GP. **Non-neoplastic tissue**: Mantle.



Microsatellite CeATC1-54 (NED). **Neoplastic tissue**: Haemolymph-GP. **Non-neoplastic tissue**: Mantle.

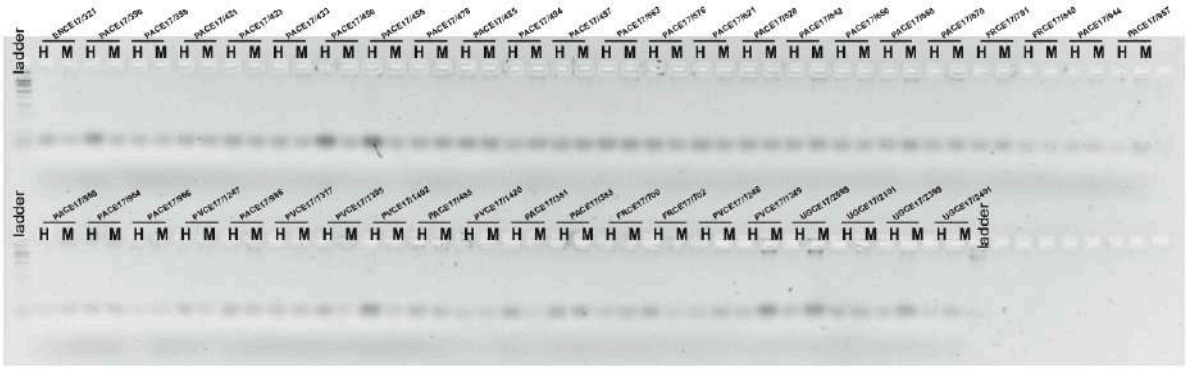


Figure continuation. Microsatellite amplifications of tumours and matched-normal tissues (part 5/6).

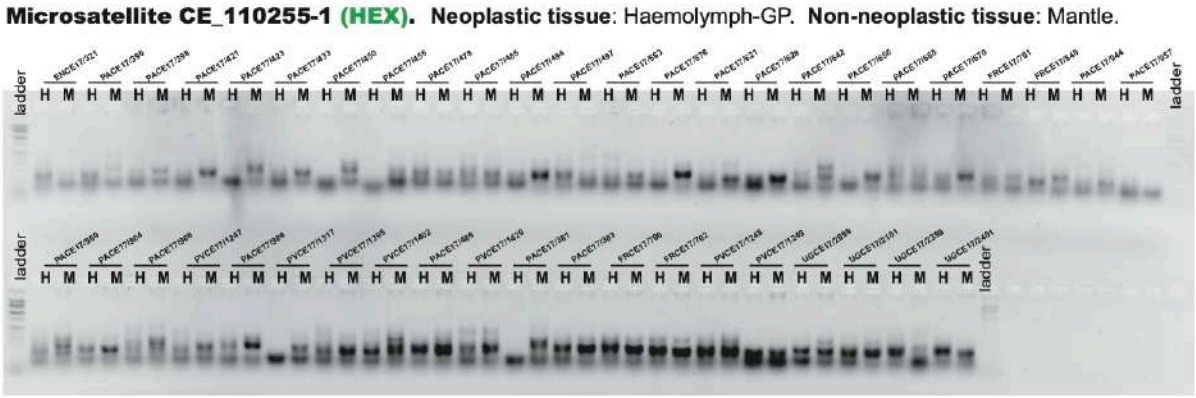


Figure continuation. Microsatellite amplifications of tumours and matched-normal tissues (part 6/6).

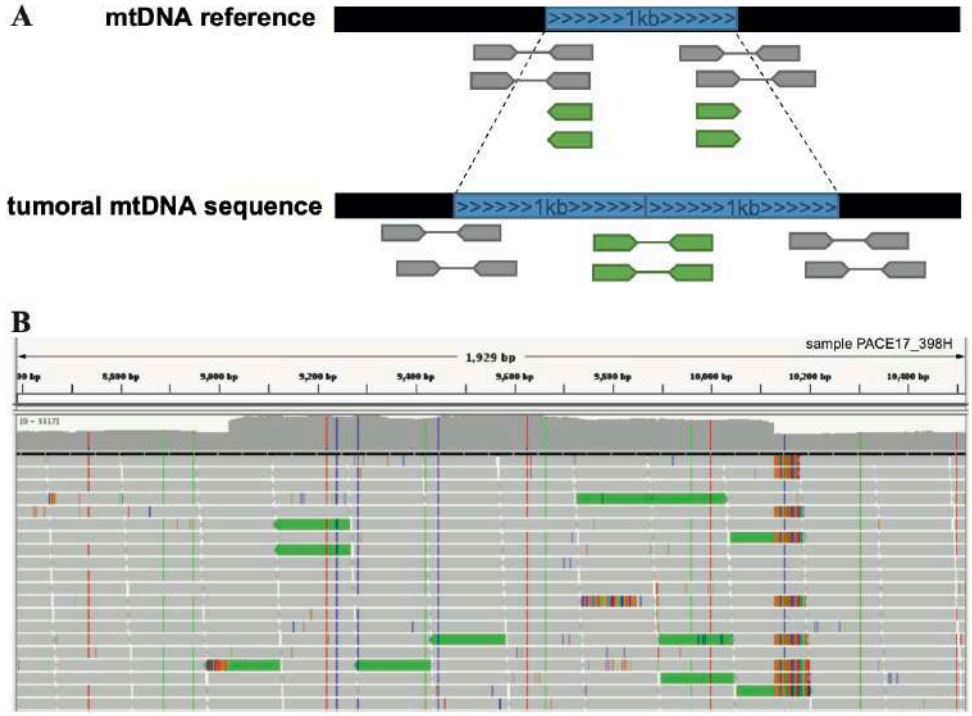


Figure 78. Copy number (CN) amplifications on cockle transmissible cancers (extension). **(A)** Colour pattern diagram of clustered paired end reads suggesting tandem amplifications. **(B)** Alignment of reads in the region where the CN amplifications were identified, coverage increases in the tumoral tissue when aligning reads against the mitogenome, green reads are displayed, and clipped reads are shown in the breakpoints. **(C)** Alignment of ONT long reads against two mitochondrial references (with and w/o tandem duplication). **(D)** Primer design strategy to detect these amplifications. **(E)** Primer design strategy of flanking the amplification. **(F)** Electrophoresis gel results of the flanking strategy design.

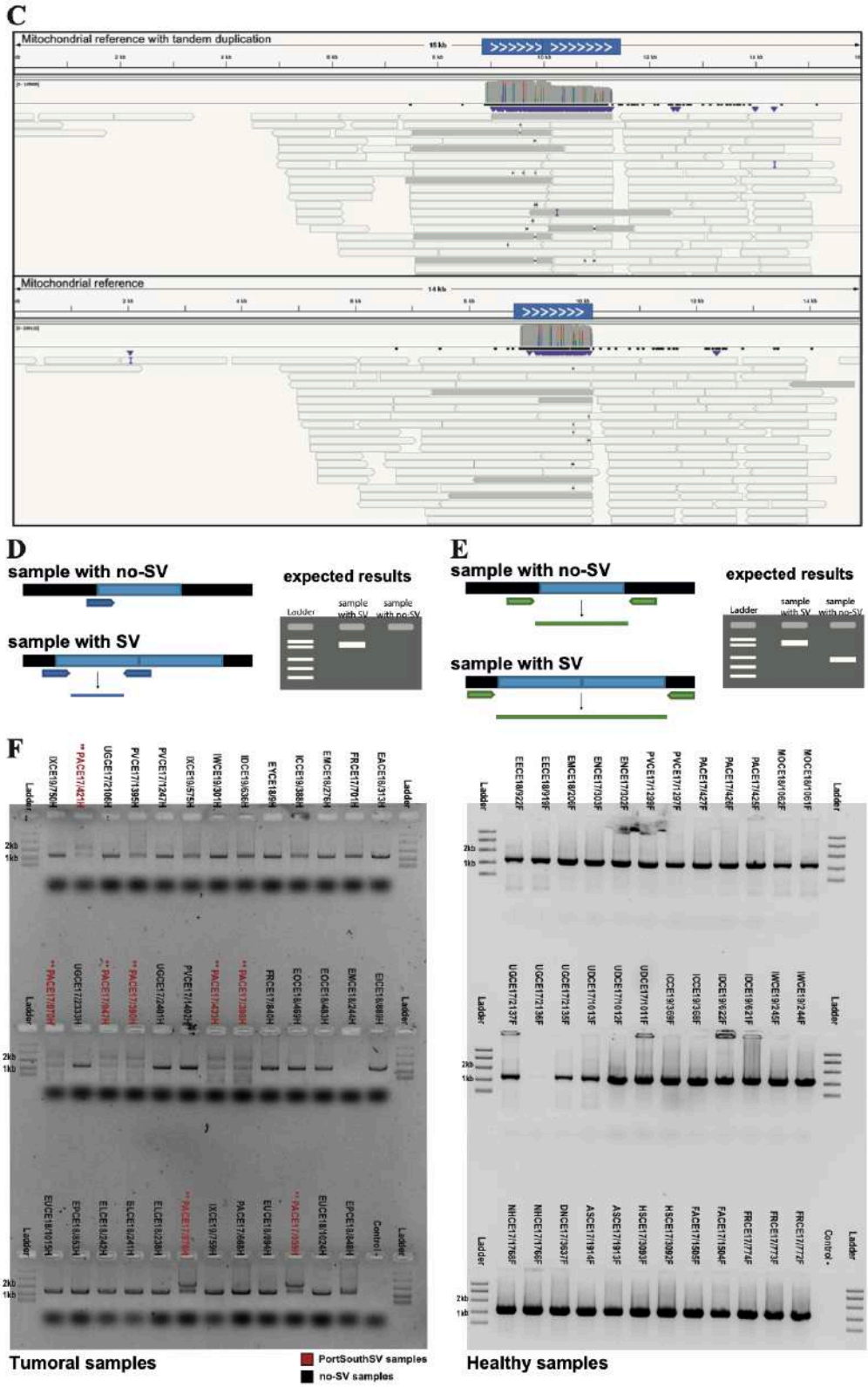


Figure continuation. Copy number (CN) amplifications on cockle transmissible cancers (extension). **(A)** Colour pattern diagram of clustered paired end reads suggesting tandem amplifications. **(B)** Alignment of reads in the region where the CN amplifications were identified, coverage increases in the tumoral tissue when aligning reads against the mitogenome, green reads are displayed, and clipped reads are shown in the breakpoints. **(C)** Alignment of ONT long reads against two mitochondrial references (with and w/o tandem duplication). **(D)** Primer design strategy to detect these amplifications. **(E)** Primer design strategy of flanking the amplification. **(F)** Electrophoresis gel results of the flanking strategy design.

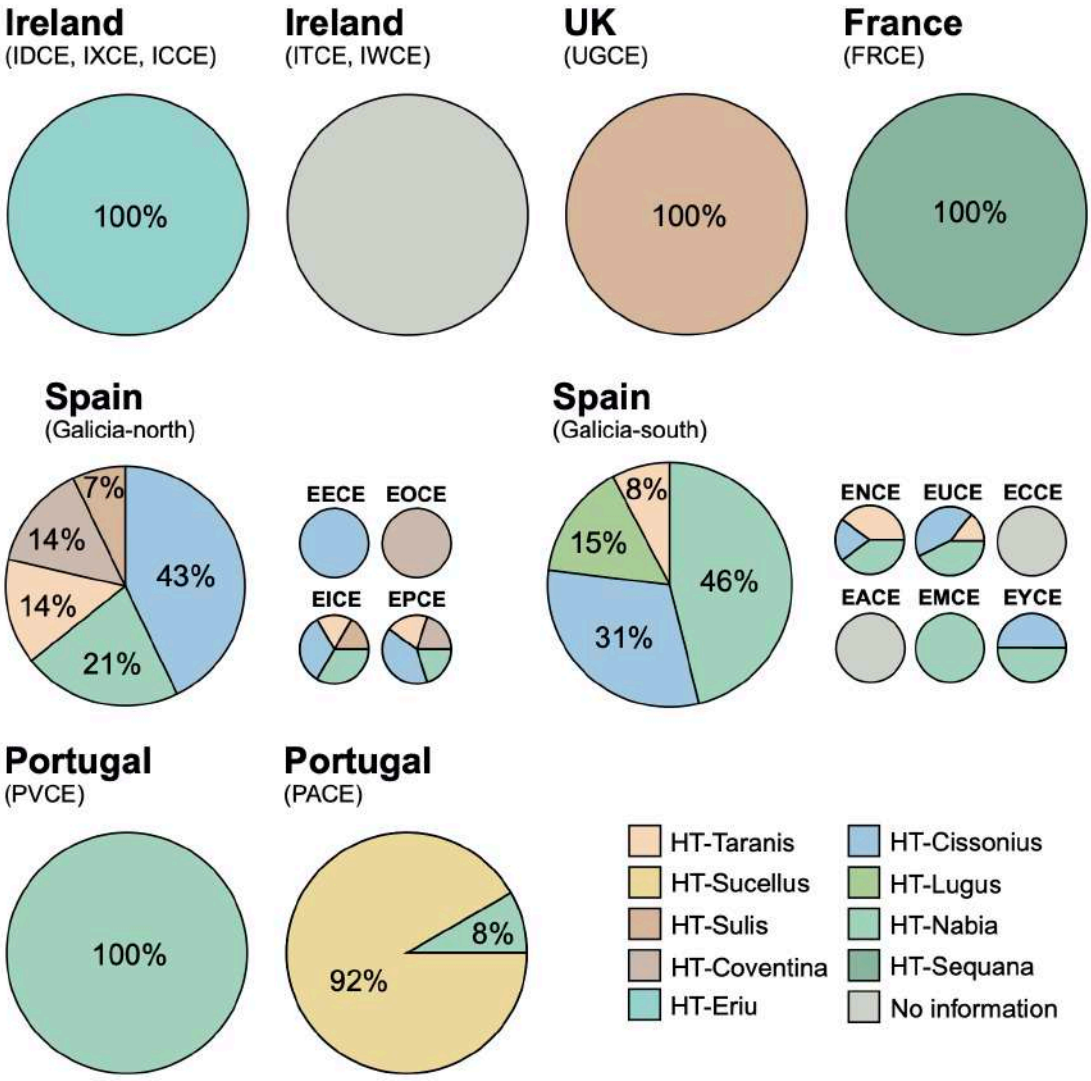


Figure 79. Piecharts of mitochondrial cancer lineages by sampling locations or areas.

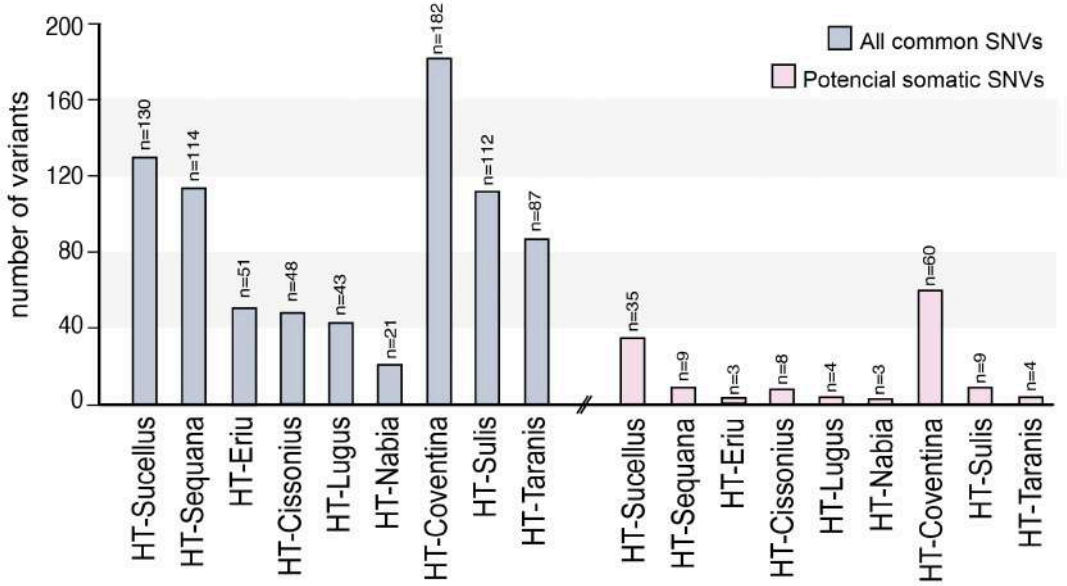


Figure 80. Barplot of common variants per mitochondrial lineage.

Nomenclature of mitochondrial Horizontal Transfers (HT)

Cockle transmissible cancers have been mainly found in Atlantic southern European countries most of them associated with a modern Celtic identity. In this thesis, nine cases of mtDNA horizontal transfer (HT) are described and named after nine Celtic deities and gods/goddess.

HT-Sequana
SEQUANA: goddess of the river Seine

French lineage

Sequana in the Musée Archéologique in Dijon, France (Source: Wikimedia commons, public domain).

HT-Sulis
SULIS: deity worshiped at the thermal spring of Bath, now in Somerset.

English lineage

Head found in 1727 and displayed at the Roman Baths, Bath (Source: Wikimedia commons, public domain).

HT-Eriu
ÉRIU: sovereign goddess of Ireland

Irish lineage

The Harp of Erin, painted by Thomas Buchanan (Source: Wikimedia commons, public domain).

HT-Lugus
LUGUS: the master of the twenty crafts or the inventor of all the arts.

Gal-smallSV lineage

Engraving of a tricephalic god, often identified as Lugus, whose bas-relief was discovered in Paris in 1867 (Source: Wikimedia commons, public domain).

HT-Coventina
COVENTINA: goddess of the waters

GalNorth lineage

Inscribed bas-relief of Coventina (Source: Wikimedia commons, public domain).

HT-Cissonius
CISSONIUS: god of trade and protector of travellers

Gal-longSV lineage

Relief of Mercury Cissonius from the Palatinate (Source: Wikimedia commons, public domain).

HT-Taranis
TARANIS: god of thunder

GalMid lineage

Taranis, France (Source: Wikimedia commons, public domain).

HT-Nabia
NABIA: goddess of rivers and water in Gallaecian and Lusitanian mythology.

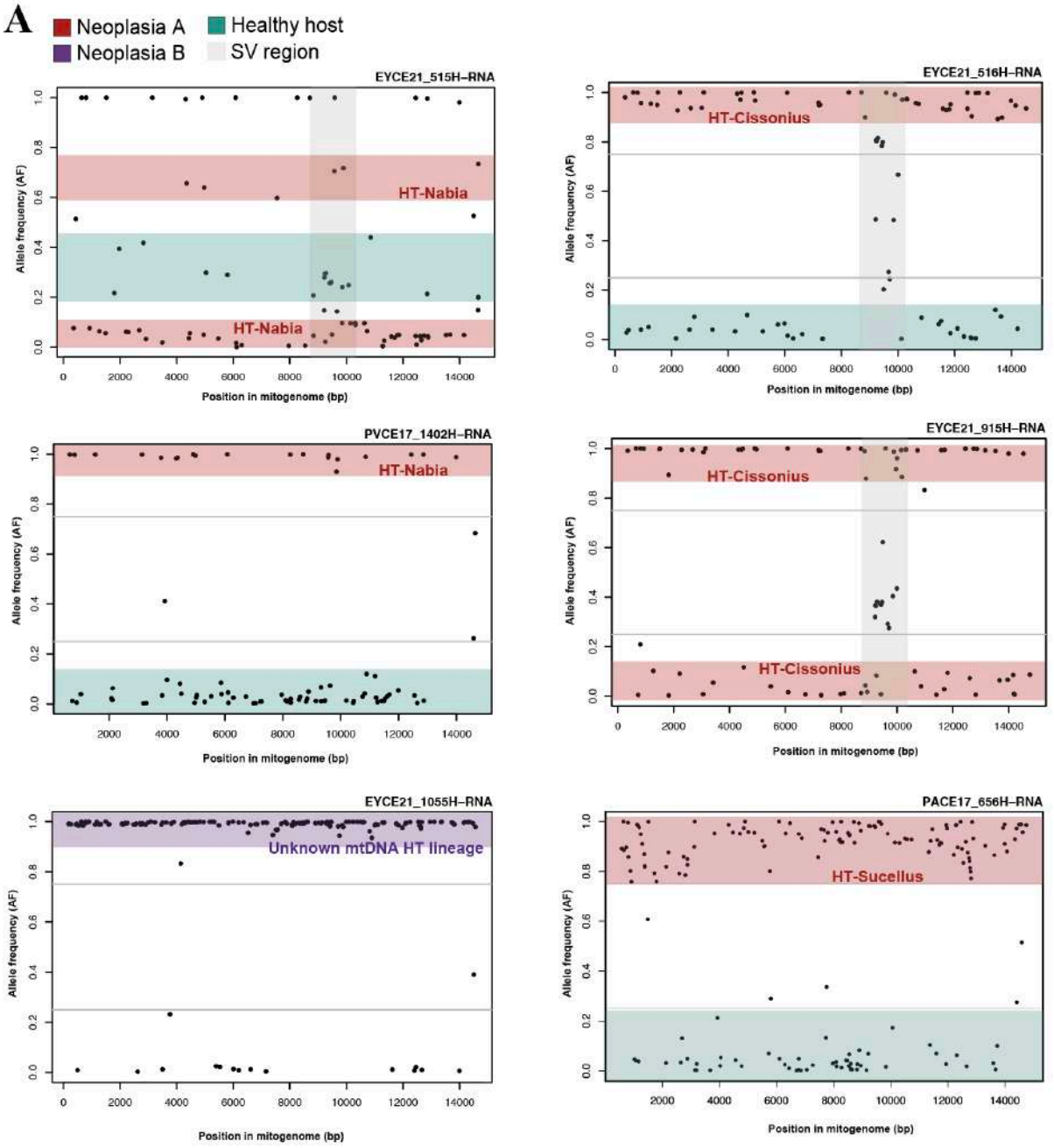
GalPor lineage

Sanctuary dedicated to the god Tongonabiago, associated with the cult of waters through the goddess Nabia. Fonte do Ídolo in Braga, Portugal (Source: Wikimedia commons, public domain).

HT-Sucellus
SUCELLUS: deity of traditional medicine, agriculture and forests. He belongs to the mythological pantheon of Lusitania.

PortSouthSV lineage

Sucellus with his characteristic hammer and *olla* at the Musee National d'Archeology (Source: Wikimedia commons, public domain).


A
Neoplasia A
Neoplasia B
Healthy host
SV region
EYCE21_515H–RNA
HT-Nabia
HT-Nabia
EYCE21_516H–RNA
HT-Cissonius
PVCE17_1402H–RNA
HT-Nabia
EYCE21_915H–RNA
HT-Cissonius
HT-Cissonius
EYCE21_1055H–RNA
Unknown mtDNA HT lineage
PACE17_656H–RNA
HT-Sucellus
Allele frequency (AF)
Position in mitogenome (bp)

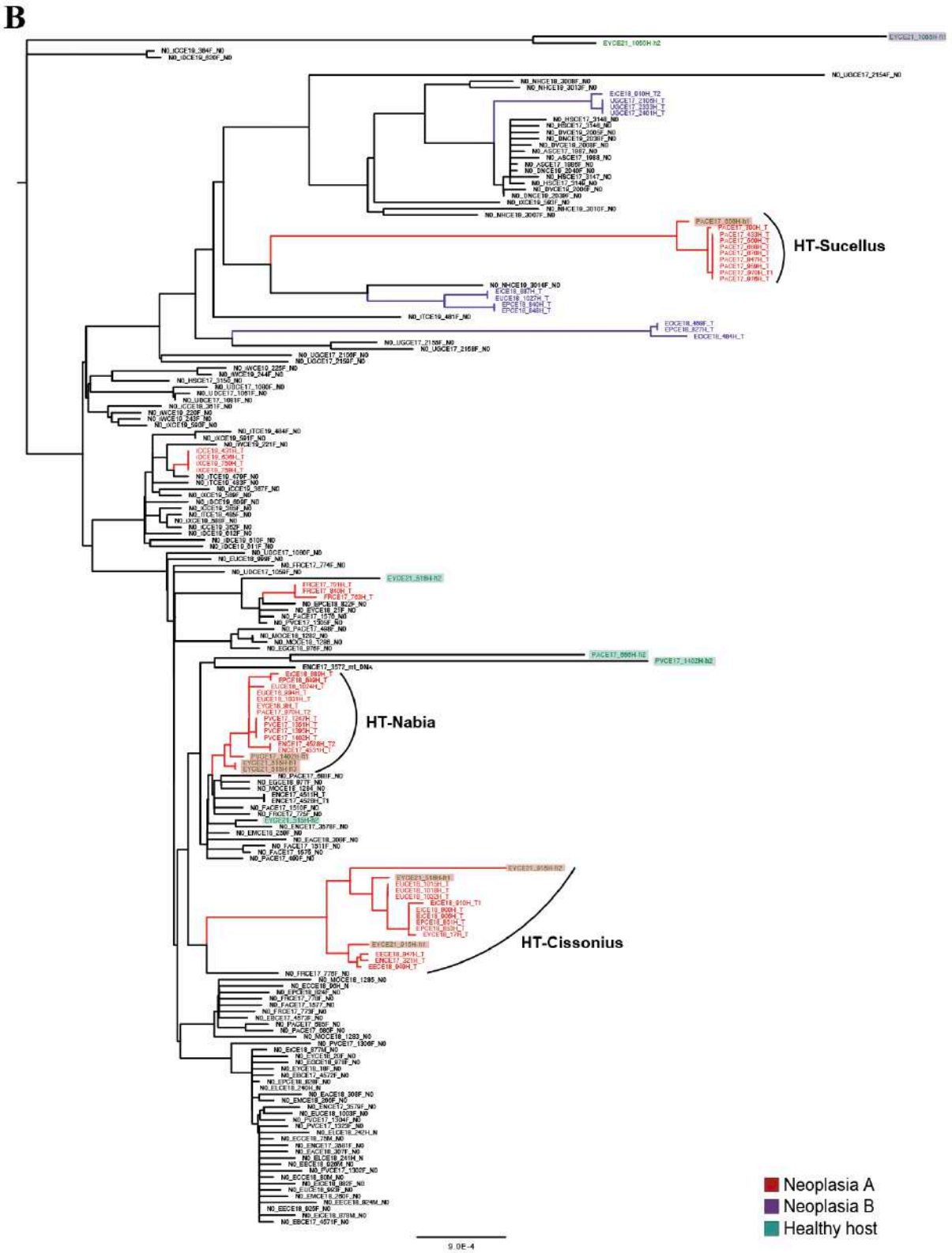


Figure 81. RNA HN samples analysis. **(A)** VAF plot of the several samples used for the RNAseq analysis, clonal deconvolution shaded in colours after checking the phylogeny. **(B)** Neighbour joining tree of the HN samples with RNA sequenced to unravel their lineage and sub lineage. Two samples are missing due to technical difficulties.

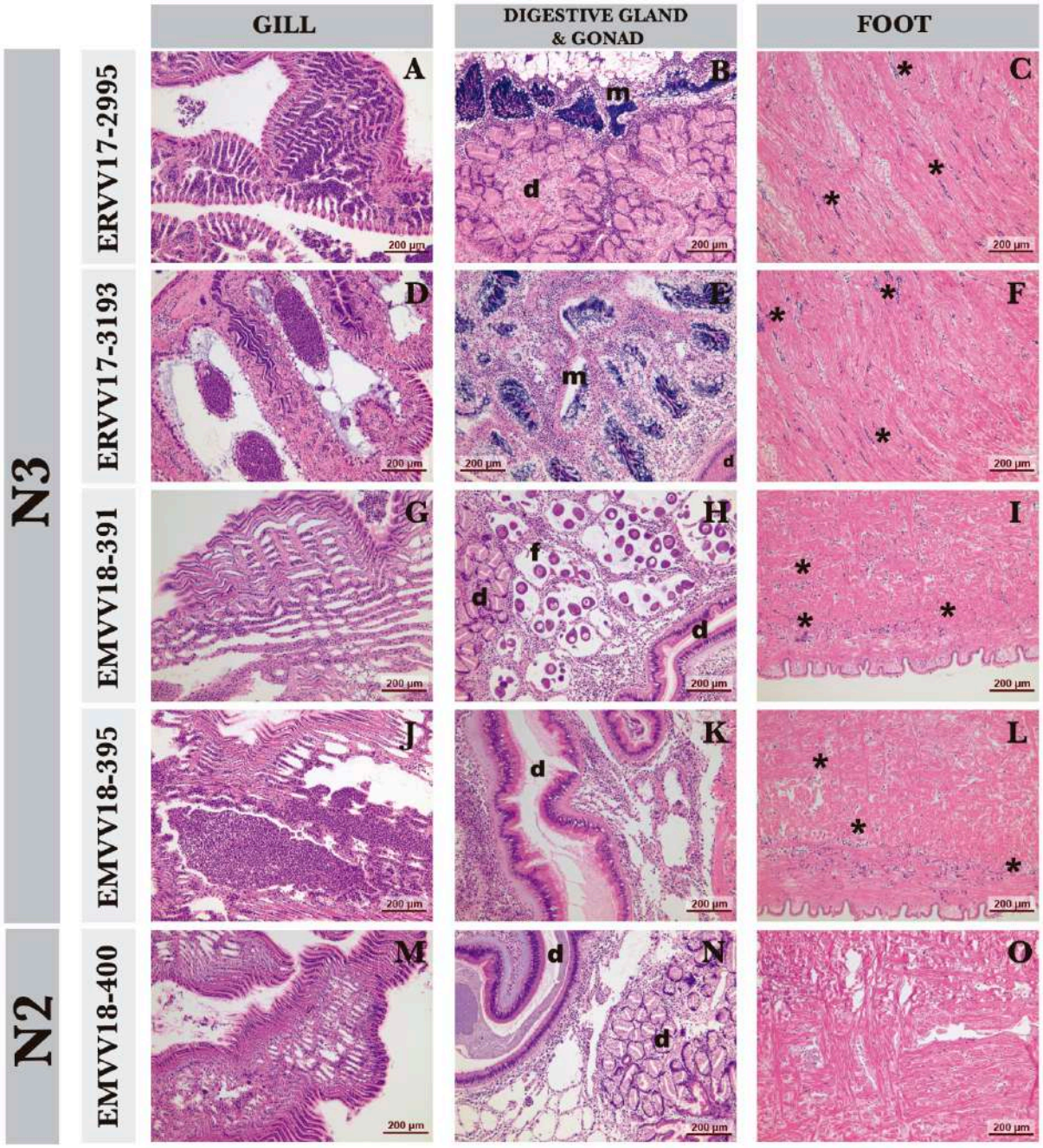


Figure 82. Histological diagnosis of hemic neoplasia in warty venus (*V. verrucosa*) specimens. Hematoxylin and eosin-stained photomicrographs of gill, digestive (d), gonad (male (m) & female (f)) and foot of warty venus specimens diagnosed with different stages of hemic neoplasia: high (N3), medium (N2), light (N1) and healthy (N0). In the N3 stage, neoplastic cells infiltrate the connective tissue and vessels of different organs (A-L), and show low infiltration of foot (C,F,I,L). In N2 stage, cell groups are observed in different organs such as gills (M) and are not detected in the foot (O). In N1 stage, groups of neoplastic or isolated cells are detected in gill sinuses (P, S, V). N0 stage is completely devoid of any trace of hemic neoplasia at either gill, digestive gland and gonad and foot (Y-AB). Arrows show isolated cells. Asterisks show groups of neoplastic cells. (Source: García-Souto et al. 2022, Copyright 2022, eLife, CC-BY 4.0).

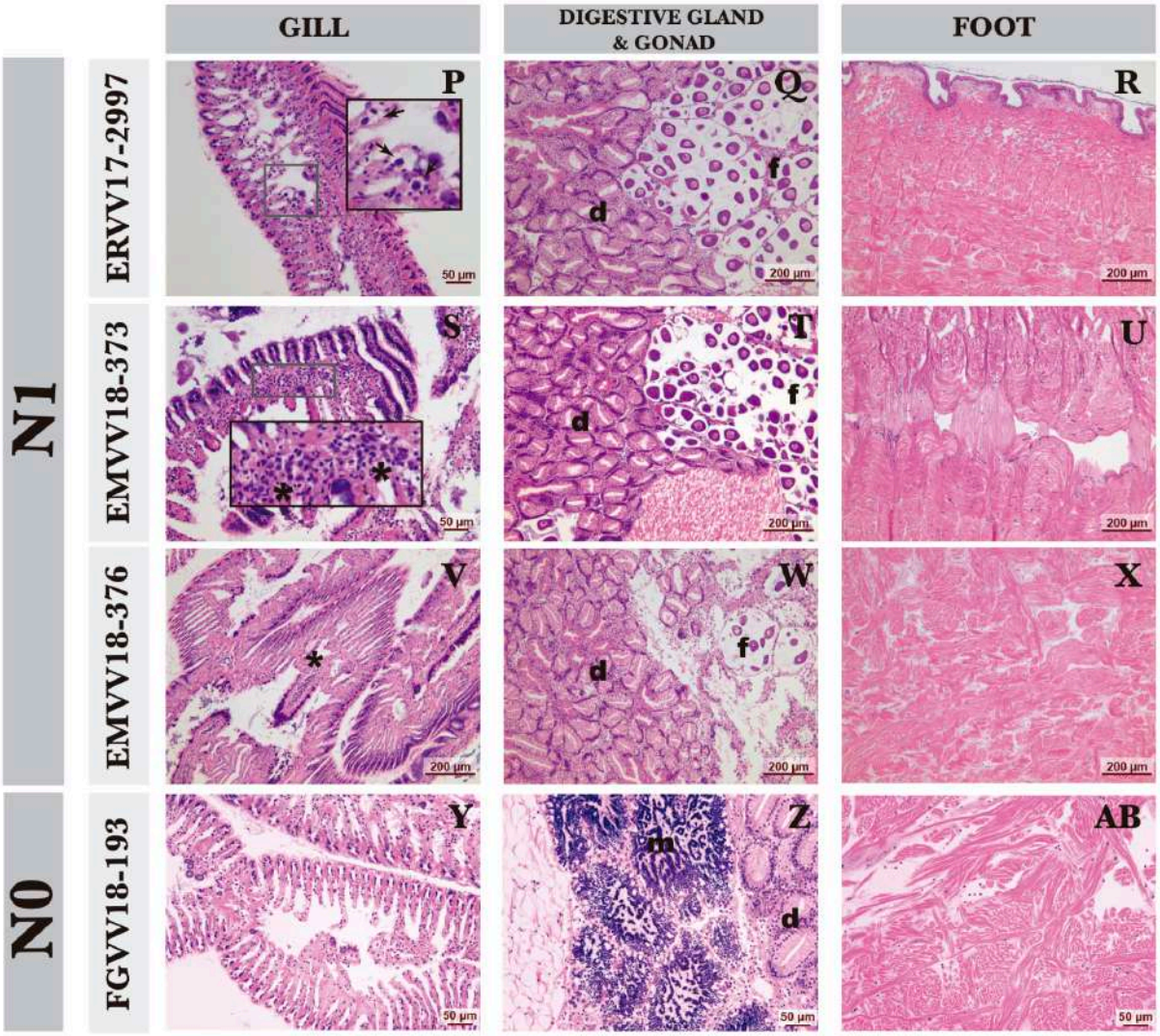
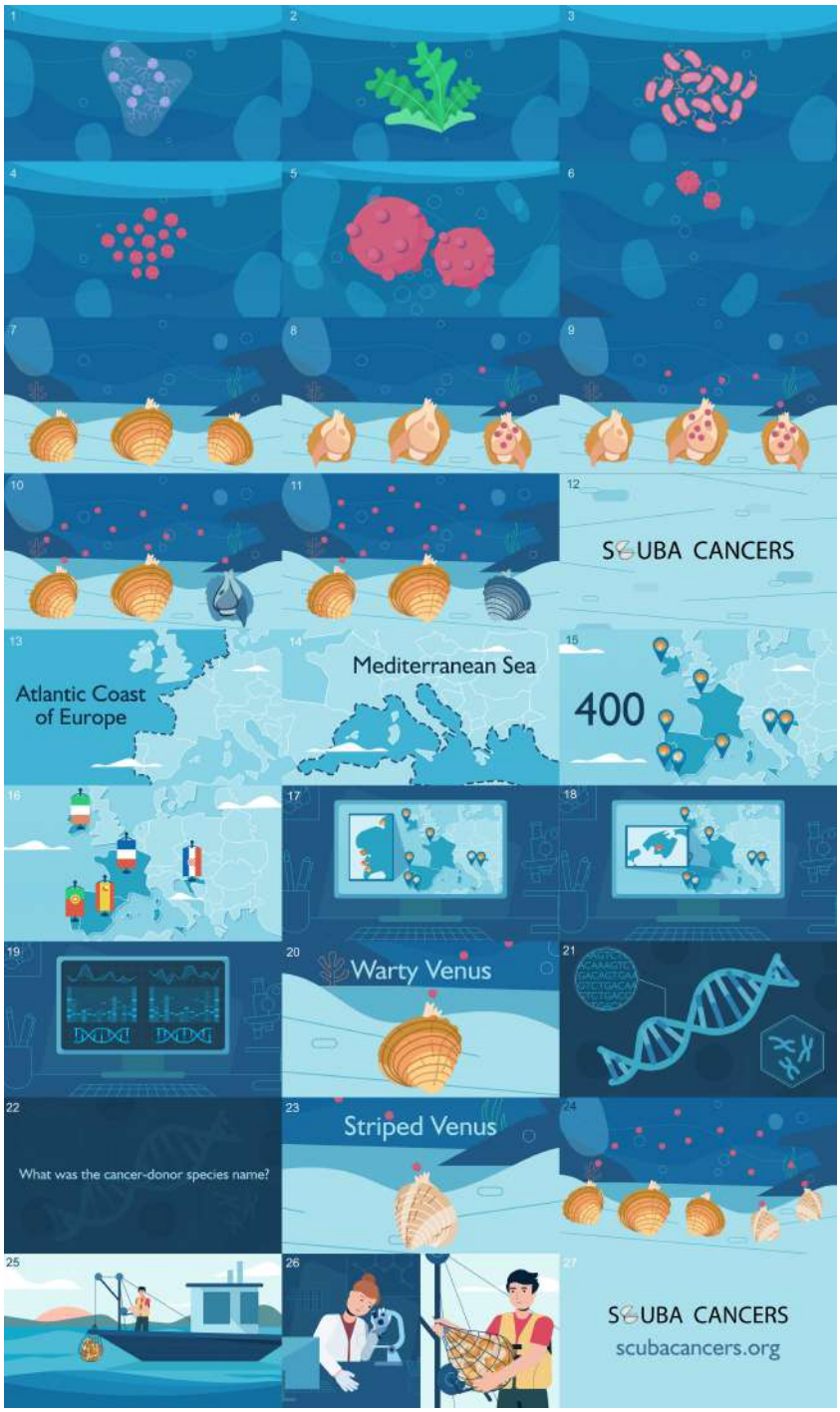


Figure continuation. Histological diagnosis of hemic neoplasia in warty venus (*V. verrucosa*) specimens. Hematoxylin and eosin-stained photomicrographs of gill, digestive (d), gonad (male (m) & female (f)) and foot of warty venus specimens diagnosed with different stages of hemic neoplasia: high (N3), medium (N2), light (N1) and healthy (N0). In the N3 stage, neoplastic cells infiltrate the connective tissue and vessels of different organs (A-L), and show low infiltration of foot (C,F,I,L). In N2 stage, cell groups are observed in different organs such as gills (M) and are not detected in the foot (O). In N1 stage, groups of neoplastic or isolated cells are detected in gill sinuses (P, S, V). N0 stage is completely devoid of any trace of hemic neoplasia at either gill, digestive gland and gonad and foot (Y-AB). Arrows show isolated cells. Asterisks show groups of neoplastic cells. (Source: García-Souto et al. 2022, Copyright 2022, eLife, CC-BY 4.0).

Video 1. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Infographic video outlining the main findings of the research carried out.

https://elifesciences.org/articles/66946/figures#video1



eLife digest. Summary cutting jargon and putting research in context to showcase the articles published in eLife.

In humans and other animals, cancer cells divide excessively, forming tumours or flooding the blood, but they rarely spread to other individuals. However, some animals, including dogs, Tasmanian devils and bivalve molluscs like clams, cockles, and mussels, can develop cancers that are transmitted from one individual to another. Despite these cancers being contagious, each one originates in a single animal, meaning that even when the cancer has spread to many individuals, its origins can be traced through its DNA.

Cancer contagion is rare, but transmissible cancers seem to be particularly common in the oceans. In fact, 7 types of contagious cancer have been described in bivalve species so far. These cancers are known as "hemic neoplasia" and are characterized by the uncontrolled division of blood-like cells, which can be released by the host they developed in, and survive in ocean water. When these cells encounter individuals from the same species, they can infect them, causing them to develop hemic neoplasia too

There are still many unanswered questions about contagious cancers in bivalves. For example, how many species do the cancers affect, and which species do the cancers originate in? To address these questions, Garcia-Souto, Bruzos, Díaz et al. gathered over 400 specimens of a species of clam called the warty venus clam from the coastlines of Europe and examined them for signs of cancer. Clams collected in two regions of Spain showed signs of hemic neoplasia: one of the populations was from Mahón in the Mediterranean Sea, while the other came from the Atlantic coast of north-western Spain.

Analysing the genomes of the tumours from each population showed that the cancer cells from both regions had likely originated in the same animal, indicating that the cancer is contagious and had spread through different populations. The analysis also revealed that the cancer did not originally develop in warty venus clams: the cancer cells contained DNA from both warty venus clams, and another species called striped venus clams. These two species live close together in the Mediterranean Sea, suggesting that the cancer started in a striped venus clam and then spread to a warty venus clam. To determine whether the cancer still affected both species, Garcia-Souto, Bruzos, Díaz et al. screened 200 striped venus clams from the same areas, but no signs of cancer were found in these clams. This suggests that currently the cancer only affects the warty venus clam.

These findings confirm that contagious cancers can jump between clam species, which could be threat to the marine environment. The fact that the cancer was so similar in clams from the Atlantic coast and from the Mediterranean Sea, however, suggests that it may have emerged very recently, or that human activity helped it to spread from one place to another. If the latter is the case, it may be possible to prevent further spread of these sea-borne cancers through human intervention.

Appendix B: Publications reproduced in this thesis

Chapter 4 is the complete reproduction of the following publication:

Title: Mitochondrial genome sequencing of marine leukemias reveals cancer contagion between clam species in the Seas of Southern Europe.

Authors: 20 authors, * denotes equal contribution
Daniel Garcia-Souto* [1,2,3], **Alicia L. Bruzos*** [1,2], Seila Diaz* [1], Sara Rocha[4], Ana Pequeño[1], Camila F. Roman-Lewis[5], Juana Alonso[5,7], Rosana Rodriguez[7], Damian Costas[7], Jorge Rodriguez-Castro[1], Antonio Villanueva[7], Luis Silva[8], Jose Maria Valencia, Giovanni Annona, Andrea Tarallo, Fernando Ricardo, Ana Bratos Cetinic, David Posada, Juan Jose Pasantes, Jose M. C. Tubio.

Affiliations:
[1] Genomes and Disease, Centre for Research in Molecular Medicine and Chronic Diseases (CiMUS), Universidade de Santiago de Compostela, Santiago de Compostela, Spain.
[2] Department of Zoology, Genetics and Physical Anthropology, Universidade de Santiago de Compostela, Santiago de Compostela, Spain.
[3] Cancer Ageing and Somatic Mutation Programme, Wellcome Sanger Institute, Cambridge, United Kingdom.
[4] Phylogenomics Lab, Universidade de Vigo, Vigo, Spain.
[5] CINBIO, Universidade de Vigo, Vigo, Spain.
[6] Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, Vigo, Spain.
[7] Centro de Investigación Mariña, Universidade de Vigo, ECIMAT, Vigo, Spain
[8] Instituto Español de Oceanografía (IEO), Centro Oceanográfico de Cádiz, Cádiz, Spain.
[9] Laboratori d'Investigacions Marines i Aqüicultura, (LIMIA) - Govern de les Illes Balears, Port d'Andratx, Balearic Islands, Spain.
[10]Instituto de Investigaciones Agroambientales y de Economía del Agua (INAGEA) (INIA-CAIB-UIB), Palma de Mallorca, Balearic Islands, Spain.
[11]Stazione Zoologica Anton Dohrn, Napoli, Italy.
[12]ECOMARE, Centre for Environmental and Marine Studies (CESAM), Department of Biology, University of Aveiro, Santiago University Campus, Aveiro, Portugal.
[13]Department of Aquaculture, University of Dubrovnik, Dubrovnik, Croatia.
[14]Department of Biochemistry, Genetics and Immunology, Universidade de Vigo, Vigo, Spain.
[15]Centro de Investigación Mariña, Universidade de Vigo, Vigo, Spain.

Year of publication: 2022
Volume and pages: 11:e66946
Journal: eLife
ISSN: 2050-084X
Publisher: eLife Sciences Publications Ltd
DOI: https://doi.org/10.7554/eLife.66946

Indexed in Web of Science – JCR 2022 impact factor: not available.
2020 impact factor: 8.146 – Q1, Biology

Indexed in Scopus – SJR 2022 impact factor: not available.
2021 impact factor: 4.71 – Q1, Biochemistry, Genetics and Molecular Biology.

Specific contribution of the PhD candidate to the article:
Conception and design of the study, collection and diagnosis of samples, analysis, and interpretation of data, drafting of the manuscript and revision after peer-review.

Appendix C: Extended abstract (Galician language)[10]

O cancro (grego "*karkinos*") recibiu este nome pola semellanza dun tumor de mama coa forma dun cangrexo, un animal que podemos atopar na area das praias ou mesmo dentro das cunchas dun berberecho. O termo foi utilizado por primeira vez por Hipócrates ao redor do 400 a. C. e os gregos comprenderon algúns conceptos clave que influíron na medicina ao longo dos séculos. Por exemplo, déronse conta de que, ao eliminar un tumor, este podía voltar, é dicir, describiron a migración do cancro dunha rexión do corpo a outra que, hoxe en día, coñecemos como metástase.

Catro milenios nos separan da primeira descrición do cancro atopada nun papiro exipcio. Desde entón, a ciencia iluminou moitos aspectos sobre a orixe e o desenvolvemento desta enfermidade. Porén, aínda queda moito por entender, especialmente a respecto dos mecanismos biolóxicos do proceso metastásico que se estima responsable do 90 % de mortes por cancro actualmente.

O *primeiro capítulo* desta tese doutoral introduce conceptos e teorías fundamentais sobre a xenómica do cancro con algunhas racións de historia para dotar ao lector dun estado da arte que lle axude a mergullarse nos capítulos de investigación.

As células cancerosas acumulan mutacións que lles permiten medrar sen control e, eventualmente, adquiren a capacidade de metastizar. Os cancros transmisibles ou contaxiosos son metástases a gran escala en que as células cancerosas se propagan a outros individuos alén do corpo que as orixinou. Ningún virus, bacteria ou parasito infecta o novo hóspede, é a propia célula cancerosa a que se establecerá no novo individuo e despois se dividirá para formar un novo tumor, é dicir, estas células cancerosas adquiren a capacidade de contaxio ou transmisibilidade. Dita capacidade equivale á creación dun novo "parasito" infeccioso: a célula "parasitaria" e cancerosa infectará un individuo diferente ao inicial, dividirase e as súas células fillas continuarán infectando outros individuos.

O primeiro indicio dun cancro transmisible aparece co estudo dun tumor venéreo canino que se coñece desde 1876. A teoría da transmisión deste cancro xurdiu dos experimentos de transmisión artificial e do descubrimento de marcadores xenéticos no século XX. Non obstante, non foi até o século XXI cando estudos de secuenciación demostraron que os xenomas das células de cancro de diferentes individuos eran moi similares entre si, e diferentes aos das células saudables dos seus hóspedes. Por tanto, os cancros contaxiosos adoitan estudarse desde un punto de vista xenético para esclarecer a súa natureza transmisible.

O contaxio de cancro é un fenómeno raro na natureza e a maioría dos cancros permanecen dentro do organismo que os orixinou; no entanto, a pesar do recente descubrimento de cancros contaxiosos, xa se atoparon en varias especies. Hai tres tipos de cancros transmisibles de orixe natural. Un deles, parecido á leucemia que se acha en varias especies de bivalvos mariños, chámase neoplasia diseminada ou neoplasia hémica e hipotetízase que é transmitido pola auga. Os outros dous afectan a mamíferos –o tumor venéreo en cans e un sarcoma facial nos diaños de Tasmania– que, para se contaxiar, requiren do contacto físico entre animais por coito ou mordedura, respectivamente.

[10] Thank you to *Sergio Couso Núñez* for his language corrections and suggestions regarding.

Nesta tese empréganse como modelo para o estudo das metástases contaxiosas os cancros contaxiosos en bivalvos. Desde finais do século XX, describíronse neoplasias hémicas en máis de quince especies de bivalvos, sendo a histoloxía o primeiro método empregado para a súa diagnose, xa que as células cancerosas poden ser diferenciadas morfolóxicamente. Non foi até o 2015 que Metzger e colaboradores estableceron a natureza transmisible da neoplasia diseminada na ameixa de Nova Inglaterra (*Mya arenaria*). Nos anos seguintes, ampliouse para algunhas outras especies coma os berberechos comúns (*Cerastoderma edule*).

A metástase nun novo hóspede non só significa superar as barreiras físicas, senón tamén a resposta inmunolóxica do hóspede, é dicir, para que o contaxio de cancro se produza as células cancerosas precisan de ser éxitosas ao (1) abandonar o hóspede que as orixinou, (2) sobrevivir na auga do mar, (3) alcanzar un novo hóspede e (4) invadilo e adaptarse para evitar as súas respostas inmunes. Non obstante, as neoplasias diseminadas de bivalvos son os únicos cancros contaxiosos de orixe natural que puideron infectar animais dunha especie diferente da que orixinou o cancro, o cal suxire unha certa susceptibilidade destes animais a crear e/ou ser infectados por cancros contaxiosos.

Para alén do anterior, adícase unha sección deste capítulo aos xenomas de referencia de bivalvos mariños, dadas as análises xenómicas descritas nesta tese. As ensamblaxes de calidade para xenomas de bivalvos adoitan ser un reto debido a varios factores, como a composición de elementos repetitivos e os altos niveis de heterocigosidade. Porén, a secuenciación dun xenoma de referencia ofrece información valiosa sobre os xenes implicados en resistencia ás enfermidades e permite comprender as alteracións xenéticas que conducen á infección.

Evolución dos cancros transmisibles en berberechos

A neoplasia hémica dos berberechos (HN polas súas siglas en inglés) descubriuse na década de 1980 simultaneamente en Irlanda e Francia, mais non foi até 2016 que se corroborou a súa natureza contaxiosa de xeito similar a como se revelou o contaxio de cancro nas ameixas de Nova Inglaterra. As análises de microsatélites e ADN mitocondrial de células tumorais e saudables illadas de seis berberechos enfermos revelaron a existencia de, polo menos, dúas liñaxes clonais de cancro independentes que se correspondían con diferenzas histopatolóxicas previamente descritas. Este achado demostrou a orixe polifilética do cancro contaxioso en berberecho, o cal suxeriu que os berberechos están predispostos xenética ou condutualmente a desenvolver cancros transmisibles.

O *segundo capítulo* desta tese doutoral presenta a historia evolutiva da HN describindo a prevalencia desta enfermidade en 6719 berberechos de 36 poboacións ao longo do rango de distribución da especie. O estudo desta historia evolutiva, desentraña e caracteriza múltiples transferencias horizontais de mitocondrias e reporta diversas co-infeccións de dous cancros contaxiosos que afectan a un só individuo.

Os berberechos comúns distribúense de Marrocos a Rusia por toda a costa atlántica de Europa; porén, observouse unha disparidade de prevalencia de HN entre as poboacións de berberecho, con áreas onde a enfermidade alcanza taxas de prevalencia elevadas e outras sen enfermidade ningunha. Foron recollidos berberechos en 36 puntos de mostraxe pertencentes a once países e observouse unha prevalencia global da enfermidade do 5,3 %; emporiso, só se diagnosticou nas rexións do sur da costa atlántica europea (Portugal, España, Francia, Inglaterra e Irlanda). Os nosos resultados mostran unha distribución principalmente continua de HN no sur de Europa con algunhas localizacións esporádicas onde non se encontrou ningún caso de HN (ex. Plymouth UDCE, Arcachon FACE, Bilbao EBCE, Grove EGCE, Placeres ELCE), o cal non significa necesariamente que non exista HN porque, por exemplo, HN en Arcachon

(Francia) xa foi descrito na literatura. Nótese que non achamos ningún HN nos países do norte ou en Marrocos, cuxa costa está fronte á zona portuguesa do Algarve, onde foi observada a maior prevalencia da enfermidade.

Segundo a estrutura poboacional dos berberechos comúns, a variación xenética desta especie caracterízase por dous grupos homoxéneos e diferenciados –sur e norte– e un grupo central heteroxéneo que pode ser unha barreira para a propagación do HN, xa que non encontramos HN nas poboacións do norte. Xunto cos patróns de fluxo xenético das poboacións de berberecho, a densidade e a distancia desas poboacións, as correntes oceánicas ou as condicións físicas mariñas (é dicir, temperatura, salinidade, pH, presión ou CO_2) tamén poden explicar a distribución do HN.

En termos de gravidade, nunca encontramos estadios graves ou medios sen estadios precoces de cancro nun lugar de mostraxe determinado; non obstante, ás veces achamos os tres estadios (ex. Roscoff en Francia, FRCE), mentres que noutros só observamos a fase inicial (ex, Moaña en España, EMCE). No xeral, o 58 % de todas as mostras de cancro recollidas para este estudo foron clasificadas como fase inicial e só o 15 % estaban en fase grave, sendo estas últimas as ideais para a secuenciación, posto que máis do 75 % das células da hemolinfa son cancerosas.

Realizamos a secuenciación do xenoma completo de 70 tumores, que representan o 20 % da nosa colección de tumores, considerando a pureza do cancro, a calidade dos ácidos nucleicos (isto é, integridade, pureza e concentración do ADN) e que todas as poboacións con cancro diagnosticado fosen incluídas. Cando as mostras non cumprían os requisitos de calidade de ADN para a secuenciación, utilizamos un protocolo de amplificación do xenoma completo. Para comparar os xenomas tumorais co fondo xenético da especie, poder filtrar a maior variación posible da liña xerminal e estimar a historia evolutiva do cancro, construímos un panel de 481 individuos normais, incluíndo todas as poboacións recollidas, mesmo aquelas nas que non se diagnosticou ningún cancro. O xenoma de referencia do berberecho común foi ensamblado a nivel cromosómico e anotado cos xenes codificantes e rexións repetitivas. O tamaño de dito xenoma é de 0,8 xigabases, o cal representa un terzo do xenoma humano e está dentro do rango de tamaños dos xenomas de bivalvos.

Para descubrir se varios mitoxenomas (hóspede e tumor) estaban presentes nos berberechos diagnosticados con cancro, analizamos a frecuencia das variantes nucleotídicas. O 63 % das mostras tumorais tiñan dous haplotipos mentres que todos os berberechos sans tiñan só un haplotipo a frecuencia 1. As frecuencias correspondíanse aproximadamente coa cantidade de células que se podían ver na hemocitoloxía. Deconvolucionamos os haplotipos e inferimos unha filoxenia para ver as relacións entre todos os xenomas de berberechos sans e berberechos con cancro empregando catro métodos diferentes. Todas as árbores mostraron nove liñaxes monofiléticas de haplotipos de mitoxenomas do cancro con mitoxenomas de berberechos saudables separándoas. Os haplotipos de berberechos saudables confirmaron os patróns xeográficos de variación xenética descritos anteriormente na literatura (dous grupos homoxéneos no sur e no norte, e un grupo heteroxéneo central). Algunhas liñaxes contan con máis mostras e están máis distribuídas xeograficamente (ex. a liñaxe *HT-Nabia* en España e Portugal ou a liñaxe *HT-Sulis* en Inglaterra e España), e outras encontráronse só nunha poboación (ex. a liñaxe *HT-Sequana* en Francia). O 40 % das poboacións onde se encontrou cancro tiña máis dunha liñaxe de cancro, sendo a liñaxe *HT-Cissonius* a máis estendida (cinco poboacións).

As relacións filoxenéticas das liñaxes do cancro cos seus taxóns irmáns mostraron disparidades. Todas as mostras de cancro acháronse en localidades do sur de Europa, pero algunhas liñaxes agrúpanse principalmente con mostras do norte, mentres que a maioría se agrupan con mostras do sur. Isto, xunto co número de mutacións e intervalos de estimacións de tempo que son amplos, dannos unha idea do antepasado común das liñaxes de cancro, mostra liñaxes moi antigas e outras mais recentes. Ademais, empregáronse análises topolóxicas que descartaron relacións monofiléticas en oito das nove liñaxes atopadas.

Para indagar se as nove liñaxes estaban presentes tamén no ADN nuclear, intentouse empregarse microsatélites como marcadores que diferencian os berberechos sans daqueles que padecen cancro, pero, debido a varios factores que impedían a marcaxe de todas as mostras de cancro, este enfoque foi descartado. Porén, todas as mostras clasificáronse en dous fenotipos (tipo A ou B) atendendo a criterios histopatolóxicos que xa foran previamente asignados a dúas liñaxes clonais empregando marcadores nucleares. As liñaxes mitocondriais de cancro non se correspondían con fenotipos histolóxicos (nove fronte a dous) e non se achou ningunha mestura de fenotipos en ningunha liñaxe mitocondriais de cancro (Figuras 34 e 35), o cal suxire que estas nove liñaxes mitocondrias son transmisións horizontais (HT polas súas siglas en inglés) de mitocondrias capturadas polas células de cancro.

Encontráronse nove mostras con coinfeccións de dúas liñaxes mitocondriais de cancro, o 13 % das mostras tumorais secuenciadas, de modo que a coinfección é relativamente frecuente nos cancros transmisibles de berberecho. Sorprendentemente, unha mostra coinfectada tiña unha liñaxe de cancro mitocondrial ou transferencia horizontal pertencente ao tipo A (*HT-Cissonius*), mentres que a outra pertencía ao tipo B (*HT-Sulis*), podendo ser confirmada por métodos histolóxicos (Figura 36).

En definitiva, o ADN mitocondrial, o nuclear e a histopatoloxía parecen contar diferentes historias do mesmo conto porque, aínda que se encontraron nove liñaxes de cancro analizando os mitoxenomas, non se viron indicios desas liñaxes ao observar o seu fenotipo ou marcadores nucleares. Estes resultados non se poden explicar simplemente por unha alta taxa de mutación nas mitocondrias de mostras de HN e suxiren que as liñaxes HN adquiren periodicamente as mitocondrias dos seus hóspedes, como xa se comprobou que ocorre no cancro transmisible que afecta a cans. Os cancros adoitan caracterizarse por unha alta taxa metabólica (e polo tanto taxa de mutación) e, no caso dos cancros transmisibles que teñen unha vida útil máis longa, as mitocondrias acumulan mutacións nocivas que permiten a selección das células cancerosas que capturan as mitocondrias do seu hóspede.

Análise transcriptómica da orixe histolóxica dos cancros transmisibles en berberechos

A neoplasia hémica é un cancro que afecta a moitas especies de bivalvos en todo o mundo e que se caracteriza pola proliferación de células circulantes anormais de orixe descoñecida que se diseminan polo sistema circulatorio e infiltran diversos tecidos. A nomenclatura neoplasia hémica utilizouse a finais da década de sesenta e o seu uso foi en detrimento porque implicaba a histoxénese da mesma na hemolinfa. No xeral, este cancro considérase que é un sarcoma, é dicir, unha neoplasia dos tecidos derivados do mesodermo, aínda que tamén se propuxo unha orixe hematopoiética e gonadal.

A pesar dos descubrimentos da etioloxía do cancro en canto á súa transmisión, a célula de orixe das células cancerosas no fundador do cancro continúa sendo descoñecida. Dada a existencia de dúas liñaxes clonais de cancro diferenciadas cito-histolóxicamente, dilucidar a histoxénese das mesmas podería axudar a comprender os cambios evolutivos que subxacen a unha célula para converterse en cancerosa e desenvolver un comportamento metastásico que

vai alén dos límites do corpo. Curiosamente, a histopatoloxía e os perfís de expresión xénica dos tumores adoitan permanecer relativamente estables durante a progresión do tumor primario á metástase, proporcionando un bo escenario para investigar a orixe destas células cancerosas mediante a análise transcriptómica.

No *terceiro capítulo*, estudamos a expresión xénica de transcriptoma completo e unha selección de xenes específicos. Para iso, empregamos datos de sete tecidos diferentes de berberechos saudables, catro estadios larvarios e oito animais con cancro clasificados en dous fenotipos (A e B). As análises foron consistentes coa hipótese de que as células cancerosas das dúas liñaxes son derivadas do mesodermo e apuntan a unha orixe hemocitaria.

Unha orixe hemocitaria dos cancros contaxiosos de berberecho contrasta coa orixe do cancro transmisible dos cans, que se propón que é de orixe histiocítica ou coa orixe do cancro transmisible do diaño de Tasmania nunha célula de Schwann.

Contaxio de cáncer entre diferentes especies de ameixa nos mares de Europa do Sur

No *cuarto capítulo* reprodúcese un artigo publicado na revista eLife no que describimos un novo cancro contaxioso que infecta a ameixas nos mares do sur de Europa. Nalgúns dos puntos onde recolliamos berberechos, atopamos ameixas carneiro (*Venus verrucosa*) en que non había reportes de neoplasia hémica. Para indagar se podían estar infectadas de cancro e descubrir a especie de orixe de dito cancro, recollemos máis de 345 exemplares de ameixa carneiro (tamén coñecida como "escupiña" pola costa das illas baleares ou "bolo" polo sur de España) en nove puntos de cinco países do sur de Europa. As ameixas carneiro recollidas en Galicia (noroeste de España, costa Atlántica) e nas Illas Baleares (este de España, costa Mediterránea) presentaban signos de neoplasia hémica.

A análise dos mitoxenomas e de dous xenes nucleares de copia única (*DEAH12* e *TFIIH*) mostrou que as células cancerosas eran similares entre si e diferentes das células saudables da ameixa carneiro hóspede, o cal indicaba que o cancro é contaxioso. Curiosamente, as células cancerosas das dúas rexións con cancro son moi similares, de modo que este cancro estendeuse a diferentes poboacións situadas a máis de 1000 millas de distancia.

A análise das mutacións non revelou diversidade entre os sete tumores secuenciados, o cal apunta a que todos pertencen á mesma liñaxe tumoral que se espalla entre as ameixas carneiro nos mares do sur de Europa. Aínda que ignoramos a idade deste cancro, podemos confirmar que xurdiu antes de 2011, cando se recolleu unha das mostras tumorais analizadas.

A análise tamén revelou que o cancro non se desenvolveu orixinalmente nas ameixas carneiro. As células cancerosas contiñan ADN de dúas ameixas: da ameixa carneiro, tal e como se esperaba, e doutra especie chamada ameixa chirla que cohabita coas carneiro no mar Mediterráneo. O mais probable é que o cancro se orixinase nunha ameixa chirla e saltase a infectar ameixas carneiro. Á luz deste achado, para determinar se o cancro aínda afecta a ambas especies, recollemos e analizamos 200 ameixas chirla, pero non se atoparon signos de cancro nelas, polo que actualmente o cancro só infecta á ameixa carneiro. Para obter máis evidencias sobre a orixe na ameixa chirla, realizamos un cribado de repeticións en tándem nos xenomas das dúas especies mediante hibridación fluorescente *in situ* que marcaron as células cancerosas e as células saudables das ameixas chirla, pero non as células saudables das ameixas carneiro.

O feito de o cancro ser tan semellante nas ameixas da costa atlántica e do mar Mediterráneo fai pensar que puido xurdir moi recentemente ou que a actividade humana axudoulle a se estender dun lugar a outro. Se este último é o caso, pode ser posible evitar unha maior propagación destes cancros mariños mediante a intervención humana.

Para fechar a tese doutoral, o *quinto capítulo* ofrece ao lector unha perspectiva xeral dos achados descritos nos capítulos anteriores, discute as leccións aprendidas xunto coas implicacións e limitacións dos experimentos así como as análises desta tese, e remata coas direccións futuras desta liña de investigación.

En síntese, esta tese doutoral avanza na comprensión do cancro transmisible en bivalvos proporcionando un marco evolutivo robusto para a transferencia horizontal de mitocondrias e informando sobre novos achados non coñecidos anteriormente como a co-infección, a histoxénese ou o cancro que infecta a ameixas carneiro.

Finalmente, iniciei esta tese definindo a etimoloxía do cancro e relacionándoa cos nosos protagonistas -os berberechos- xa que os cangrexos pódense atopar dentro deles. Curiosamente no século XVII, unha pasta barata de ollos de cangrexo era popular para tratar o cancro sen éxito; porén, a investigación mellorou notablemente os nosos tratamentos contra o cancro e estamos de camiño para erradicar o emperador de todas as enfermidades. Este é o meu modesto paso nese camiño.

Conteo de palabras: 3.042

Appendix D: List of collaborators

The list below contains the names and institutional affiliations of the scientists and technical staff who collaborated in Scuba Cancers project which is presented on this thesis or that helped me at some point during the thesis. To all of you, thanks for making this thesis a reality. Apologies are due to any persons who may have been accidentally omitted from this list despite their involvement or collaboration in the project.

Adrián Báez, Dr.	Wellcome Trust Sanger Institute, United Kingdom.
Aisling Smith	Marine Biological Association, United Kingdom.
Alix Harvey	Marine Biological Association, United Kingdom.
Alba Hernández	Toralla Marine Station, Universidade de Vigo, Spain.
Alberte Román	Campus do Mar, Universidade de Vigo, Spain.
Alex Viña	Universidade de Santiago de Compostela, Spain.
Ana Bratos Cetinic, Dr.	Unversity of Dubrovnik, Croatia.
Ana Copena Soutelo	Universidade de Santiago de Compostela, Spain.
Ana Isabel Vidal Garnelo	Cofradía de Pescadores San Martín, Spain.
Ana M. Insua Pombo, Dr.	Universidade da Coruña, Spain.
Ana Margarida Amaral, Dr.	Centro de Ciências do Mar CCMAR, Portugal.
Ana Pequeño	Universidade de Santiago de Compostela, Spain.
Andrea Estrella Arias	Universidade de Santiago de Compostela, Spain.
Andrea Lema	Universidade de Santiago de Compostela, Spain.
Andrés Simón Gómez	Cofradía de Pescadores de San Telmo, Spain.
Antonio Villalba, Dr.	Marine Research Centre CIMA, Spain.
Antonio Villanueva	Toralla Marine Station, Universidade de Vigo, Spain.
Artemis Ntoula	University of Patras, Greece.
Asunción Cao Hermida, Dr.	Marine Research Centre CIMA, Spain.
Birgit Hussel	Alfred Wegener Institute, Germany.
Bouchra El Khalfi, Dr.	Université Hassan II Casablanca, Morocco.
Camila Rolán Lewis	Universidade de Vigo, Spain.
Carla Coedo	Universidade de Santiago de Compostela, Spain.
Carlos Canchaya, Dr.	Universidade de Vigo, Spain.
Carlos García de Leaniz	Swansea University, United Kingdom.
Carmen M. Vidal Álvarez	Cofradía de Pescadores de Camariñas, Spain.
Damian Costas	Toralla Marine Station, Universidade de Vigo, Spain.
Daniel García-Souto, Dr.	Wellcome Trust Sanger Institute, United Kingdom.
Daniel Rey	Campus do Mar, Universidade de Vigo, Spain.
David Iglesias Estepa, Dr.	Marine Research Centre CIMA, Spain.
David Posada, Dr.	Universidade de Vigo, Spain.
Dolores Gondar Meis	Cofradia de Pescadores San Martiño, Spain.
Elisa Gándara	Toralla Marine Station, Universidade de Vigo, Spain.
Fernando Ricardo, Dr.	Universidade de Aveiro, Portugal.
Iago Otero Coto	Universidade da Coruña, Spain.
Ian Probert	Station Biologique de Roscoff, France.
Javier Temes	Universidade de Santiago de Compostela, Spain.
Joán L. Ferreiro Caramés	Cofradía de Pescadores de Barallobre, Spain.
Jorge Alfaya Masó	Cofradía de Pescadores San Francisco, Spain.
Jorge Rodríguez-Castro	Universidade de Santiago de Compostela, Spain.
Jorge Zamora, Dr.	Universidade de Santiago de Compostela, Spain.
Jose A. Santiago Amoedo	Cofradía de Pescadores de Baiona, Spain.

Jose Tubío, Dr.	Universidade de Santiago de Compostela, Spain.
Juanjo Pasantes, Dr.	Universidade de Vigo, Spain.
Katarzyna Smolarz, Dr-	University of Gdansk, Poland.
Kieran O'Halloran	National University Ireland Galway, Ireland.
Laura Iglesias Carballo	Aloya Superior School, Spain.
Laura Tomás López	Universidade de Vigo, Spain.
Lene Friis Møller	National Institute of Aquatic Resources, Denmark.
Leyre Aramburu	Universidade de Santiago de Compostela, Spain.
Liliana Solís	Cofradía de Pescadores San Bartolomé de Noia, Spain.
Maciej Wolowicz, Dr.	University of Gdansk, Poland.
Manuel D. Formoso	Cofradía de Pescadores de Muros, Spain.
Manuel Garcia Graña	Cofradia de Pescadores de Espasante, Spain.
María J. Carballal, Dr.	Marine Research Centre CIMA, Spain.
María Luisa Martínez, Dr.	Universidade da Coruña, Spain.
Maria Skazina	Saint-Petersburg State University, Russia.
Marisa Yonemitsu	Pacific Northwest Research Institute, United States.
Mark Johnson, Dr.	National University Ireland Galway, Ireland.
Marta Ruíz Arribas	Bachelor Degree Thesis, Universidade de Vigo, Spain.
Martín Santamarina	Universidade de Santiago de Compostela, Spain.
Merchi Rodríguez	Universidade de Vigo, Spain.
Michael Metzger, Dr.	Pacific Northwest Research Institute, United States.
Monserrat Camiña, Dr.	Universidade de Santiago de Compostela, Spain.
Nicolas Pade	Marine Biological Association, United Kingdom.
Nita Alonso	Universidade de Vigo, Spain.
Noé Sar	Agrupación de Mariscadoras "Río Anllóns", Spain.
Pablo Balseiro	Universitetet i Bergen, Norway.
Pablo Hortal	Pix Videos Production Company, Spain.
Paula Ferreira	Aloya Superior School, Spain.
Pilar Alvariño	Universidade de Vigo, Spain.
Rachael Giersch	Pacific Northwest Research Institute, United States.
Rachel Parks	Centre for Environment, Fisheries and Aquaculture Science, Scotland.
Ricardo Calado, Dr.	Universidade de Aveiro, Portugal.
Rosana Rodríguez	Toralla Marine Station, Spain.
Sara Lafuente	Universidade da Coruña, Spain.
Sara Rocha, Dr.	Universidade de Vigo, Spain.
Seila Díaz-Costas, Dr.	Universidade de Santiago de Compostela, Spain.
Sergio Permuy Leal	Cofradia de Pescadores Santiago Apostol de Carril, Spain.
Sofía Venzel	*NA*
Sonia Prado, Dr.	Universidade de Vigo, Spain.
Tamara Prieto Fernández	Universidade de Vigo, Spain.
Tim Verstraeten	Ghent University, Belgium.
Thorolf Magnesen	Universitetet i Bergen, Norway.
Urtzi Izagirre	Centre for Experimental Marine Biology & Biotechnology, Spain
Xavier de Montaudouin	Université de Bordeaux, France.
Yasmina Jamardo	Practicum, Aloya Superior School, Spain.
Young Seok Ju, Dr.	Korea Advanced Institute of Science and Technology, South Korea.
Yunah Lee	Korea Advanced Institute of Science and Technology, South Korea.
Zemin Ning, Dr.	Wellcome Trust Sanger Institute, United Kingdom.

Thank you very much · Muchísimas gracias · Moitísimas grazas

Appendix E: Funding

The doctoral candidate Alicia L. Bruzos was supported by a predoctoral fellowship from the **Spanish Ministry of Economy, Industry and Competitiveness (BES2016/078166)** from April 2017 to September 2021. In addition, a short-term research stay in Seattle (USA) was covered with this grant.

The research presented here was funded by **European Research Council (ERC) Starting Grant no. 716290 SCUBA CANCERS** of Jose MC Tubio.

In 2017, the doctoral candidate was awarded an **EuroMarine Young Scientist Fellowship** to attend the training "DNA phylogenies and genealogies: reconstruction and applications" from Universitat de Barcelona (Spain).

In 2018, a short-term research stay at Pacific Northwest Research Institute in Seattle (USA) was covered with the predoctoral fellowship of the Spanish Ministry of Economy, Industry and Competitiveness (BES2016/078166).

In 2019, a short-term research stay at National University of Ireland Galway in Galway (Ireland) was funded by the **ASSEMBLE PLUS Transnational Access program** (European Union's Horizon 2020 research and innovation program, Grant Agreement No. 730984) to develop the project entitled "Finding the clonal structure of cockle's cancer on Ireland".

In 2021, a short-term research stay at Korea Advanced Institute of Science and Technology (KAIST) in Daejeon (South Korea) was funded with a travel grant from **Boheringer Ingelheim Fonds**.

In 2022, my participation in Falling Walls Lab with a pitch about this thesis was funded with a travel grant from the Sociedad Española de Bioquímica y Biología Molecular (SEBBM).

Appendix F: Academic Curriculum Vitae of doctoral candidate

EDUCATION

2016 – 2022* **PhD. in Molecular Medicine**, Centre for Research in Molecular Medicine and Chronic Diseases (CiMUS), Universidade de Santiago de Compostela, Spain. **expected*

2015 – 2016 **MSc. in Bioinformatics**, Universitat Autònoma de Barcelona, Spain.

2011 – 2015 **BSc. in Biology**, Universidade de Santiago de Compostela, Spain.

CURRENT POSITION

2021 – 2024 **Research Assistant.** The Francis Crick Institute, University College of London, UK.

PREVIOUS POSITIONS

2018 – 2021 **Research Assistant.** University of Santiago de Compostela, Spain.

2016 – 2018 **Research Assistant.** University of Vigo, Spain.

Fall 2014 **Museum staff.** Environmental Interpretation Centre of Compostela, Spain.

Summer 2014 **Research Assistant.** Université Libre de Bruxelles, Belgium.

Summer 2013 **Childminder of a French boy.** Aupair World Organization, France.

Summer 2012 **Camp Instructor of Australian students.** Montemar Summer School, Spain.

RESEARCH STAYS (3 international, 1 national)

2021 **Korea Advanced Institute of Science and Technology** (KAIST), Daejeon (South Korea). *Advisor*: Prof. Young Seok Ju. Theme: Analyzing expression and cell origin using RNA-seq data. Funded by a travel grant of Boheringer Ingelheim Fonds from January to April.

2020 **University of Vigo** (UVIGO), Vigo (Spain). *Advisor*: Prof. David Posada. Theme: Phylogenetic inference of mitogenomes, topology testing, dNdS per branch and time estimations. From September to December

2019 **National University of Ireland Galway** (NUIG), Galway (Ireland). *Advisor*: Prof. Mark Johnson. Theme: Studying the clonal structure of contagious cancers in Irish cockles. Funded by AssemblePlus (grant agreement no. 730984) from March to April.

2018 **Pacific Northwest Research Institute** (PNRI), Seattle (USA). *Advisor*: PhD. Michael Metzger. Theme: Gene editing of bivalve genomes using CRISPR. Funded by the Spanish Ministry of Science (Ref. BES-2016-078166) from September to December.

LANGUAGES

English	C1*	IELTS certification in July 2021; CAE certification in December 2014
French	B2*	DELF certification in May 2020
Galician	C1*	Celga certification in July 2011
Spanish	native	

*according to The Common European Framework of Reference for Languages

CERTIFICATIONS

2017 Craft Skipper. Consellería do Mar, Xunta de Galicia, Spain.

2017 Animal Experimentation Certificate (A+B+C+D+E). Centro de Estudios Biosanitarios, Spain.

2014 Open Water Diver License. Scuba Schools International (SSI), Spain.

AWARDS, GRANTS & FELLOWSHIPS (last 10 years)

2022 **Travel grant**. Sociedad Española de Bioquímica y Biología Molecular, Spain.
2022 **3MT Video Award: Cuentame11F**. Citizen initiative 11 febrero, Spain. *First place*.
2022 **Women's Name Award: promotion of scientific vocations**. Fundación Merck Salud, Spain.
2021 **Travel grant**. Boehringer Ingelheim Fonds, Germany.
2019 **Best Scientific Panel Prize**. XXII Foro dos Recursos Mariños e da Acuicultura das Rías Galegas (ForoACUI), Spain.
2017 **EuroMarine Young Scientist Fellowship.** Universidade de Vigo & European Marine Research Network, Spain.
2017 **Best Scientific Panel Prize**. XX Foro dos Recursos Mariños e da Acuicultura das Rías Galegas (ForoACUI), Spain. *First place*.
2017 **National doctoral fellowship (2017-2021).** Spanish Ministry of Economy, Industry and Competitivity (Ref. BES-2016-078166). Universidade de Vigo, Spain.
2015 **National Speech Competition: Transmito, luego existo**. Club de Debate Compostela. *Finalist*.
2013 **Erasmus Scholarship**. Université Libre de Bruxelles, Belgium. One year to study abroad.
2013 **Open Software Translation Award**. Engineering Technical Superior School of Universidade de Santiago de Compostela. *First place*.

SCIENTIFIC PUBLICATIONS

PhD publications:

García-Souto, D.#, **Bruzos, A.L.#**; Díaz, S.# et al. 2022. "Mitochondrial genome sequencing of marine leukemias reveals cancer contagion between clam species in the Seas of Southern Europe." BioRxiv. ***eLife,*** 11:e66946. Q1, IF (2020) = 8.14, open access (# denotes co-first authorship) | *A novel hemic neoplasia is described for the warty venus clam and it is further investigated revealing a cross-species cancer transmission found in the Mediterranean and Atlantic coasts of Spain.* DOI: https://doi.org/10.7554/eLife.66946

PhD publications. In preparation:

Bruzos, A.L.#; Díaz-Costas, S.#; Santamarina, M.# et al. 2022 "The evolutionary history of a marine contagious cancer spreading along the Atlantic coast of Europe." Target journal: Nature/Science (Q1, open access) | *The flagship of the Scuba Cancers project that describes the main scientific findings from the analyses of ~500 whole genomes to find the genetic causes of the leukaemia-like cancers transmitted among cockle populations. I have been involved in all the experiments, analysis and discussions that gave birth to this paper, some parts are included in this doctoral thesis.*

Bruzos, A.L. et al. 2022* "Transcriptomic analyses of cockle transmissible cancers reveal the cell-of-origin of a long lifespan cancer." Target journal: Communications Biology (Q1, open access) | *Here, we analysed gene expression of transcriptome data and a selection of tissue specific genes to study the cell-of-origin of cockle contagious cancer lineages providing fundamental insights into the histogenesis and opening a framework to investigate the origin of marine contagious metastases.*

Díaz, S. et al. (includes **Bruzos, A.L.**). 2022* "Molecular markers for the diagnosis of two histological cancer types of cockle transmissible cancers" Target journal: unknown. | *This manuscript will describe a genetic assay designed with nuclear markers to differentiate neoplasia A and B of cockle contagious cancers.*

Other publications not related with my PhD thesis theme:

The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (includes **Bruzos A.L.**). 2020. "Pan-Cancer analysis of whole genomes". ***Nature***, 578:82-93. Q1, IF = 43.070, open access | *The flagship of the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium that describes the main scientific findings from the analyses of ~3,000 cancer whole genomes and their matched normal tissues across 38 tumour types. I did bioinformatic analysis during my Master Thesis that ended up being one of the figures of this paper.* DOI: https://doi.org/10.1038/s41586-020-1969-6

Rodriguez-Martin, B. et al. (includes **Bruzos A.L.**). 2020. "Pan-Cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition". ***Nature Genetics***, 52:306-319. Q1, IF = 25.455, open access | *Here we identify a new mutational mechanism in cancer by with aberrant integration of L1 elements promote chromosomal rearrangements such as Megabase-size genomic deletions that on occasion remove tumour suppressor genes, contributing to cancer progression. I catalogued the retrotransposition events of LINE-1.* DOI: https://doi.org/10.1038/s41588-019-0562-0

Álvarez, E. G. et al. (includes **Bruzos A.L.**). 2021. "Aberrant integration of Hepatitis B virus DNA promotes major restructuring of human hepatocellular carcinoma genome architecture." ***Nature Communications***, *12:6910*. Q1, IF (2020) = 14.92, open access | *Here, we identify and characterize a remarkable mutational mechanism in human hepatocellular carcinoma caused by Hepatitis B virus, by which DNA molecules from the virus are inserted into the tumour genome causing dramatic changes in its configuration, including non-homologous chromosomal fusions and megabase-size telomeric deletions. I participated in the wet-lab validation of the mechanism.* DOI: https://doi.org/10.1038/s41467-021-26805-8

Ricardo, F.; Mamede, R.; **Bruzos, A.L.**; Díaz, S.; Thébault, J.; Ferreira da Silva, E.; Patinha, C.; Calado, R. 2021. "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,*** 814:152304 Q1, IF (2020) = 7.96, open access | *Here, we used elemental fingerprints of cockle shells of more than 25 sampling locations to trace the geographic origin of the samples.* DOI: http://dx.doi.org/10.1016/j.scitotenv.2021.152304

Albuixech-Martí, S.; Lynch, S.A.; Díaz, S.; **Bruzos, A.L.**; Skujina, I.; Ironside, J. E.; Culloty, S.C. 2022. "Geographical distribution and abundance of significant pathogens associated with *Cerastoderma edule* along the Irish and Welsh coasts." ***PhD thesis chapter*** of Albuixech-Martí, S. (2022) entitled "Disease dynamics and parasitic transmission between *Cerastoderma edule* and shorebirds in the Irish coast". University College Cork, Ireland.

Hermida, M.; Robledo, D.; Díaz, S.; Costas, D.; **Bruzos, A.L.**; Blanco, A.; The Cockle's Consortium; Martínez, P. 2022. "The broad shell colour variation in common cockle (*Cerastoderma edule*) from Northeast Atlantic relies on a major QTL revealed by GWAS using a new high- density genetic map." ***BioRxiv,*** *preprint* | In this work we identified a QTL on chromosome 13 that explains most of the variation for shell colour patterns of cockles. DOI: https://doi.org/10.1101/2022.04.13.48 8192

CONFERENCES AND CONGRESSES[11]

Oral communications (18)

Bruzos, A. L.; Díaz, S.; Santamarina, M; Rocha, S; Otero, I; Zamora, J; García-Souto, D.; Pequeño, A; Rodríguez-Castro, J.; Temes, J; Posada, D; Tubío, J. (2022, July). *Unlocking the evolutionary history of cockle contagious metastases*. Oral communication in **SEB Annual Conference**. Montpellier (France). ◉

Bruzos, A. L.; Díaz, S.; Rocha, S; Zamora, J; Santamarina, M; Otero, I; García-Souto, D.; Rodríguez-Castro, J.; Pequeño, A; Temes, J; Posada, D; Tubío, J. (2022, July). *Beyond the limits of metastasis: marine contagious cancers*. Oral communication in **IX SRUK/CERU International Symposium**. Oxford (UK). ◉

Bruzos, A. L.; Díaz, S.; Rocha, S; Zamora, J; Santamarina, M; Otero, I; García-Souto, D.; Rodríguez-Castro, J.; Pequeño, A; Temes, J; Posada, D; Tubío, J. (2022, March). *Heterogeneity of cockle transmissible cancers*. Oral communication in **19th Portugaliae Genetica**. Lisbon, virtual edition (Portugal). ◉

Bruzos, A. L.; Díaz, S.; García-Souto, D.; Rocha, S; Temes, J; Zamora, J; Rodríguez-Castro, J.; Posada, D; Tubío, J. (2021, April). *Scuba Cancers: finding the genetic causes of contagious metastases under the sea*. Oral communication in Schwarz Lab Seminars at Max Delbrück Center for Molecular Medicine, Berlin, virtual edition (Germany). ◉

Bruzos, A. L.; García-Souto, D.; Díaz, S.; Rocha, S; Pequeño, A; Roman-Lewis, C; Alonso, J; Rodriguez, R; Costas, D; Rodríguez-Castro, J.; Villanueva, A; Silva, L;Valencia, JM; Annona, G; Tarallo, A; Ricardo, F; Bratos, A; Posada, D; Pasantes, J; Tubío, J. (2021, July). *Contagious Cancers: sequencing reveals a pandemic*

[11] Highlighting: Underlined author presenting the communication; bold CV's author and conference name; ◉ international (excluding Spain); ★award; ▶ book chapter with ISBN.

affecting clams in our seas. Oral communication in **Centro de Ciências do Mar (CCMAR)**. Universidade do Algarve, virtual edition (Portugal). ◉

Bruzos, A. L.; Díaz, S.; García-Souto, D.; Rocha, S; Temes, J; Zamora, J; Rodríguez-Castro, J.; Posada, D; Tubío, J. (2021, May). *Multiple cancer clones metastasize the Atlantic Coast of Europe*. Oral communication in **VIII Young Researchers Meeting**. Universidade de Santiago de Compostela, Santiago de Compostela (Spain).

Bruzos, A. L.; Díaz, S.; García-Souto, D.; Rocha, S; Temes, J; Zamora, J; Rodríguez-Castro, J.; Posada, D; Tubío, J. (2021, April). *Unravelling the genetic story of marine contagious metastases*. Oral communication in **CIBIO-InBIO Seminar Series**. Universidade de Porto, virtual edition (Portugal). ◉

Bruzos, A. L.; Díaz, S.; García-Souto, D.; Temes, J.; Zamora, J.; Tubío, J. (2019, October). *What can we learn from transmissible cancers to treat cancer?* Oral communication in **VI Edición Investigación Biomédica del Cáncer en Galicia (IBCG)**. A Coruña (Spain).

Díaz, S.; **Bruzos, A. L.;** Temes, J.; Johnson, M.; Tubío, J. (2019, September). *Distribution of cockle disseminated neoplasia in Ireland: evaluation of their clonal structure*. Oral communication in **19th International Conference on Diseases of Fish and Shellfish**. Porto (Portugal). ◉

Bruzos, A. L.; Díaz, S.; García-Souto, D.; Temes, J.; Zamora, J.; Tubío, J. (2019, July). *Scuba Cancers: Finding the genetic causes of contagious metastases under the sea*. Oral communication in **Evolution and Ecology of Cancer**. Wellcome Genome Campus, Hinxton-Cambridge (UK). ◉

Bruzos, A. L.; Díaz, S.; Temes, J.; Rodríguez-Castro, J.; García-Souto, D.; Zamora, J.; Tubío, J. (2019, July). *Single-molecule sequencing of whole mitochondrial genomes reveals the clonal structure of cockle transmissible cancers*. Oral communication in **II Annual CIMUS workshop**. Center for Research in Molecular Medicine and Chronic Diseases (CIMUS), Santiago de Compostela (Spain).

Bruzos, A. L.; Díaz, S.; Rodríguez-Castro, J.; Tubío, J. (2019, May). *Colchicine effect on DNA integrity for the study of cockle transmissible cancers*. Oral communication in **VII Young Researchers Meeting**. Universidade de Santiago de Compostela, Santiago de Compostela (Spain).

Bruzos, A. L.; Díaz, S.; Temes, J.; García-Souto, D.; Zamora, J.; Tubío, J. (2018, December). *Bivalve transmisible cancers: excellent models to study metastasis*. Oral communication in **International Young Investigator Meeting (IYI)**. Museo Nacional de Ciencia y Tecnología (MUNCYT), A Coruña (Spain).

Bruzos, A. L.; García-Souto, D.; Díaz, S.; Zamora, J.; Temes, J., Rodríguez-Castro, J.; Tubío, J. (2018, June). *Cockle Reference Genome: DNA isolation challenges*. Oral communication in **VI Young Researchers Meeting-Health Science**. Universidade de Santiago de Compostela, Santiago de Compostela (Spain).

García-Souto, D.; **Bruzos, A. L.**; Díaz, S.; Santamarina, M.; Zamora, J.; Pasantes, J.; Tubío, J. (2018, June). *Molecular cytogenetic analysis of bivalve tumors*. Oral communication in **VI International Symposium on Marine Sciences** (ISMS), Vigo (Spain).

García-Souto, D.; Carpena, M.; **Bruzos, A. L.;** Díaz, S.; Santamarina, M.; Zamora, J.; Tubío, J.; Pasantes, J. (2018, June). *Análisis citogenético molecular en tumors de bivalvos*. Oral communication in **X Seminario de Citogenética de la SEG**. Sociedad Española de Genética, A Coruña (Spain).

Bruzos, A. L.; Díaz, S.; García-Souto, D.; Zamora, J.; Tubío, J. (2017, December). *Cockle reference genome: size estimation*. Oral communication in **33º Encuentro de Jóvenes Investigadores**. Instituto de investigaciones científicas y ecológicas (INICE), Salamanca (Spain). ► Book chapter in: ***Jóvenes investigadores***, 47- 51. ISBN: 978-84-945079-9-1; legal deposit: S.469-2017. | *Size estimation of the cockle reference genome to study cockle transmissible cancers*.

Bruzos, A. L.; Tubío, J. (2017, May). *Finding the genetic causes of contagious metastases under the sea*. Oral communication in **Assemblée Générale EMBRC-France**. EMBRC, Bannyuls-sur-mer (France). ◉

Poster communications (13)

Barberan Martín, S.; Polubothu, S.; **Bruzos, A. L.**; Bulstrode, N.; Spence, G.; Kinsler, V. (2022, April). Mosaic BRAF fusions are a recurrent cause of multiple Congenital Melanocytic Naevi. Poster at **AACR Annual Meeting**. New Orleans (United States). ◉

Bruzos, A. L.; Díaz, S.; Santamarina, M.; Rocha, S; Otero, I.; García-Souto, D.; Zamora, J; Pequeño, A.; Temes, J; Rodríguez-Castro, J.; Posada, D; Tubío, J. (2021, November). *Clonal Structure of Cockle Transmissible*

Cancers. Poster at **3rd ASEICA Educational Symposium**. Spanish Association for Cancer Research (ASEICA), Virtual.

Barberan Martín, S.; Polubothu, S.; **Bruzos, A. L.**; Bulstrode, N.; Spence, G.; Kinsler, V. (2021, November). Mosaic BRAF fusions are a recurrent cause of multiple Congenital Melanocytic Naevi. Poster at **5th Crick Autumn Science Meeting**. Francis Crick Institute, London (United Kingdom). ◉

Bruzos, A. L.; Díaz, S.; García-Souto, D.; Rocha, S; Temes, J; Zamora, J; Rodríguez-Castro, J.; Posada, D; Tubío, J. (2021, June). *Multiple cancer clones metastasize the Atlantic Coast of Europe*. Poster at **EACR 2021 Congress, Innovative Cancer Science: Better Outcomes through Research**. European Association for Cancer Research (EACR), Virtual. ◉

Bruzos, A. L.; Díaz, S.; García-Souto, D.; Rodríguez-Castro, J.; Tubío, J. (2019, October). *WGS y cromosomas de berberechos con cáncer transmissible: efectos en la integridad del ADN*. Poster at **XXII Foro dos Recursos Mariños e da Acuicultura das Rías Galegas**. Universidade de Santiago de Compostela, O Grove (Spain). ★ **Best Scientific Panel Prize** (second prize). ▶ Book chapter in: Rey-Méndez M., et al. Foro dos Recursos Mariños e da Acuicultura das Rías Galegas. 22: 265- 273. ISBN: 978-84-09-19360-8; legal deposit: C 2014-2016 | *Comparison and results of several DNA isolation protocols in cockle cancer samples to meet the requirements of whole genome sequencing (WGS).*

Bruzos, A. L.; Lafuente, S.; Tubío, J.; Díaz, S. (2019, October). *Braquiuros en berberechos: ¿parasitismo o amensalismo?* Poster at **XXII Foro dos Recursos Mariños e da Acuicultura das Rías Galegas**. Universidade de Santiago de Compostela, O Grove (Spain). ★ **Best Scientific Panel Prize** (first prize). ▶ Book chapter in: Rey-Méndez M., Fernández Casal J, Lastres M.A., Padín X.A. (Eds.). Foro dos Recursos Mariños e da Acuicultura das Rías Galegas. 22: 257-264. ISBN: 978-84-09-19360-8; depósito legal: C 2014-2016 | *Description of crayfish that we have found inside cockles' shells during the samplings of scuba cancers project.*

Díaz, S.; **Bruzos, A. L.;** García-Souto, D.; Tubío, J. (2019, May). *Germinoma en berberechos* Cerastoderma edule *de Dinamarca: caracterización histopatológica y genética*. Poster at **XVII Congreso Nacional de Acuicultura**. Cartagena (Spain). ▶ Book chapter in: Martínez, F.J., et al. Libro de resúmenes CNA. 2019:200-201. ISBN: 978-84-09-11292-0 | *First description of gonadal neoplasia in cockles from Denmark.*

Díaz, S.; **Bruzos, A. L.;** García-Souto, D.; Román Lewis, C.; Rodríguez-Castro, J.; Tubío, J. (2019, May). *Neoplasia diseminada en* Venus verrucosa*: ¿un nuevo caso de cáncer transmisible?* Poster at **XVII Congreso Nacional de Acuicultura**. Cartagena (Spain). ▶ Book chapter in: Martínez, F.J., et al. Libro de resúmenes CNA. 2019:198-199. ISBN: 978-84-09-11292-0 | *First description of disseminated neoplasia in the clam Venus verrucosa.*

Bruzos, A. L.; García-Souto, D.; Díaz, S.; Zamora, J.; Temes, J.; Rodríguez-Castro, J.; Tubío, J. (2018, June). *Optimizing DNA extraction protocols to study bivalve transmissible cancers*. Poster at **I Annual CIMUS workshop**. Center for Research in Molecular Medicine and Chronic Diseases (CIMUS), Santiago de Compostela (Spain).

Díaz, S.; **Bruzos, A. L.**; García-Souto, D.; Tubío, J. (2018, June). *Features of transmissible cancer cells of Cerastoderma edule*. Poster at **I Annual CIMUS workshop**. Center for Research in Molecular Medicine and Chronic Diseases (CIMUS), Santiago de Compostela (Spain).

García-Souto, D.; **Bruzos, A. L.**; Díaz, S.; Santamarina, M.; Zamora, J.; Pasantes, J.; Tubío, J. (2018, June). *Epigenetic DNA methylation study of bivalve tumors*. Poster at **VI International Symposium on Marine Sciences** (ISMS), Vigo (Spain).

Santamarina, M.#; **Bruzos, A. L.#**; Díaz, S.; García-Souto, D.; Zamora, J.; Tubío, J. (2017, October). *Transmissible Cancers: a new paradigm in Cancer Evolution*. Poster at **I Annual Meeting CINBIO**. Centro de Investigaciones Biomédicas (CINBIO), Vigo (Spain). #equal contribution

Bruzos, A. L.; Díaz, S.; García-Souto, D.; Zamora, J.; Tubío, J. (2017, October). *Primeros pasos para decodificar el genoma del berberecho C. edule*. Poster at **XX Foro dos Recursos Mariños e da Acuicultura das Rías Galegas**. Universidade de Santiago de Compostela, O Grove (Spain). ★ **Best Scientific Panel Prize** (first prize). ▶ Book chapter in: Rey-Méndez M., et al. Foro dos Recursos Mariños e da Acuicultura das Rías Galegas. 20: 227-237. ISBN: 978-84-09- 01474-3 | *Preliminary results of the cockle reference genome required to study cockle transmissible cancers.*

MENTORING

2020/2021 "Molecular diagnostic of different clonal lineages of transmissible cancer." Andrea E. Arias Díaz. **Degree Thesis for the BSc. in Biology of Universidade de Santiago de Compostela**. *Supervisors*: Alicia L. Bruzos and Jose Tubío. Awarded for linguistic quality by the USC Language Normalization Service in the category of *sciences*.

2019/2020 "Identification of driver mutations in marine transmissible cancers through exome sequencing analysis". Iago Otero Coto. **Master Thesis for the MSc. in Bioinformatics from Universidade da Coruña**. *Supervisors*: Alicia L. Bruzos and Jose Tubío.

2019/2020 "Characterization of a genomic molecular marker of a marine transmissible cancer". Ana Copena Soutelo. **Degree Thesis for the BSc. in Biology of Universidade de Santiago de Compostela**. *Supervisors*: Alicia L. Bruzos and Jose Tubío.

TEACHING EXPERIENCE (total: 90 hours)

2021/2022 **Medical Student-Selected Component**. University College of London. MBBS BSc. in Medicine. Hours: 8 *(English)*

2020/2021 **Genetics II.** Universidade de Santiago de Compostela. BSc. in Biotechnology. Hours: 8 *(Spanish)*

2020/2021 **Genetics I.** Universidade de Santiago de Compostela. BSc. in Biology. Hours: 40 *(Spanish)*

2019/2020 **Human Genetics.** Universidade de Santiago de Compostela. BSc. in Biology. Hours: 2 *(Spanish)*

2019/2020 **Genetics I.** Universidade de Santiago de Compostela. BSc. in Biology. Hours: 13 *(Spanish)*

2018/2019 **Genetics II.** Universidade de Santiago de Compostela. BSc. in Biology. Hours: 16 *(Spanish)*

2017/2018 **Genetics II.** Universidade de Vigo. BSc. in Biology. Hours: 3 *(English)*

INSTITUTIONAL CITIZENSHIP

Service

26-28/05/2021 *Conference organization*. Member of the organizing and scientific committee for the **VIII Youth Researchers Meeting** in Santiago de Compostela (Spain) with 300 attendees. The programme comprised three plenary talks, three parallel sessions gathering 135 short oral communications and exhibited 67 posters.

2021 – 2022 *PhD alumni representative*. Spokeswoman of doctoral students on the **International Doctoral School Direction Committee** (EDIUS). Universidade de Santiago de Compostela, Spain.

2017 – 2018 *PhD alumni representative*. Spokeswoman of doctoral students on the **International Doctoral School Committee** (EIDO) and representative of the Quality Commission of the same body. Universidade de Vigo, Spain.

2016 – 2018 *Consortium secretariat.* Secretary of the **International Common Cockle Genome Consortium** (ICCGC). Three face-to-face symposiums were organized to address the goals of building a reference genome for the cockle species.

Professional membership

2021- **Society of Spanish Researchers in the United Kingdom** (SRUK/CERU). Active member of the Wom=n Equity & Research Committee.

2021- **Spanish Society for Biochemistry and Molecular Biology** (SEBBM). Member; number: 8706; proposing partner Dr. María Mayán.

2020- **Confederación Intersindical Galega** (CIG). Member of this labor union.
2019- **InvestiGal**. Member.
2019- **Asamblea de Investigadoras de Compostela** (AIC). Member.
2019- **European Association for Cancer Research** (EACR). Member and Ambassador; number: 28748.
2019- **Asociación Española de Investigación sobre el Cáncer** (ASEICA). Member; number: 1746.

Others

Official invitation to the regional government. Meeting with the President of Xunta de Galicia chaired by Alberto Núñez Feijóo and the Minister of Economy, Business and Innovation chaired by Francisco Conde López *(19/02/2020)* for the advances made in the field of cancer genomics.

OUTREACH

Activities

2022 *Falling Walls Lab*. Falling Walls is an international pitch competition about science, business, politics, arts and society. I was finalist in the Spanish national competition with the talk "Breaking the Walls of Cancer Metastasis".

2022 *Pint of Science Festival*. PoS is an annual science festival that aims to communicate contemporary scientific developments to the public by bringing scientists to pubs, cafés and other public places to share their research and findings. 9-11 May 2022. I gave the talk "The journey of a cancer cell" at Ink@84, London, United Kingdom. https://pintofscience.co.uk/event/demystifying-diseases

2022 *British Science Week*. BSW is a ten-day celebration of science, technology, engineering and maths that took place between 11-20 March 2022. I gave two 15-min talks addresed to ~100 scholars on a primary school (Rotherfield Primary School, London, UK) on March 14th.

2022 *Women in Science talk*. **11 de febrero** is a citizen initiative to commemorate the International Day of Women and Girls in Science through activities to visibilize the work of women who are dedicated to STEM areas and create female role models for children who can contribute to the choice of these areas as professional careers. My talk was addresed to students of a public vocational trainning (IES Federica Montseny, Valencia, Spain).

2022 *Women in Science talk*. **Conócelas-ASEICA.** A project to encourage girls to study STEM organized by the Spanish Association of Cancer Research. I gave a talk entitled "Una detective del cáncer" to 8-9 years old scholars of a public primary school (CEIP Monte dos Postes, Santiago de Compostela, Spain).

2021 *Skype a Scientist*. **Beaulieu Convent School** (Jersey, UK). In pandemic time, this imitative started to connect real-life scientists with classrooms across the globe.

2021 *Women in STEM*. **Acland Burghley School** (London, UK). School's networking event for Year 9 female students.

2021 *Children's colouring science book*. The project Scientists Meet the Artists joined 12 scientists and 12 illustrators to design a series of drawings for children to colour and learn marine scientific concepts. The book was presented on the **World Oceans Day** (June 8th).

2021 *High school talk*. Semana de orientación laboral. To help students in their **career orientation**, a talk focused on STEM and particularly in biology, biotechnology, and biochemistry was given at Colegio M. Peleteiro (Santiago de Compostela, Spain).

2021 *Women in Science talk*. **Conócelas-ASEICA.** To encourage girls to study STEM, the Spanish Association of Cancer Research organizes the programme "conócelas" where I gave a talk entitled ¿Qué nos puede enseñar un berberecho sobre el cáncer? (Santiago de Compostela, Spain).

2020 *Research promotional film*. Our regional government financed **short films of outstanding research projects** that were developed in Galicia. Link: www.youtube.com/watch?v=Ig3-LggH9Rs

2020 *Research promotional film*. Our research institute (CiMUS) recorded a promotional video with the participation of the authors involved in the **Pan-cancer initiative** when the results were published in high impact journals. Link: www.youtube.com/watch?v=1fm9kL94xn0

2019 *Open doors day*. Participation as an instructor in workshops, conferences and debates for kids and adults in the event **Ciencia Sigular of CiMUS** (Santiago de Compostela, Spain).

2019 *Research video*. **ASEICA video competition**. Youth scientists talk about their research in three minutes, I reached more than 11,600 views in Twitter. Link: www.twitter.com/BruzosAliciaL/status/1196606566365089792

2019 *Science outreach talk*. "Understanding metastasis through transmissible cancers" **Happy Fridays** (Santiago de Compostela, Spain).

2019 *Workshop for children*. Monitor of activities for pupils aged 8-10 organized for 9 primary-school classes in the **International DNA Day** (Compañía de María, Santiago de Compostela, Spain).

2018 *Science outreach talk*. "Unravelling cancer evolution using cockles" **Café con Sal** (Vigo, Spain).

Opinion Editorials and Popular Science articles

2022 *Opinion editorial*. Brief opinion article for the regional newspaper *La Voz de Galicia* entitled "El cáncer se contagia en el mar". La Voz de Galicia (20/01/2022, page 13).

2021 *Popular science article*. Overview about cancers that can be contagious, article entitled "El cancer se puede contagiar (al menos en animales)" published in **The Conversation** (open access news source). https://theconversation.com/el-cancer-se-puede-contagiar-al-menos-en-animales-163529

2021 *Opinion editorial*. Brief opinion article for the regional newspaper *La Voz de Galicia* entitled "Virus y cancer: una pareja peligrosa". La Voz de Galicia (20/12/2021, page 11).

2021 *Institutional bulletin*. Brief article to encourage doctoral students to plan **research stays abroad** entitled "Como che vai na túa estadía?" nEDIUS 3:15.

2020 *Popular science blog article*. "How my master's thesis on jumping genes became part of an article in Nature" **The Cancer Researcher** (EACR online magazine). https://magazine.eacr.org/how-my-masters-thesis-on-jumping-genes-became-part-of-an-article-in-nature/

2019 *Popular science article*. "Las claves de las metástasis enterradas en la arena." **Encuentros en la Biología** 169: 5–7. ISBN 2254-0296.

2018 *Institutional blog article*. "Cáncer transmisible de bivalvos para desentrañar la evolución del cáncer." **Océano Ecimat**. Link: www.oceanoecimat.wordpress.com/2018/05/11/cancer-transmisible-de-bivalvos-para-desentranar-la-evolucion-del-cancer/

MEDIA AND PRESS COVERAGE

TV

TVG. Short interview for the **Midday Newscast** of the regional television of Galicia about the Pan-cancer initiative (06/02/2020)

Radio

BBC Cambridgeshire and *BBC5 live*. Interview to talk about marine contagious cancers (12/04/2022 and 17/04/2022).

CRTVG. Interview for the radio show **Convivir** to talk about my research within the framework of Scuba Cancers project (21/01/2022).

CUAC FM. Interview for the radio show **Ciencia es Femenino** to talk about my research and the role of women in science (20/06/2021)

CRTVG. Interview for the radio show and podcast **Efervesciencia** to talk about science, my PhD project and our recent publications in the framework of the Pan-cancer initiative (20/02/2020)

CRTVG. Live interview in the show **Galicia Por Diante** of the regional radio of Galicia about the Pan-cancer initiative (06/02/2020)

Newspapers

El Correo Gallego. Short interview and picture for the local newspaper to the scientific committee organizing the VIII Youth Researchers Meeting in Santiago de Compostela which I was part of. (02/06/2021)

GCiencia. Recorded interview about Scuba Cancers project for a video published in the scientific newspaper (09/05/2019)

La Voz de Galicia. Interview and picture for the regional newspaper about the Pan-cancer initiative (06/02/2020)

La Gaceta. Interview and picture for Salamanca's newspaper about research in science with cockles (10/12/2017)

Research publications which I co-authored were mentioned in more than 40 different newspapers that can be checked here: https://genomesdisease.tech/media

Social networks and podcasts

VOCES11F. Interview to talk about the **3MT Video Award** (13/06/2022).

NAKED SCIENTIST. Interview to talk about marine contagious cancers (12/04/2022).

ELIFE PODCAST. Brief interview to highlight the findings of our recent publication of a novel contagious cancer among clams on the seas of southern Europe (Episode 79, March 2022)

TWICH. Interview for the show **Ciencia e tal**, a programme that gives insights into the latest scientific research that is being done (31/08/2021)

INSTAGRAM. Interview for #Pintíficas initiative organized by the **Pint of Science** on the International Day of Girls and Women in Science (11/02/2020).

Appendix G: Animal welfare

All animal experiments and the collection included in this doctoral thesis are part of the project Scuba Cancers funded by the European Research Council Starting Grant no. 716290 and therefore, reviewed and approved by the *Standing Committee on Conflict of Interest, Scientific Misconduct and Ethical Issues* (CoIME).

Use of invertebrate mollusc species, such as *Cerastoderma edule, Cerastoderma glaucum, Venus verrucosa, Chamelea gallina* or *Chamelea striulata* included in this doctoral thesis, is exempt from the European Animals Scientific Procedures Directive 2010/63/EU and the Spanish RD/53/2013, on the protection of animals used for scientific purposes, and experimental manipulation. Although cockles are uncovered by these legislations; we followed our institutional guidelines for the use and welfare of laboratory animals and the research team was trained on animal experimentation.

Collection of animals from natural sand beds was carried out after obtaining the permits required by local/national authorities in the countries/locations that were necessary.

Transport was monitored according to European Commission Decision 2003/623/EU, 599/2004/EU and 1251/2008/EU and Spanish RD/542/2016. The Intra Trade Animal Health Certificate (TRACES) was obtained, reference INTRA.NO.2017.0001201-V1.

Maintenance in seawater tanks was carried out in two facilities: *Toralla Marine Science Station,* Universidade de Vigo (ECIMAT s/n, Illa de Toralla, Vigo, Spain; REGA: ES360570181401) and in the *Aquatic Facilities of the Faculty of Biology*, Universidade de Santiago de Compostela (Rúa Constantino Cadeira s/n, Campus Vida, Santiago de Compostela, Spain; REGA: ES150780263301). As we were aware of the potential ecological threat of working with cockles affected by contagious cancers, in terms of environmental protection, international specimens were carefully processed in a biosecurity facility (ISO 9001:2015) to minimise the potential biological risks.

Sacrifice of animals was carried out following the standards and requirements of the European Commission, national governmental agencies, and our institution.

Genetic modification of cockle samples has not been performed although it was initially planned as preliminary findings showed several limitations. Preliminary experiments were perform in the US during a short-term research stay of this doctoral thesis.

Formation in animal welfare and experimentation was attended by the doctoral candidate. In March 2017, she obtained the Animal Experimentation Certificate (140 hours course) from Centro de Estudios Biosanitarios (Madrid, Spain) for the following activities included in the Spanish legislation ECC/566/2015: *care of animals* (function A), *euthanasia of animals* (function B), *performance of procedures* (function C), *design of projects and procedures* (function D) and *supervision of animal welfare* (function E). After completing 190 hours of supervised work, Alicia L. Bruzos was officially capacitated (CAP-1691-18) by Consejería de Medio Ambiente, Admisnistración Local y Ordenación del Territorio of Comunidad de Madrid (Spain) in October 2018.

Certificate of Animal Experimentation (A+B+C+D+E)

Centro de Estudios Biosanitarios

La doctora, Dña. Mónica López Barahona,
Directora General Académica del Centro de Estudios Biosanitarios,

CERTIFICA QUE:

Dª. ALICIA LÓPEZ BRUZOS
con D.N.I./Pasaporte número: 33545624-D

Ha superado los estudios correspondientes a los **Cursos de Experimentación Animal de las Funciones A, B, C, D y E**, según Orden ECC/566/2015, con una duración de 140 horas (70 teóricas y 70 prácticas). La alumna ha iniciado este curso de teleformación el 23 de enero de 2017 y lo ha finalizado el 07 de marzo de 2017.

Esta actividad docente de enseñanza está reconocida por la Comunidad de Madrid para todos los grupos de especies animales incluidas en el Anexo II de la Orden ECC/566/2015.

Fdo.: Mónica López Barahona
Directora General Académica
Centro de Estudios Biosanitarios

Capacitation for Animal Experimentation

CONSEJERÍA DE MEDIO AMBIENTE, ADMINISTRACIÓN LOCAL Y ORDENACIÓN DEL TERRITORIO
Comunidad de Madrid

Certificado de reconocimiento de la capacitación para manejar animales utilizados, criados o suministrados con fines de experimentación y otros fines científicos, incluyendo la docencia. Orden ECC/566/2015, de 20 de marzo

D./Dña. ALICIA LÓPEZ BRUZOS, con DNI/NIE 33545624D ha obtenido el reconocimiento de la capacitación para realizar las funciones de:

DISEÑO DE LOS PROYECTOS Y PROCEDIMIENTOS

ASUNCIÓN DE LA RESPONSABILIDAD DE LA SUPERVISIÓN "IN SITU" DEL BIENESTAR Y CUIDADO DE LOS ANIMALES

en los siguientes grupos de especies animales:

SIN LIMITACIÓN DE ESPECIES

Nº de certificado: CAP-1691-18

ORGANISMO QUE EXPIDE EL CERTIFICADO

Dirección General de Agricultura, Ganadería y Alimentación. Consejería de Medio Ambiente y Ordenación del Territorio. Comunidad de Madrid

El reconocimiento de la capacitación para la realización de las funciones relacionadas en este certificado surtirá efecto en todo el territorio nacional.

Fecha, 8 de octubre de 2018	**Sello**

COMUNIDAD DE MADRID · DIRECCIÓN GENERAL DE AGRICULTURA, GANADERÍA Y ALIMENTACIÓN · CONSEJERÍA DE MEDIO AMBIENTE Y ORDENACIÓN DEL TERRITORIO

EL DIRECTOR GENERAL DE AGRICULTURA, GANADERÍA Y ALIMENTACIÓN

(P.D.F Resolución de 15 de junio de 2018)

EL SUBDIRECTOR GENERAL DE PRODUCCIÓN AGROALIMENTARIA Y BIENESTAR ANIMAL.

Fdo: Jesús Carpintero Hervás

Appendix H: Declarations

Declaration of interests

I, Ms. Alicia L. Bruzos, declare that I have no conflict of interest or any economic, personal, political, financial, or academic relationship that may influence this doctoral thesis.

Image use rights

Permission to reproduce and/or adapt any images and figures for which I, Alicia L. Bruzos, do not own the copyright have been requested. An table containing source, copyright holder and date of permission grating can be found in the following pages.

En Londres, 8 de junio de 2022.

Table 20. Figures summary table with permissions for third party copyright works and own designs.

Type of work	Name of work	Source	Use	Copyright holder	Permission requested	Others
Graphical abstract	Evolution of bivalve transmissible cancers	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 1A	Karyotypes human cancer cell vs. healthy cell	National Human Genome Research Institute, genome.gov	Reproduction	National Human Genome Research Institute, genome.gov	Public domain.	More information: https://www.genome.gov/about-nhgri/Policies-Guidance/Copyright
Figure 1B	Karyotypes human cancer cell vs. healthy cell	French CA. Pathogenesis of NUT midline carcinoma. Annual Review of Pathology: Mechanisms of Disease. 2012;7:247–65.	Adaptation	The Annual Review of Pathology: Mechanisms of Disease	YES. CCC marketplace. May 9th, 2022.	Order License ID: 1219421-1
Figure 2.	Germline and somatic variation in a population	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 3.	Hallmarks of metastasis	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 4A	Cancer is a clonal evolving disease	Nik-Zainal, S. *et al.* (2012) 'The life history of 21 breast cancers', *Cell*, 149(5), pp. 994–1007. doi: 10.1016/j.cell.2012.04.023.	Adaptation	Elsevier Ltd. CC BY 3.0	Not required	More information: https://www.cell.com/fulltext/S0092-8674(12)00527-2
Figure 4B	Cancer is a clonal evolving disease	Rübben, A. and Araujo, A. (2017) 'Cancer heterogeneity: converting a limitation into a source of biologic information', *Journal of Translational Medicine*, 15(190), pp. 1–10. doi: 10.1186/s12967-017-1290-9.	Adaptation	Springer Nature. CC BY 4.0	Not required	More information: https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-017-1290-9#rightslink
Figure 5	Accumulation of driver and passenger somatic mutations	Stratton, M. R. (2013) 'Journeys into the genome of cancer cells', *EMBO Mol Med*, 5, pp. 169–172.	Adaptation	John Wiley and Sons, Ltd on behalf of EMBO, CC BY 3.0	Not required	More information: https://www.embopress.org/doi/full/10.1002/emmm.201202388
Figure 6	Wildlife cancers reported across the tree of life	Aktipis, C. A. *et al.* (2015) 'Cancer across the tree of life: cooperation and cheating in multicellularity', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, p. 20140219. doi: 10.1098/rstb.2014.0219.	Adaptation	Royal Society, CC BY 4.0	Not required	-

Figure 7	Cancer types regarding the scope of their metastasis	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 8	Horizontal spread of a clonally transmissible cancer affecting cockles	Strakova, A. and Murchison, E. P. (2015) 'The cancer which survived: Insights from the genome of an 11000 year-old cancer', *Current Opinion in Genetics and Development*, 30, pp. 49–55. doi: 10.1016/j.gde.2015.03.005.	Adaptation	Elsevier Ltd.	YES. CCC marketplace. May 9th, 2022.	License number: 5304831121638
Figure 9	Sequencing data analysis of contagious cancers	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 10A	Canine Transmissible Venereal Tumour.	Strakova, A. and Murchison, E. P. (2014) 'The changing global distribution and prevalence of canine transmissible venereal tumour', *BMC Veterinary Research*, 10(1), pp. 1–10. doi: 10.1186/s12917-014-0168-9.	Reproduction	BioMed Central Ltd., CC BY 4.0	Not required	-
Figure 10B	Canine Transmissible Venereal Tumour	Strakova, A. and Murchison, E. P. (2015) 'The cancer which survived: Insights from the genome of an 11000 year-old cancer', *Current Opinion in Genetics and Development*, 30, pp. 49–55. doi: 10.1016/j.gde.2015.03.005.	Adaptation	Elsevier Ltd.	YES. CCC marketplace. May 10th, 2022.	License number: 5305301209770
Figure 10C	Canine Transmissible Venereal Tumour	Báez, A. (2021) 'As cancer grows old', *Science*, 374(6571), p. 1066. doi: 10.1126/science.abm8137.	Adaptation	AAAS	YES. Email May 12th, 2022.	-
Figure 11A	Devil Facial Tumour Disease	Stammnitz, M. R. *et al.* (2018) 'The Origins and Vulnerabilities of Two Transmissible Cancers in Tasmanian Devils', *Cancer Cell*, 33(4), pp. 607-619.e15. doi: 10.1016/j.ccell.2018.03.013.	Reproduction	Elsevier Ltd., CC BY 4.0	Not required	-
Figure 11B	Devil Facial Tumour Disease	Pye, R. J. *et al.* (2016) 'A second transmissible cancer in Tasmanian devils', *Proceedings of the National Academy of Sciences*, 113(2), pp. 374–379. doi: 10.1073/pnas.1519691113.	Adaptation	PNAS	YES. Email May 12th, 2022.	-
Figure 12A-E	Human cancer contagions	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 12F	Human cancer contagions	Muehlenbachs, A. *et al.* (2015) 'Malignant Transformation of *Hymenolepis nana* in a Human Host', *New England Journal of Medicine*, 373(19), pp. 1845–1852. doi: 10.1056/NEJMoa1505892.	Reproduction	Massachusetts Medical Society	Not required	More information: https://www.nejm.org/about-nejm/permissions

Figure 13	Worldwide distribution of bivalve contagious cancer lineages	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 14	Bivalve transmissible cancers spread hypothesis	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Courtesy: Jose Tubío	-
Figure 15	Histology of two cockles HN affected by neoplasia A and B	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Courtesy: Seila Díaz	-
Figure 16	Histological comparison of HN and non-cancer tissues of cockles	European Research Council Starting Grant no. 716290 Scuba Cancers	Own work	Universidade de Santiago de Compostela	Not applicable	-
Figure 17	Cytological severity scale for the diagnosis of HN in cockles	European Research Council Starting Grant no. 716290 Scuba Cancers	Own work	Universidade de Santiago de Compostela	Not applicable	-
Figure 18	Micrographs of metaphases of healthy and cancerous cockles	Diaz, S. *et al.* (2013) 'Disseminated neoplasia causes changes in ploidy and apoptosis frequency in cockles Cerastoderma edule', *Journal of Invertebrate Pathology*, 113(3), pp. 214–219. doi: 10.1016/j.jip.2013.03.010.	Adaptation	Elsevier Ltd.	YES. CCC marketplace. May 11th, 2022.	License number: 5305601276671
Figure 19	Analysis of transmissible cancer in the soft-shell clam	Metzger, M. J. *et al.* (2015) 'Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams', *Cell*, 161(2), pp. 255–263. doi: 10.1016/j.cell.2015.02.042.	Adaptation	Elsevier Ltd.	YES. CCC marketplace. May 11th, 2022.	License number: 5305610086534
Figure 20	Dates of some releases of references genomes	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 21	Interspecies transmission scenarios	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 22A-B-C	Overview of defence mechanisms and immune responses in bivalves.	European Research Council Starting Grant no. 716290 Scuba Cancers	Own design and work	Universidade de Santiago de Compostela	Not applicable	-
Figure 22D	Overview of defence mechanisms and immune responses in bivalves.	Chakraborty, S., Ray, M. and Ray, S. (2021) 'Bivalve haemocyte adhesion, aggregation and phagocytosis: A tool to reckon arsenic induced threats to freshwater ecosystem', Fish and Shellfish Immunology, 114(March), pp. 229–237. doi: 10.1016/j.fsi.2021.05.008.	Adaptation	Elsevier Ltd.	YES. CCC marketplace. May 29th, 2022.	License number: 5318221363590

Figure 22E	Overview of defence mechanisms and immune responses in bivalves.	Wootton, E. C., Dyrynda, E. A. and Ratcliffe, N. A. (2006) 'Interaction between non-specific electrostatic forces and humoral factors in haemocyte attachment and encapsulation in the edible cockle, Cerastoderma edule', *Journal of Experimental Biology*, 209(7), pp. 1326–1335. doi: 10.1242/jeb.02118.	Adaptation	Company of Biologists Ltd.	YES. CCC marketplace. May 29th, 2022.	Order Number: 1226623
Figure 23	Histological section of cancer encapsulation in cockles	Díaz, S. et *al*. (2016) 'Long-term epidemiological study of disseminated neoplasia of cockles in Galicia (NW Spain): Temporal patterns at individual and population levels, influence of environmental and cockle-based factors and lethality', Journal of Fish Diseases, pp. 1–16. doi: 10.1111/jfd.12436.	Adaptation	John Wiley and Sons	YES. CCC marketplace. May 29th, 2022.	License number: 5318220936398
Figure 24	Shells of two cockle species	Pictures taken by Olivier Caro.	Reproduction	Olivier Caro	Courtesy: Olivier Caro	Email communication.
Figure 25	Analysis of cockle transmissible cancers	Metzger, M. J. *et al*. (2016) 'Widespread transmission of independent cancer lineages within multiple bivalve species', *Nature*, pp. 1–11. doi: 10.1038/nature18599.	Adaptation	Springer Nature	YES. CCC marketplace. May 11th, 2022.	License number: 5305610563828
Figure 26	Sample collection and processing	This thesis	Own work	Universidade de Santiago de Compostela	Not applicable	-
Figure 27	Schematic representation of the steps performed from diagnosis to read alignment for the three sample types sequence	This thesis	Own work	Universidade de Santiago de Compostela	Not applicable	-
Figure 28	Sequenced samples characteristics	This thesis	Own work	Universidade de Santiago de Compostela	Not applicable	-
Figure 29	Distribution and prevalence of HN	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 30	Species determination of 259 samples	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 31A	Cockle reference genome	This thesis	Own work	Universidade de Santiago de Compostela	Not applicable	-

Figure 31B	Cockle reference genome	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Courtesy: Jorge Zamora	-
Figure 31C-D	Cockle reference genome	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 32	Sequencing dataset of cockle transmissible cancers.	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 33	Sequencing dataset of healthy cockles (PoN)	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 34	Mitogenome alignment, variants and deconvolution	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 35	Maximum-likelihood phylogeny of cockle transmissible cancers based on mitogenomes.	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 36	Structure of mitochondrial clonal lineages	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 37	Microsatellite analysis of cockle transmissible cancers	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 38	Two microsatellite loci across the dataset	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 39	Bayesian phylogeny of cockle transmissible cancers based on mitogenomes	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 40	Geographical distribution of HTs and CedBTN lineages	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 41A	Coinfection of type A and B in a single cockle	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 41B	Coinfection of type A and B in a single cockle	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Courtesy: Seila Díaz	-

Figure 42A	Coinfection of two type A cancer lineages in a single cockle	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 42B	Coinfection of two type A cancer lineages in a single cockle	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Courtesy: Seila Díaz	-
Figure 43	Copy number amplifications on cockle transmissible cancers	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 44	Overall variants called along cockle's mtDNA	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 45	Life cycle and anatomy co common cockles	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 46A	Healthy cockle tissues, organs and larval stages	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 46B-G	Healthy cockle tissues, organs and larval stages	European Research Council Starting Grant no. 716290 Scuba Cancers	Own work	Universidade de Santiago de Compostela	Not applicable	-
Figure 46H-I	Healthy cockle tissues, organs and larval stages	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 47	Transcriptomic profiling.	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 48A	Sampling of clam specimens	Natural History Museum Rotterdam (www.hetnatuurhistorisch.nl)	Reproduction	Natural History Museum Rotterdam, CC-BY SA	Not required	-
Figure 48B	Sampling of clam specimens	Natural History Museum Rotterdam (www.hetnatuurhistorisch.nl)	Reproduction	Natural History Museum Rotterdam, CC-BY SA	Not required	-
Figure 48C	Sampling of clam specimens	Natural History Museum Rotterdam (www.hetnatuurhistorisch.nl)	Reproduction	Natural History Museum Rotterdam, CC-BY SA	Not required	-
Figure 49	Sampling of clam specimens	European Research Council Starting Grant no. 716290 Scuba Cancers	Own work	Universidade de Santiago de Compostela	Not applicable	-

Figure 50	Diagnosis of HN in warty venus clams	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1–20.	Adaptation	eLife. CC BY 4.0	Not required	-
Figure 51	Histological diagnosis of hemic neoplasia in warty venus (V. verrucosa) specimens.	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1–20.	Adaptation	eLife. CC BY 4.0	Not required	-
Figure 52	Transmission electron microscopy (TEM) of healthy and neoplastic V. verrucosa clams	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1–20.	Reproduction	eLife. CC BY 4.0	Not required	-
Figure 53	Ploidy analysis by flow cytometry	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Courtesy: Seila Díaz	-
Figure 54A	Chromosomes of healthy and tumoral cells of warty venus clams	García-Souto, D. *et al.* (2015) 'Divergent evolutionary behavior of H3 histone gene and rDNA clusters in venerid clams', *Molecular Cytogenetics*, 8(1), pp. 1–10. doi: 10.1186/s13039-015-0150-7.	Reproduction	Springer Nature. CC BY 4.0	Not required	-
Figure 54B	Chromosomes of healthy and tumoral cells of warty venus clams	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1–20.	Reproduction	eLife. CC BY 4.0	Not required	-
Figure 55	Mitochondrial Cox1 gene sequencing reveals cancer contagion	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1–20.	Reproduction	eLife. CC BY 4.0	Not required	-

Figure 56	Draft reference mitochondrial DNA (mtDNA) genome assemblies	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1–20.	Reproduction	eLife. CC BY 4.0	Not required	-
Figure 57	Comparison of read coverage	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1–20.	Reproduction	eLife. CC BY 4.0	Not required	-
Figure 58	Molecular phylogeny	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1–20.	Reproduction	eLife. CC BY 4.0	Not required	-
Figure 59	Single-nucleotide variants (tumours, V. verrucosa, C. gallina, and C. striatula)	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1–20.	Reproduction	eLife. CC BY 4.0	Not required	-
Figure 60	Maximum Likelihood molecular phylogenies	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1–20.	Reproduction	eLife. CC BY 4.0	Not required	-
Figure 61	Multispecies coalescent (MSC) tree	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1–20.	Reproduction	eLife. CC BY 4.0	Not required	-
Figure 62	Fluorescence in situ hybridization (FISH) to specifically detect the satellite DNA	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1–20.	Reproduction	eLife. CC BY 4.0	Not required	-

Figure 63	Scenarios of two independent cancer lineages or two subclones of a cancer lineage	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 64	Mitochondria capture versus mtDNA capture.	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 65	Schematic diagram summarising the interspecies contagion	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 66	Interspecies metastases	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 67	Somatic mutations filtration	This thesis	Own design	Universidade de Santiago de Compostela	Not applicable	-
Figure 68A	Chromosome comparison of healthy and cancerous cockles	AboElkhair, M., Siah, A., et al. (2009) 'Reverse transcriptase activity associated with haemic neoplasia in the soft-shell clam Mya arenaria', Diseases of Aquatic Organisms, 84(1), pp. 57–63. doi: 10.3354/dao02038.	Adaptation	Elsevier Ltd.	YES. CCC marketplace. June 5th, 2022.	License number: 5322460439636
Figure 68B	Chromosome comparison of healthy and cancerous cockles	Matias, A. M. *et al.* (2014) 'Karyotype variation in neoplastic cells associated to severity of disseminated neoplasia in the cockle Cerastoderma edule', *Aquaculture*, 428–429, pp. 223–225. doi: 10.1016/j.aquaculture.2014.03.017.	Adaptation	Elsevier Ltd.	YES. CCC marketplace. June 5th, 2022.	License number: 5322460106800
Figure 69	Schematic workflow of sample processing.	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 70	Schematic compendium of rules for the biobank of scuba cancers.	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 71	VAF plots of tumour and matched-normal mitogenomes (part 1/4).	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 72	Full mitogenome RaxML phylogeny of tumours, matched-normal and healthy cockles	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-

Figure 73	Full mitogenome BEAST phylogeny of tumours, matched-normal and healthy cockles	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 74	Full mitogenome MrBayes phylogeny of tumours, matched-normal and healthy cockles	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 75	Full mitogenome RaxML phylogeny of healthy cockles (1000 bootstraps, mid-rooted).	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 76	Full mitogenome Bayesian phylogeny (BEAST inference) of healthy cockles.	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 77	Microsatellite amplifications of tumours and matched-normal tissues	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 78	Copy number (CN) amplifications on cockle transmissible cancers (extension).	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 79	Piecharts of mitochondrial cancer lineages	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 80	Barplot of common variants	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 81	RNA analysis of HN samples	European Research Council Starting Grant no. 716290 Scuba Cancers	Reproduction	Universidade de Santiago de Compostela	Not applicable	-
Figure 82	Histological diagnosis of hemic neoplasia in warty venus (V. verrucosa) specimens.	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1–20.	Adaptation	eLife. CC BY 4.0	Not required	-

Video 1	Mitochondrial genome sequencing of marine leukaemias	Garcia-Souto D, Bruzos AL, Diaz S, Rocha S, Pequeño-Valtierra A, Roman-Lewis CF, et al. Mitochondrial genome sequencing of marine leukaemias reveals cancer contagion between clam species in the Seas of Southern Europe. Elife. 2022;11:1-20.	Adaptation	eLife. CC BY 4.0	Not required	-
Cover 1	Chapter 1 cover	*Scientists Meet Artists* of Campus do Mar https://campusdomar.gal/en/scientists-meet-artists/	Reproduction	Illustration created by Sofía Venzel for Universidade de Vigo (Spain)	YES. Email April 5th, 2022.	-
Cover 2	Chapter 2 cover	This thesis	Reproduction	Doctoral candidate.	YES. Email April 5th, 2022.	People on the picture have given written consent for its use in this thesis.
Cover 3	Chapter 3 cover	This thesis	Reproduction	Doctoral candidate.	Not applicable	-
Cover 4	Chapter 4 cover	European Research Council Starting Grant no. 716290 Scuba Cancers	Adaptation	Scuba Cancers ERC grant (no. 716290) and PixVideos	YES. Email April 5th, 2022.	-
Cover 5	Chapter 5 cover	*Scientists Meet Artists* of Campus do Mar https://campusdomar.gal/en/scientists-meet-artists/	Adaptation	Illustration created by Sofía Venzel for Universidade de Vigo (Spain)	YES. Email April 5th, 2022.	-
Cover 6	Bibliography	This thesis	Reproduction	Doctoral candidate.	Not applicable	-

scuba cancers

¿Y si, amigo mío, la respuesta está enterrada en la arena?

What if, my friend, the answer is buried in the sand?

alicia l bruzos

Cancer cells accumulate mutations that allow them to grow uncontrollably and eventually acquire the ability to metastasize, that is, spread to other parts of the body. Transmissible or contagious cancers are large-scale metastases in which the cancer cells spread to other individuals beyond the body that originated them. This doctoral thesis provides further insights into the evolution of transmissible cancers in bivalves through the inspection of 7,290 cockles and clams and genomic and transcriptomic analyses of 643 bivalves. The findings reported include multiple mitochondrial horizontal transfers, co-infections of two contagious cancer lineages affecting a single individual, histogenesis for two independent cancer lineages and the description of a novel interspecific contagious cancer.
Enjoy the reading!