



**HAL**  
open science

## Prevalence and polymorphism of a mussel transmissible cancer in Europe

Maurine Hammel, Alexis Simon, Christine Arbiol, Antonio Villalba, Erika A.V. Burioli, Jean-françois Pépin, Jean-baptiste Lamy, Abdellah Benabdelmouna, Ismael Bernard, Maryline Houssin, et al.

► **To cite this version:**

Maurine Hammel, Alexis Simon, Christine Arbiol, Antonio Villalba, Erika A.V. Burioli, et al.. Prevalence and polymorphism of a mussel transmissible cancer in Europe. *Molecular Ecology*, 2022, 31 (3), pp.736-751. 10.1111/mec.16052 . hal-03335752

**HAL Id: hal-03335752**

**<https://hal.science/hal-03335752v1>**

Submitted on 6 Sep 2021

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

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

## Prevalence and polymorphism of a mussel transmissible cancer in Europe

Maurine Hammel<sup>(1,2)\*</sup>, Alexis Simon<sup>(1)</sup>, Christine Arbiol<sup>(1)</sup>, Antonio Villalba<sup>(3,4,5)</sup>, Erika AV Burioli<sup>(2,6)</sup>, Jean-François Pépin<sup>(7)</sup>, Jean-Baptiste Lamy<sup>(8)</sup>, Abdellah Benabdelmouna<sup>(8)</sup>, Ismael Bernard<sup>(9)</sup>, Maryline Houssin<sup>(6)</sup>, Guillaume M Charrière<sup>(2)</sup>, Delphine Destoumieux-Garzon<sup>(2)</sup>, John J. Welch<sup>(10)</sup>, Michael J Metzger<sup>(11)</sup>, Nicolas Bierne<sup>(1)</sup>

(1)ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France

(2)IHPE, Univ Montpellier, CNRS, Ifremer, Univ Perpignan, Via Domitia, France

(3)Centro de Investigacións Mariñas, Consellería do Mar, Xunta de Galicia, Vilanova de Arousa, Spain

(4) Departamento de Ciencias de la Vida, Universidad de Alcalá, Alcalá de Henares, Spain.

(5) Research Centre for Experimental Marine Biology and Biotechnology (PIE), University of the Basque Country (UPV/EHU), Plentzia, Basque Country, Spain

(6)LABÉO, Caen, France

(7)Laboratoire Environnement ressources des Pertuis Charentais, IFREMER, La Tremblade, France

(8)Santé, Génétique, Microbiologie des Mollusques, IFREMER, La Tremblade, France

(9)Eurêka Mer, Lézardrieux, France

(10)Department of Genetics, University of Cambridge, Downing Street, Cambridge, UK

(11)Pacific Northwest Research Institute, Seattle, USA

### Abstract

Transmissible cancers are parasitic malignant cell lineages that acquired the ability to infect new hosts from the same species, or sometimes related species. First described in dogs and Tasmanian devils, transmissible cancers were later discovered in some marine bivalves affected by a leukemia-like disease. In *Mytilus* mussels, two lineages of Bivalve Transmissible Neoplasia (BTN) have been described to date (MtrBTN1 and MtrBTN2), both emerged in a *M. trossulus* founder individual. Here, we performed an extensive screening of genetic chimerism, a hallmark of transmissible cancer, by genotyping 106 SNPs of 5907 European *Mytilus* mussels. The genetic analysis allowed us to simultaneously obtain the genotype of hosts - *M. edulis*, *M. galloprovincialis* or hybrids - and the genotype of tumors of heavily infected individuals. In addition, a subset of 222 individuals were systematically genotyped and analysed by histology in order to screen for possible non-transmissible cancers. We detected MtrBTN2 at low prevalence in *M. edulis*, and also in *M. galloprovincialis* and hybrids although at a much lower prevalence. No MtrBTN1 or new BTN were found, but 8 individuals with non-transmissible neoplasia were observed at a single polluted site on the same sampling date. We observed a diversity of MtrBTN2 genotypes that appeared more introgressed or more ancestral than MtrBTN1 and reference healthy *M. trossulus* individuals. The observed polymorphism is most likely due to somatic null alleles caused by structural variations or point mutations in primer-binding sites leading to enhanced detection of the host alleles. Despite low prevalence, two sublineages divergent by 10% fixed somatic null alleles and one non-synonymous mtCOI substitution, are co-spreading in the same geographic area, suggesting a complex diversification of MtrBTN2 since its emergence and host species shift.

**Keywords:** transmissible cancer, disseminated neoplasia, *Mytilus* sp. mussels, genetic chimerism, genetic polymorphism

## 1. Introduction

Transmissible cancers are malignant cell lines, derived from a founder host, which can spread and colonize other individuals by transmission of living cancer cells. Several lineages of transmissible cancer have been described: one in dogs (Canine Transmissible Venereal Tumor, CTVT, Murgia et al. 2006, Rebbeck et al. 2009), two in Tasmanian devils (Devil Facial Tumor Disease, DFT1 and 2, Pearse et al. 2006, Pye et al. 2016) and, more recently, seven in different bivalve species (Bivalve Transmissible Neoplasia, BTN, Metzger et al. 2015, 2016, Yonemitsu et al. 2019, Garcia-Souto et al. 2021). The three lineages found in vertebrates are the best studied to date. In dogs, CTVT arose about 4000-8000 ybp and spread worldwide (Leathlobhair et al. 2018, Baez-Ortega et al. 2019). In Tasmanian devils two independent transmissible cancer lineages have emerged in the last 30 years. In both cases, although at different spatio-temporal scales, genome studies have shown early diversification into sublineages and secondary co-existence of some divergent sublineages (Baez-Ortega et al. 2019, Kwon et al. 2020; Patton et al. 2020). Studies in marine bivalves also showed multiple emergences (two in *Cerastoderma edule*, Metzger et al. 2016 and two in *Mytilus trossulus*, Yonemitsu et al. 2019) and for the first time in transmissible cancer research, transmissions crossing the species barrier (BTN lineage in *Politiapes aureus* has a *Venerupis corrugata* origin, Metzger et al. 2016, MtrBTN2 in *M. edulis* and *M. chilensis* has a *M. trossulus* origin, Yonemitsu et al. 2019, and BTN lineage in *Venus verrucosa* has a *Chamelea gallina* origin, Garcia-Souto et al. 2021).

The *Mytilus edulis* complex of species is composed of three species in the northern hemisphere: *M. edulis*, *M. galloprovincialis* and *M. trossulus*, and three species in the southern hemisphere: *M. chilensis*, *M. platensis* and *M. planulatus*. They are incompletely reproductively isolated and can hybridize when they come into contact, either naturally or via human-induced introductions (Fraisse et al. 2016, Simon et al. 2020, Popovic et al. 2020). Disseminated neoplasia (DN) has been previously observed in four of these species around the world: *M. trossulus* (Rasmussen 1986, Sunila 1987, Moore et al. 1991, Elston et al. 1992, Usheva and Frolova 2000, Vassilenko and Baldwin 2014), *M. chilensis* (Campalans et al. 1998, Cremonte et al. 2015, Lohrmann et al. 2019), *M. edulis* (Farley 1969, Lowe and Moore 1978, Green and Alderman 1983, Elston et al. 1988, Galimany and Sunila 2008) and *M. galloprovincialis* (Figueras et al. 1991, Villalba et al. 1997, Carrasco et al. 2008, Gombac et al. 2013, Carella et al. 2013, Matozzo et al. 2018). Cancer cells are characterised by rounded, basophilic and polyploid cells with a high nucleus-to-cell ratio (for a review see Carballal et al. 2015) and their presence is determined by histological observation, and cytology or flow cytometry of hemolymph (Benabdelmouna and Ledu 2016, Benabdelmouna et al. 2018, Burioli et al. 2019). However, in order to demonstrate that a cancer is transmissible, a genetic study is required in order to identify (i) genetic differences between tumor and host cells and (ii) genetic similarity between tumors of different individuals (the hallmark of transmissible cancer). Thus, two independent lineages of BTN from two *M. trossulus* founder mussels have been described in three mussel species: one lineage (MtrBTN1) found in East Pacific *M. trossulus* populations, and the other lineage (MtrBTN2) with a worldwide distribution in West Pacific *M. trossulus* populations, European *M. edulis* and South American *M. chilensis* hosts (Metzger et al. 2016, Yonemitsu et al. 2019, Skazina et al. 2021). MtrBTN1 was discovered in *M. trossulus* populations in Western Canada by genetic similarity of neoplastic hemocytes at one mitochondrial and one nuclear locus (Metzger et al. 2016). Although disseminated neoplasia was first

described 50 years ago in *Mytilus* sp. (Farley 1969) we do not know if cases of disseminated neoplasia reported before Metzger et al. (2016) correspond to transmissible cancers. A population genetic study of French mussels detected chimeric individuals with an abnormal presence of *M. trossulus* alleles in *M. edulis* populations devoid of *M. trossulus* individuals, leading the authors to hypothesize the presence of a *M. trossulus* transmissible cancer (Riquet et al. 2017). The presence of *M. trossulus* transmissible cancer in *M. edulis* was later confirmed by the detection of a *M. trossulus* genetic signal in the hemolymph of cancerous *M. edulis* grown in Brittany (Burioli et al. 2019). By comparing two nuclear and two mitochondrial loci from MtrBTN1 tumors, and tumors found in South-American *M. chilensis* and European *M. edulis*, Yonetmistu et al. (2019) showed the latter are an independent emergence, called MtrBTN2. One mitochondrial marker indicated that MtrBTN2 from two *M. chilensis* mussels have the *M. chilensis*-derived haplotype indicating that recombination between host and cancer mitochondria might occur in this cancer lineage (Yonetmistu et al. 2019). Very recently, MtrBTN2 has been observed in *M. trossulus* hosts in the Sea of Japan (Skazina et al. 2021) with divergent mtDNA sequences and a lower ploidy (~4N) than the hyperploids MtrBTN2 found in *M. edulis* (>8N, Burioli et al. 2019).

Here, we performed a large genetic screening of European mussels (*M. edulis*, *M. galloprovincialis* and their hybrids) in order to detect genetic evidence of transmissible cancer (chimerism) of any origin and compare the prevalence between species. For this study we performed a dedicated sampling and also reanalysed data from previous population genetics studies. We finally inferred the genotypes of tumors of the most cancerous individuals and studied polymorphism of these tumors. Most candidate chimeric individuals were subsequently observed in histology, cytology or flow cytometry in order to obtain a direct validation that the chimerism was the consequence of a transmissible cancer. Because disseminated neoplasia could also be nontransmissible, a subsample of 222 mussels was systematically inspected by histology for the presence of neoplasia before genetic screening. MtrBTN2 was found at a low prevalence in Europe, mostly in *M. edulis* and also in *M. galloprovincialis* and their hybrids. Some neoplastic mussels did not show evidence of genetic chimerism suggesting that conventional nontransmissible cancers were also present. Finally, the study of MtrBTN2 genotypes from the 6 most infected *M. edulis* hosts revealed polymorphism and allows us to describe two divergent sublineages that are co-spreading in French populations.

## **2. Material and methods**

### **a. Sampling and genotyping**

We analysed a total of 5907 European mussels (*Mytilus edulis*, *M. galloprovincialis*, and some hybrids, subsequently classified with SNPs, see below), collected between 2005 and 2019 (Figure 1, Table S1). For this study  $n = 1749$  mussels were sampled on two tissues: hemolymph (higher amount of cancer cells, increasing with cancer severity) and mantle (higher amount of host cells, decreasing with cancer severity). We also analysed  $n = 4158$  mussels sampled for population genetics studies at only one tissue (gill or hemolymph). For comparison, we also analysed two known MtrBTN1 samples and their *M. trossulus* hosts (described in Metzger et al. 2016), and 32 healthy *M. trossulus* (10 from British Columbia and 10 from Saint-Laurent, Canada and 8 from the Baltic Sea, Gdansk, Poland) (Table S1).

Solid tissues (mantle and gill) were either fixed in 96% ethanol or directly used for DNA extraction. For hemolymph samples, collected from adductor muscle (1ml syringe, 26G needle), cells were centrifuged (10 min at 2500g) and the supernatant

removed and replaced by 96% ethanol. DNA extraction was done with the NucleoMag™ 96 Tissue kit (Macherey-Nagel) in combination with a Kingfisher Flex™ extraction robot (serial number 711-920, ThermoFisher Scientific). We followed the kit protocol with modified volumes of some reagents: beads 2x diluted in water, 200µL of MB3 and MB4, 300µL of MB5 and 100µL of MB6.

A total of 106 biallelic single nucleotide polymorphisms (SNPs) described in Simon et al. 2020 were analysed. However 32 SNPs produced too much missing data and unsatisfactory fluorescence level and were removed for subsequent analyses. The list of 74 nuclear SNPs used and their sequences are provided in Table S2. They were identified to be strongly differentiated between species and lineages of the *Mytilus edulis* complex of species in the study of Fraisse et al. (2016), in order to develop an ancestry-informative panel of SNPs, and its efficacy was confirmed by the analysis of large datasets (Simon et al. 2020, 2021). Genotyping was subcontracted to LGC genomics (Hoddensdon, UK) and performed with the KASP™ method (Cuppen, 2007, Semagn et al. 2014). This method puts two “allele specific” primers with two different fluorochromes in competition during a PCR amplification and gives two values of fluorescence level, one for allele X and the other for allele Y. Homozygous genotypes have a high fluorescence for allele X or allele Y only, while heterozygous genotypes have an intermediate fluorescence for each allele (Figure 2). The genotype classification at each SNP of each sample was determined by two LGC genomics experts. Assays with too low fluorescence were classified as badly amplified (“bad”) and those with an unbalanced intermediate fluorescence, that were compatible with neither homozygous nor heterozygous genotypes of a standard diploid genotype, were classified as “uncalled” (Figure 2). In order to assign mussels to species (*M. edulis*, *M. galloprovincialis* or hybrid), we performed a Principal Component Analysis (PCA, *dudi.pca()*, package *ade4* in R, Jombart et al. 2011) based on the 74 nuclear SNPs and used the first axis coordinate to infer a hybrid index (Väinölä and Strelkov 2011; Brisbin et al. 2012).

### ***b. Transforming KASP genotyping data for transmissible cancer genetic diagnostics***

In contrast to healthy mussels, mussels affected by transmissible cancer are chimeric, with their infected tissues composed of a mix of neoplastic and host cells, usually with a proportion deviating from 50:50 (a minority of cancerous cells mainly in the hemolymph at the earlier stages of the disease, and a majority of cancerous cells in every tissue at the later stages but usually with a greater proportion of cancer cells in the hemolymph). As a consequence, the KASP™ amplification is likely to result in an unbalanced amplification of the two alleles at SNPs for which the genotype is different between cancerous and host cells. Such deviations from standard diploid genotype callings are likely to result in the genotype being classified as “uncalled” (Figure 2). Consequently, we chose to use allele fluorescence values directly rather than genotype calling. To be able to compare samples from independent genotyping experiments, we first had to standardize the X and Y fluorescence values to have all of them in a range from 0 to 1 by using the following formulas within each experimental plate:  $x_{i,j} = (X_{i,j} - \min(X)) / (\max(X) - \min(X))$  and  $y_{i,j} = (Y_{i,j} - \min(Y)) / (\max(Y) - \min(Y))$ , where  $X_{i,j}$  and  $Y_{i,j}$  are the allele X and allele Y fluorescence values respectively for the  $i$ th SNP of  $j$ th sample. Then, following the method of Cuenca et al. (2013) for genotyping polyploid samples, we transformed the data to obtain a measure of the proportion of allele Y fluorescence at each SNP, with a value ranging from 0 when allele X ( $x_{i,j}$ ) fluorescence dominates, to 1 when allele Y ( $y_{i,j}$ ) fluorescence dominates (Figure 2).

Furthermore, SNP alleles were oriented in order for allele Y to be defined as the most frequent in *M. trossulus* and allele X the most frequent in compound *M. edulis* and *M. galloprovincialis* reference samples (see Table S2). To do this, we used allele dosage measured by the following formula  $y'_{ij} = y_{ij}/(x_{ij} + y_{ij})$  where  $y_{ij}$  and  $x_{ij}$  are the fluorescence values of allele X and Y respectively for the  $i$ th SNP of  $j$ th sample (Figure 2). We called  $y'_{ij}$  the “Variant Allele Fluorescence Fraction (VAFF)” (analogously to the Variant Allele Fraction statistics used with NGS dataset, but with the fluorescence signal used instead of the read coverage).

To filter the data out from badly amplified low fluorescence level SNPs and samples, we calculated an index of amplification following this formula  $a_{ij} = X_{ij} + Y_{ij}$ , where  $X_{ij}$  and  $Y_{ij}$  are the allele X and allele Y original fluorescence values. Fluorescence levels were removed from the dataset when  $a_{ij}$  was below 0.3. Samples with an average index of amplification below 0.8 were also removed.

### **c. Genetic diagnosis: 3 indexes to detect chimerism**

We developed two indexes to identify chimeric mussels (a hallmark of transmissible cancer), and a third one designed to specifically identify chimera with an excess of *M. trossulus* allele fluorescence (Figure 2).

When two tissues of the same individual were sampled we were able to investigate genetic differences between the two tissue samples within individuals. We define the two-tissue genetic distance (2TD) as the Euclidean distance between the hemolymph and the mantle based on allele fluorescences  $x_i$  and  $y_i$  ( $dist()$ , package *stats* in R, R Core Team 2020) (Figure 2). Chimerism was considered detected when 2TD was above 1.54 (highest percentile, see results). We expected a higher proportion of cancer cells in the hemolymph than in hard tissues, at least in the early stage of the disease. However, we noticed that for more infected individuals, chimerism was observed in both tissues with sometimes little difference in fluorescence rates between hemolymph and mantle. We therefore developed a second index to identify fluorescence rates that deviate from a balanced amplification of the two alleles. The proportion of unbalanced allelic fluorescence (PUF) was computed as the proportion of SNPs classified as “uncalled” in an individual tissue (classification performed by two LGC experts) (Figure 2). For individuals genotyped at two tissues we obtained two values of PUFs, one for each tissue. Chimerism was considered detected when PUF was above 0.20 (highest percentile, see results), corresponding to a high number of “uncalled” SNPs in a sample. Because only MtrBTN2 has been found in Europe to date (Riquet et al. 2017, Yonemitsu et al. 2019, Burioli et al. 2019) and as MtrBTN2 originated in a *M. trossulus* host, we computed a “*M. trossulus* fluorescence fraction” index (MtFF) to specifically detect *M. trossulus* chimerism (Figure 2). We selected 10 SNPs of the array that are diagnostic between *M. trossulus* and both *M. edulis* and *M. galloprovincialis* (allele Y fixed or nearly fixed in *M. trossulus* populations and absent or very rare in *M. edulis* and *M. galloprovincialis* populations, see Table S2). We computed the average VAFF ( $y'_{ij}$ ) at these 10 SNPs in order to detect samples with abnormal excess of *M. trossulus* fluorescence; recall that *M. trossulus* is absent from the areas investigated in this study. As for PUFs, we obtained two values of MtFF for individuals genotyped at two tissues. We considered the *M. trossulus* fluorescence signal as positive when MtFF was above 0.22 (highest percentile, see results).

Because apparent chimerism can result from experimental artifacts rather than transmissible cancer, we inferred a clustering tree to determine if positive individuals grouped together, as expected for a clonally propagating transmissible cancer, but not

for independent contaminations. We used the Neighbor-Joining method ( $nj()$ , package *ape* in R, Paradis and Schliep 2018) based on Euclidean distance of VAFF value ( $y'_{ij}$ ).

#### **d. Method evaluation: analysis of in silico chimera**

We evaluated our ability to detect non-*trossulus* chimerism, and conversely our ability to conclude that neoplasia are non-transmissible, using the 2TD and PUF statistics, by computing the expected distribution of in-silico generated chimera with an increasing fraction of cancer cells in the sample. For the 2TD statistic we first had to estimate the allele fluorescence of chimera (“chim”). We combined the fluorescence values of two randomly selected hemolymph samples (“host” and “cancer”) following this formula:  $x_{i,chim} = (1-F)*x_{i,host} + F*x_{i,cancer}$  and  $y_{i,chim} = (1-F)*y_{i,host} + F*y_{i,cancer}$ , where  $x_i$  and  $y_i$  are the allele X and allele Y fluorescence values respectively for the  $i$ th SNP of the “host” and “cancer” samples and F the fraction of cancer cells. Then, we calculated the 2TD as the Euclidean distance between the computed “chimera” hemolymph sample and the “host” mantle sample (as described previously). For the PUF statistic we classified as “uncalled” each SNP for which the “host” and “cancer” hemolymph samples had different genotypes called, and calculated the proportion of “uncalled” SNPs for each chimera (as described previously). Because transmissible cancer can be either intraspecific or interspecific, we generated different categories of *in silico* chimera by selecting “host” and “cancer” samples from the same or different species (*M. edulis* - *M. edulis*, *M. edulis* - *M. galloprovincialis*, *M. edulis* - *M. trossulus*, *M. galloprovincialis* - *M. galloprovincialis* and *M. galloprovincialis* - *M. trossulus*).

#### **e. Additional diagnosis methods: histology, cytology and flow cytometry**

A total of  $n = 222$  mussels sampled in two tissues were screened for neoplasia using histology (Figure 2, Table S3) before genetic diagnosis. Shells were opened by cutting the adductor muscle and the soft tissues were removed. An approximately 5 mm thick transverse section of mussel tissue containing mantle lobes, visceral mass (gut, digestive gland) and gills was excised, fixed in Davidson’s fixative for 48h and then preserved in 70% ethanol. Fixed tissues were dehydrated through an ascending ethanol series and embedded in paraffin wax. Five  $\mu$ m thick sections were produced with a rotary microtome and stained with Harris’ hematoxylin and eosin (HHE) (Howard et al. 2004). Histological sections were examined with a light microscope Olympus BX50. Following a semi-quantitative scale previously described (Mix 1983, Diaz et al. 2016), neoplastic mussels were classified into three severity stages of the disease. These severity stages were “light”: a few isolated neoplastic cells in vessels or a single organ; “moderate”: a higher number of neoplastic cells in multiple organs; “heavy”: very large number of neoplastic cells in most or all organs.

In order to validate genetic diagnosis,  $n = 17$  mussels were investigated by histology ( $n = 11$ ), cytology ( $n = 5$ ) and/or flow cytometry ( $n = 6$ ) after the genetic screening. The methods used are referenced in Table S3.

#### **f. real-time PCR to validate MtFF genetic diagnosis**

In order to validate the MtFF genetic diagnosis of  $n = 18$  mussels (see results), we performed the MtrBTN2 specific real-time PCR (Elongation factor locus, EF1 $\alpha$ -H) described in Yonemitsu et al. (2019). Genomic DNA was amplified with sensiFAST SyBR No-ROX Kit (Bioline) using the primer pair design by Yonemitsu et al. (2019) (forward primer: 5'-GCAAAAGTGGCTGAAAACAGATTCTA, reverse primer: 5'-GTAAAAAAGTTAAAATTTCTTTTAGTCACACAAT, amplicon length: 180bp). Real-

time PCR was performed using LightCycler® 480 Real-Time PCR System on 3 cycling stages (95°C for 15s, 50°C for 15s and 60°C for 30s, respectively) for 40 cycles.

### **g. Inference of cancer genotypes and their genetic relationship**

Determining the genotype of a cancer lineage is complicated by the presence of host cells. However, 6 samples exhibited a MtFF close to our *M. trossulus* reference samples consistent with the majority of their cells being cancerous (55% < MtFF < 90%). We used these 6 samples to infer the genotypes of tumors. In order to infer genotypes, we used the x/y fluorescence plots as well as the correlation between fluorescence rate at individual SNP and MtFF (Figure S1). As we were working with *M. edulis* hosts only (see results) we used a new *M. trossulus* fluorescence fraction (MtFF2) with an extended number of 17 *M. edulis* - *M. trossulus* diagnostic SNPs (Table S2). MtFF2 was used as a proxy of the proportion of cancerous cells in the sample (see Figure S1). Genotype inference was performed independently by two users and cross-checked.

To have a global comparison of cancerous genotypes with *Mytilus* references, we performed a maximum parsimony (MP) tree using MEGA v7.0.26 (Kumar, Stecher and Tamura 2016) with all sites and Subtree-Pruning-Regrafting method. SNP genotypes were coded with two nucleotides per position (a single SNP was represented by two positions in a “pseudo-fasta” file) to be able to represent heterozygous and homozygous genotypes and then we combined all SNPs in a single “pseudo-sequence” for each sample before converting to fasta format (*as.fasta()* package *bio3d* v2.4-1 in R, Grant et al. 2006).

### **h. mtCOI sequences**

Individuals STBRI\_69, L19\_8, L9\_3, L11\_4 and CamaretB\_1 affected by MtrBTN2, were sequenced for a parallel research project and short reads produced by Illumina sequencing were available (Novaseq, paired-end 150 bp). We skimmed mitochondrial reads by mapping them to a set of eight *Mytilus* mitogenomes, using bwa-mem2 (v2.0, Md et al. 2019) and SAMtools (v1.10, Li et al. 2009) with a minimal mapping quality of 5. The mitochondrial genomes from female and male *M. galloprovincialis* (FJ890849 and FJ890850, respectively), *M. edulis* (DQ198231 and AY823623, respectively), *M. trossulus* (AY823625 and HM462081, respectively), and *M. californianus* (JX486124 and JX486123, respectively) were retrieved from GenBank and used as references. Note that the *M. edulis* female mitochondrion is attributed to a *M. trossulus* individual in GenBank but corresponds to an introgression of the *M. edulis* mitochondrion in the Baltic Sea. For each cancerous individual, mitochondrial reads were then assembled with MEGAHIT (v1.2.9, Li et al. 2016). Contigs produced with megahit were finally processed with MitoFinder (Allio et al. 2020) to automatically retrieve contigs corresponding to mitochondria and annotate them (using the invertebrate mitochondrial code as organism type). Then, we extracted COI sequences from all mitochondrial contigs for each individual, and aligned them with sequences of Yonemitsu et al. (2019), Skazina et al. (2021) as well as *M. trossulus* mtCOI sequences gathered from Genbank (Breton et al. 2006, Lesser et al. 2010, Marko et al. 2010, Layton et al. 2014, Zbawicka et al. 2014, Crego-Prieto et al. 2015, Metzger et al. 2016, Laakkonen et al. 2021, see Table S4). We also extracted COI sequences from the mitogenome of Mtx10A, a 100,000 year old *M. trossulus* mussel (Der Sarkissian et al. 2020, European Nucleotides Archive SAMEA6367443). Finally, we performed a maximum parsimony (MP) tree using MEGA v7.0.26 (Kumar, Stecher and Tamura 2016) with all sites and Subtree-Pruning-Regrafting method.



### 3. Results

#### a. Transmissible cancer genetic diagnosis of individuals sampled on two tissues

We first analysed samples of two tissues from  $n = 1749$  mussels in order to evaluate the prevalence of genetic chimerism in this dataset and the ability of our indexes to identify mussels infected by a transmissible cancer either of *M. trossulus* origin or other *Mytilus* spp. origins.

We found 11 positive MtFF outlier individuals (*M. trossulus* fluorescence fraction), 5 for both tissues and 6 for hemolymph only (Figure 3A, Table S5). Ten of them were also positive at both PUF (proportion of unbalanced allelic fluorescence, 5 for both tissues and 5 for hemolymph only) and 2TD (2-tissue genetic distance), and the last one was positive at PUF but not 2TD (Figure S2, Table S5). MtFF values were higher in the hemolymph samples than the mantle samples but the difference was sometimes low (Table S5). Those 11 individuals are probably infected by the MtrBTN2 lineage previously described in these populations (Yonemitsu et al. 2019). Ten of these 11 mussels were also investigated by at least one of the additional diagnostic methods (histology, cytology and/or flow cytometry), which confirmed the presence of cancer cells for all of them (Figure 3A, red triangle up, Table S3). Moreover, the highest MtFF values were found in heavily infected individuals as also reported by Burioli et al. (2019) (Table S3). The individual that was negative at 2TD (L11\_4) was equally infected in both tissues. Finally, we used the MtrBTN2 specific real-time PCR diagnosis of Yonemitsu et al. (2019) and found all 11 samples to be positive (Table S6).

We also found 12 individuals that were negative at MtFF but positive at both PUF and 2TD (Figure 3A, Figure S2, see green point, Table S5). Those individuals could be candidates for a new transmissible cancer that emerged in *M. edulis* or *M. galloprovincialis*. We obtained additional diagnostic information (histology, cytology and/or flow cytometry) for 9 of these 12 mussels and found no evidence of the presence of cancer cells for 8 of them, suggesting the chimerism is not related to transmissible cancer, but probably to experimental artifacts such as contamination (Table S3). Only one individual (CamaretA\_61) revealed the possible presence of a few cancer cells (Figure 3A, green triangle up, Table S3). Although the hemolymph MtFF of this individual was not high enough to be classed as an outlier, the hemolymph MtFF was nonetheless high, and close to the detection limit (Figure 3A, green triangle up). This suggests that it could be the *M. trossulus* cancer found in other mussels, likely MtrBTN2, at a very early stage of the disease.

To further investigate the origin of non-*trossulus* chimerism we analysed the genetic similarity of samples with a cluster tree reconstruction based on VAFF Euclidean distances between individuals. All *M. trossulus* chimeric samples grouped together (Figure S3, red points) in a cluster related to *M. trossulus* reference samples. Conversely, non-*trossulus* chimeric samples were sparsely distributed across the tree. Two exceptions, CamaretA\_61 and AUN\_001\_13, were the two samples with the hemolymph MtFF close to the detection limit (and with few cancer cells for CamaretA\_61), which cluster with the *M. trossulus* chimeric group. This suggests once again that they represent the early stages of the MtrBTN2 disease. Although two other groups (of 4 and 2 samples, respectively) tended to cluster together, they were respectively sampled in a population with sympatric *M. edulis* and *M. galloprovincialis*, and a contamination can easily resemble a hybrid genotype. Note that the 3 non-*trossulus* chimeric individuals (AUN\_001\_13, Plg3 and L1\_29) for which no additional method diagnosis information was available do not cluster together in the tree (Figure S3). However, lack of clustering among the non-*trossulus* chimeric samples should not

be overinterpreted, because the fluorescence signal of non-*trossulus* BTN cells might be insufficient to dominate the signal of host cells. In that case, however, PUF and 2TD statistics would not be expected to be outlying. Our analysis of in-silico chimera with variable proportions of mixing between two individuals (Figure S4), allows us to estimate the level of infection needed to detect a BTN (over 80% for a *M. edulis* BTN in *M. edulis* host, over 50% for a *M. galloprovincialis* BTN in *M. galloprovincialis* host and over 30% for an interspecific spreading). With such proportions of cancer cells, chimeric samples would have tended to cluster together, as illustrated by examples of in-silico chimera with mixtures of 40, 60 and 80% added in the tree (see Figure S3, orange branches). Above all, we would have observed neoplastic cells by histological examination. Taking this evidence together, we conclude that non-*trossulus* chimerism was most likely due to technical artifacts.

### **b. screening of MtrBTN2 in individuals sampled on one tissue only**

Having shown that the *M. trossulus* fluorescence fraction (MtFF) provided a satisfactory diagnostic for MtrBTN2, even with mantle samples, and having found very little evidence for other transmissible cancers, we decided to analyse previously obtained genotyping data from DNA extraction of gills ( $n = 3988$ ) and hemolymph ( $n = 197$ ). This allowed us to substantially increase our sample size from 1749 to 5907 mussels. The MtFF index allowed us to identify 15 individuals with an excess of *M. trossulus* fluorescence (Figure 3B, red point, Table S5). Chimerism was confirmed by a high PUF signal for 12 of them (Figure S2, Table S5). The 3 others had a very low PUF and correspond to MtFF values close to the detection limit. We considered as candidates only the 12 positives for both MtFF and PUF. In addition, we had additional diagnostic information for 4 of them which all proved to be infected by neoplastic cells (Figure 3B, red triangle up, Table S3). Finally, for 6 of those samples we performed the MtrBTN2 specific real-time PCR diagnosis of Yonemitsu et al. (2019) and found all 6 to be positive (Table S6).

### **c. Host-specific prevalence of MtrBTN2 in Europe**

We classified individual mussels as *M. edulis*, *M. galloprovincialis* or hybrid using a PCA on all SNP VAFF values ( $y_{ij}$ ). The first axis of the PCA separates *M. edulis* and *M. galloprovincialis* individuals, with hybrids in-between, while the second axis separates these two species and *M. trossulus* (Figure S5). We used the first axis coordinate to infer a hybrid index (HI: Väinölä and Strelkov 2011; Brisbin et al. 2012) and assign mussels to *M. edulis* ( $HI > 0.85$ ), *M. galloprovincialis* ( $HI < 0.4$ ), and hybrids ( $0.4 < HI < 0.85$ ). These values were chosen because asymmetric introgression generates a larger variance of *M. edulis* ancestry in *M. galloprovincialis* than vice versa. The overall prevalence of MtrBTN2 was low ( $25/5907 = 0.42\%$ ) but was significantly higher in *M. edulis* hosts ( $22/2808 = 0.78\%$ ) than in hybrids and *M. galloprovincialis* hosts ( $1/381 = 0.26\%$  and  $2/2718 = 0.07\%$ , respectively; Fisher's exact test  $p = 0.0002$ ).

### **d. Non-transmissible cancers**

For 222 individuals we conducted histological inspection before genetic diagnosis. One individual, L9\_3, was highly infected by large neoplastic cells referred to as type A neoplasia (see micrographs A and B in Figure 4 and description below). This individual was positive at the three genetic diagnostics indexes (MtFF, PUF and 2TD) and was infected by MtrBTN2. Eight individuals from batch L6 were affected by a type of neoplasia characterised by neoplastic cells with smaller size than the type A and

referred as type B neoplasia (see micrographs C and D in Figure 4). Both disseminated neoplasia type A and type B involved the proliferation of abnormally large cells through the connective tissue and hemolymph vessels of different organs. When examining histological sections, the main morphological character for discrimination between neoplasia types was the nuclear size, which was significantly larger in type A (mean diameter  $\pm$  SE =  $7.83 \pm 0.136 \mu\text{m}$ ; range: 6.0 – 9.9  $\mu\text{m}$ ; N = 31) than in type B (mean diameter  $\pm$  SE =  $4.56 \pm 0.096 \mu\text{m}$ ; range: 3.6 – 5.1  $\mu\text{m}$ ; N = 40); both cell types showed scant cytoplasm (high N/C ratio) and mitotic figures were more abundant in type A. Genetic analysis revealed no evidence of chimerism for these individuals with neoplasia type B (low values of PUF and 2TD, Figure S2, Table S5). Note that the highest PUF and 2TD values corresponded to two different samples (L6\_49 for PUF and L6\_46 for 2TD). However, our analysis of in-silico chimera with variable proportions of mixing between two individuals (Figure S4) suggested that a BTN from an *M. edulis* founder infecting an *M. edulis* host can be missed if the neoplastic cells do not outnumber normal cells. It is therefore possible that the proportion of neoplastic cells could have been too low in each individual. We used our in-silico chimera to test the proportion of neoplastic cells needed to detect chimerism with 2TD when we combine the eight samples with a Fisher's combined probability test. More than 35% neoplastic cells in each sample would be needed to obtain a significant test. Given that almost all mussels infected were in the moderate or heavy stages of the diseases (see Table S3), the most parsimonious interpretation is that type B disseminated neoplasia are not transmissible.

#### **e. Genotypic differences and polymorphism of MtrBTN2 tumors**

Cancerous samples were a mix of host and tumor genotypes and when the proportion of cancer cells was too low we could not infer the tumor genotype accurately. Hence, we inferred the tumor genotype of 6 cancerous individuals with the highest MtFF2 values (more than 0.75) and composed of a vast majority of cancer cells (up to 90%, Figure S1 and Table S7). Most of the SNPs were associated with *M. trossulus*-state alleles (Figure 5C, Figure S6, see “fixed between lineages”), as expected for an emergence in a *M. trossulus* host.

Using the 74 nuclear SNPs, we then reconstructed a MP-tree with *M. edulis* and *M. trossulus* healthy individuals, 2 MtrBTN1 and 6 MtrBTN2 (Figure 5A). As expected, *M. edulis* and *M. trossulus* form two genetically distinct clusters. MtrBTN2 samples group together and differ from MtrBTN1, confirming the existence of two distinct lineages (Yonemitsu et al. 2019). The two *Mytilus* BTN lineages are related to the *M. trossulus* cluster also confirming the *M. trossulus* origin of both.

Even though MtrBTN2 tumors group together, many genetic differences (polymorphisms) were observed among the 6 tumors. A closer look at the subtree with only the 6 tumors (Figure 5B) shows two groups of tumors: L11\_4, Pal30 and STBR169 (sublineage a) on one side and L19\_8, L9\_3 and CamaretB\_1 (sublineage b) on the other side, separated by as many as 8 mutations (10% of SNPs, Figure 5C, see polymorphic between sublineages). Sublineage a is homozygous for 7 of these 8 mutations (mostly *M. trossulus*-state alleles) while sublineage b appears heterozygous. The 12 mutations found in each sublineage can be oriented by parsimony. Surprisingly, 10 of those 12 mutations (see polymorphic SNPs on the right of Figure 5C) are gains of heterozygosity (GOH, Figure 5B, black arrow) while only 2 are losses of heterozygosity (LOH, Figure 5B, colored arrow). However, the 8 mutations separating the two sublineages cannot be oriented without rooting the tree, which is not possible given that the host is sexually reproducing.

We can however work on alternative hypotheses given that the SNPs are diagnostics between *M. trossulus* and *M. edulis*. Under a “non-hybrid founder” hypothesis (Figure 5B, see H1) the ancestor would have been homozygous for *M. trossulus*-state alleles and we could infer excess gain of heterozygosity (GOH) in sublineage b. Under a “hybrid founder” hypothesis (Figure 5B, see H2) the ancestor would have been heterozygous and we could infer an excess loss of heterozygosity (LOH) toward the *M. trossulus*-state in sublineage a.

The mtCOI tree (Figure 6) mostly confirms the nuclear tree and sublineages with the exception of the L9\_3 mtCOI sequence, which appears divergent to every other MtrBTN2 sequence sampled in *M. edulis* or *M. chilensis* while this tumor is related to L19\_8 and CamaretB\_1 in the nuclear tree. Together with the MtrBTN2 sequences obtained from *M. trossulus* (Skazina *et al.* 2021), the L9\_3 mtCOI sequence suggests 62mc10 mitochondria came from a MtrBTN2 tumor, as extensively discussed by Skazina *et al.* (2021). Briefly, this sequence has been obtained after an extensive screening of a *M. trossulus* mitochondria in Baltic mussels known to be fixed for introgressed *M. edulis* mitochondria, and 62mc10 was found heteroplasmic for two F mitochondria (Śmietanka and Burzyński 2017). The Bf-48 haplotype that clusters with J54-1 and J161-1, as well as the SA1 and SA5 haplotypes that clusters with J111-2 and J181-2 could also be MtrBTN2 tumors as discussed by Skazina *et al.* (2021). The closest mtCOI sequences of healthy mussels were obtained from populations of the Northwest Pacific, Northwest and Northeast Atlantic but none from populations of the Northeast Pacific, as previously discussed by Skazina *et al.* (2021). In principle, molecular clock calibration using the 100Ky old Tx101A sequence (Der Sarkissian *et al.* 2020), would have allowed us to date the MRCA of every MtrBTN2. However, the closest haplotypes have no substitutions from the ancestral node and four haplotypes only one or two (Figure 6B), and the MtrBTN2 clade appears to have much longer branches, suggesting an accelerated evolutionary rate that prevents accurate datation of the MRCA.

#### 4. Discussion

We investigated the presence of transmissible cancer in samples of European *M. edulis* and *M. galloprovincialis* mussels and their hybrids. We used SNP KASP genotyping to screen a large number of mussels in order to detect transmissible cancers and at the same time studied the genotypes of the infected hosts and their tumors. We detected MtrBTN2, a previously described transmissible cancer that originated in a *M. trossulus* host, but no other transmissible cancers were found in the inspected area with our screening approach. The MtrBTN2 was known to affect *M. edulis*, (Riquet *et al.* 2017, Burioli *et al.* 2019), *M. chilensis* (Yonemitsu *et al.* 2019) and *M. trossulus* (Skazina *et al.* 2021) but we also found cases in *M. galloprovincialis* and hybrid individuals, thus extending the known spreading capacity of MtrBTN2 to an additional species of the complex. Despite an overall low prevalence (0.42%), it appeared more prevalent in *M. edulis* hosts (0.78%) than *M. galloprovincialis* (0.07%) and hybrid (0.26%) hosts. We also found neoplastic mussels without evidence of genetic chimerism, suggesting the presence of conventional non-transmissible cancers. The analysis of the genotypes of 6 tumors from *M. edulis* hosts confirmed the *M. trossulus* origin of MtrBTN2. We surprisingly observed a high amount of nuclear and mitochondrial polymorphism in MtrBTN2 samples. This polymorphism defines two sublineages that co-exist and spread in the same populations, as it has been observed in other transmissible cancers (CTVT: Baez-Ortega *et al.* 2019, DFTD: Kwon *et al.* 2020, Patton *et al.* 2020).

The MtrBTN2 prevalences found in our study were consistent with DN prevalences previously reported in *M. edulis* (around 1%, for review see Carballal et al. 2015, Muttray et al. 2018) and *M. galloprovincialis* mussels (rarer and when detected < 1%, for review see Carballal et al. 2015, Matozzo et al. 2018), but slightly lower than the prevalence of *M. trossulus* BTN found in Brittany (3.4%, Burioli et al. 2019). It is likely that we missed the very early stages of the disease. Indeed, MtFF estimates the excess of *M. trossulus* allele fluorescence and when the proportion of cancer cells is too low, the host genotype dominates the signal. The two other indexes we used (PUF and 2TD) did not help us to find additional tumors, with the exception of one sample (CamaretA\_61) that was infected by a few neoplastic cells and showed a MtFF value at the limit of the outlier threshold (Figure 4A, green triangle up). Although the prevalence of MtrBTN2 could be slightly underestimated, the detection limit of MtrBTN2 is expected to be similar for *M. edulis*, *M. galloprovincialis* and their hybrids.

The higher prevalence in *M. edulis* than *M. galloprovincialis* and their hybrids in between could reflect a lower sensitivity (higher efficient immune response) of *M. galloprovincialis* to MtrBTN2. Indeed, Fuentes et al. (2002) found a higher DN prevalence in hybrid lab crosses (*edulis* x *galloprovincialis*) than *M. galloprovincialis* crosses (*galloprovincialis* x *galloprovincialis*), but unfortunately no *M. edulis* lab crosses (*edulis* x *edulis*) were reported. Moreover, it has been shown that *M. galloprovincialis* is less affected by Pea crab and Bucephalid trematode parasitism than *M. edulis* living in the same region (Seed 1969, Coustau et al. 1991a). Also, a study of mussel hemocyte transcriptomes shows higher sequence numbers related to immune genes in *M. galloprovincialis* compared to *M. edulis* after in vitro and in vivo stimulation with pathogen associated molecular patterns and heat-inactivated *Vibrio anguillarum*, respectively (Moreira et al. 2018). Beyond that, *M. galloprovincialis* is a worldwide invader (Popovic et al. 2020) with a fitness advantage in natural hybrid zones (Skibinski 1983; Coustau et al. 1991b; Hilbish et al. 1994; reviewed in Gardner 1994) and in lab crosses (Bierne 2006) which, in different ways, appear to be more efficient than *M. edulis*.

Given the existence of conventional cancers is a prerequisite for transmissible cancer to evolve in the initial founder host, the observation of eight non-transmissible neoplasia was not unexpected. Cases of conventional non-transmissible cancer have also been described in Chile where two tumors without transmissible cancer hallmarks were found in the same area as two MtrBTN2 tumors (Yonemitsu et al. 2019). The eight tumors (type B, see micrographs C and D in Figure 4) were collected on the same date (April 2017), in the bay of Brest, where level of lead (Pb) were found above the regulatory threshold for human food (Charles et al. 2020).

Environmental contaminants have previously been proposed as an etiological cause of DN (Lowe and Moore 1978, Hillman 1993), such as biotoxins (Landsberg 1996) and viral factors (Elston et al. 1988, Moore et al. 1991, Rasmussen 1986), which could be involved in the high prevalence of type B diseased individuals in this specific area and date. However, we also sampled mussels from bay of Brest in February, when Pb levels were also high (Charles et al. 2020), but no type B cancerous mussels were found. The coexistence of a transmissible cancer with non-transmissible cancers, warn for the systematic use of molecular markers to verify the transmissible nature of DN. More generally, wildlife cancer research needs to consider a more general use of molecular markers (Leathlobhair et al. 2017).

Polymorphism in transmissible cancer at SNPs ascertained in host populations can suggest multiple emergences, or a single emergence with somatic back mutations (point or structural mutations) or a deviation from clonal propagation (horizontal gene

transfer, cell fusion, parasexual recombination, etc.). Comparison of our mtCOI tumor sequences with published MtrBTN2 mtCOI sequences (Yonemitsu et al. 2019, Skazina et al. 2021), and the grouping of the tumors with 54 fixed SNPs, suggests a single emergence is the most parsimonious interpretation. However, we have to assume that some published mtCOI sequences have been obtained from uncharacterized cancerous mussels, as argued by Skazina et al. (2021). This is likely for 62mc10 given how it has been obtained in a heteroplasmic mussel after extensive screening for recombinant mitogenomes (Śmietanka and Burzyński 2017), but only based on phylogenetic proximity for Bf-48, SA1 and SA5 (Figure 6B). Indeed, the study of Riquet et al. (2017) suggests cancer genetic signals exist in the standard population genetics literature of mussels. Thus, we cannot rigorously reject the hypothesis of a twin co-emergence in an ancestral (or contemporary unsampled) *M. trossulus* population with a different genetic composition than sampled *M. trossulus* populations. The recent independent emergence of two transmissible facial-tumour lineages (DFTD1 and DFTD2) in Tasmanian devils suggests that when favorable conditions are met, multiple emergences can succeed in a very short time period (Ujvari et al. 2016). Interestingly, we found *M. edulis*-state alleles on some SNPs which could suggest an ancient origin, in a *M. trossulus* population which was less divergent than the actual population, or an origin in a recombinant hybrid between *M. trossulus* and *M. edulis* (Figure 5B, H1 and H2). However, it is also likely that somatic null alleles caused by structural variations or point mutations in primer-binding sites lead to enhanced detection of the host alleles at these SNPs.

Given that we analysed only 74 nucleotide positions and we observed polymorphisms at 20 SNPs (27%), we believe polymorphism observed among tumors is unlikely to be due to homoplasic point back mutations at the SNP site. Such a high rate of homoplasic back substitutions would correspond to a very old divergence that is not supported by mtCOI divergence (Figure 6) and Elongation Factor  $\alpha$  gene divergence (Yonemitsu et al. 2019, Skazina et al. 2021). The polymorphisms are more likely due to mutations in primer-binding sites or structural variations, such as insertion-deletion and chromosomal losses or gains, or recombination leading to loss of heterozygosity, that are widespread in cancer (McClelland 2017, Nichols et al. 2020) and have been reported in other transmissible cancers (e.g. in DFT1 and DFT2, Kwon et al. 2020).

According to the genotyping method used (competition of the two allele fluorescences), partial or complete chromosomal losses or gains results in a modification of the ratio of fluorescences of the two alleles of hyperploid MtrBTN2 tumors and of the ratio of tumor and host genotypes. We observed that most somatic variations that could be oriented resulted in gain of heterozygosity rather than loss (10 gains mostly of *M. edulis*-allele state, 2 losses), supporting the hypothesis that the variation is due to somatic null alleles leading to enhanced detection of the host *M. edulis* alleles. The same hypothesis would also explain the divergence between the two sublineages (H1 in Figure 5B). Another, although maybe less likely, hypothesis could be horizontal gene transfer from host to tumors. This hypothesis would be supported by 5 SNPs for which the host allele state does not explain the heterozygous state of tumors (Figure 5C, black star). The same hypothesis would also explain the divergence between the two sublineages under the hypothesis of a hybrid founder genotype (H2 in Figure 5B). Indeed, spontaneous fusion between tumor and host cells (reviewed in Noubissi and Ogle 2016) or between tumor cells (Amend et al. 2019, Pienta et al. 2020) has been shown to bring tumor heterogeneity which increases invasion, migratory capacities and resistance to therapies. Recombination can also occur after cancer cell fusion

(Miroshnychenko et al. 2021). Fusion between two BTN lineages or between BTN and conventional cancers could explain hyperploidy and variable VAFFs observed. We also note that heterozygous-like SNPs, with an intermediate dosage of the two alleles, and homozygotes for *M. edulis*-state allele, do not show evidence of a decreased fluorescent signal as it could have been expected under null allele hypothesis (Figure S7). However, in the absence of further convincing evidence, we suggest somatic null alleles and mutations remains the most parsimonious explanation to the diversity observed.

Finally, one of the 6 tumors, L9\_3, had a divergent mtCOI sequence to the other MtrBTN2 mtCOI sequences sampled in *M. edulis* and *M. chilensis* hosts (Figure 6). The mismatch between the nuclear and the mitochondrial trees suggests the existence of mitochondrial capture or recombination events in MtrBTN2. This phenomenon has been described several times in CTVT where mitogenomes reveal multiple independent horizontal transfers of mitochondria from hosts (Strakova et al. 2016, 2020). In addition, recombinant mitochondria have been observed in two MtrBTN2 tumors from Chile (Yonemitsu et al. 2019). Here however, the capture/recombination would have happened from another MtrBTN2 tumor, possibly by cellular fusion as explained above. Mitochondrial capture could also be an explanation for the phylogenetic positions of published mtCOI haplotypes that cluster with MtrBTN2 haplotypes while sampled from mussels that were not characterized for DN and assumed healthy (Bf-48, SA1 and SA5 in Figure 6B). Under this hypothesis these captures would involve related mitochondria and would have probably happened early during the diversification of MtrBTN2 in the ancestral population.

## 5. Conclusion

Our study has revealed that MtrBTN2, a bivalve transmissible neoplasia that originated in a *M. trossulus* host, affects European *M. edulis* at a low prevalence and also affects *M. galloprovincialis* and hybrids although at a much lower rate. No other transmissible cancer has been identified with our approach, but conventional non-transmissible cancers have been observed, which suggests that molecular markers should be used to diagnose the transmissible status of disseminated neoplasia. Somewhat surprisingly we found polymorphism among MtrBTN2 tumors, probably due to somatic structural variation leading to null alleles. The origin of this polymorphism will need to be further investigated with alternative sequencing methods. The coexistence of sublineages in the same geographic region suggests a history of MtrBTN2 diversification with a complex biogeography, as has also been reported in CTVT and DFTD (Baez-Ortega et al. 2019, Know et al. 2020, Patton et al.2020).

## Acknowledgements

This work was supported by Montpellier Université d'Excellence (BLUECANCER project), Agence Nationale de la Recherche (TRANSCAN project, ANR-18-CE35-0009), the French DPMA (Direction des pêches maritimes et de l'aquaculture, DPMA-2017-MORBLEU Conventions N°: 16/1212569 & 17/1212952) and Conseil Régional de Normandie, LabEx CeMED (Junior Research Team funding).

## References

- Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B. & Delsuc, F. (2020). MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Molecular Ecology Resources*, 20(4), 892–905. <https://doi.org/10.1111/1755-0998.13160>
- Amend, S. R., Torga, G., Lin, K. C., KostECKa, L. G., de Marzo, A., Austin, R. H., & Pienta, K. J. (2019). Polyploid giant cancer cells: Unrecognized actuators of tumorigenesis, metastasis, and resistance. *Prostate*. John Wiley and Sons Inc. <https://doi.org/10.1002/pros.23877>
- Baez-Ortega, A., Gori, K., Strakova, A., Allen, J. L., Allum, K. M., Banske-Issa, L., ... Murchison, E. P. (2019). Somatic evolution and global expansion of an ancient transmissible cancer lineage. *Science*, 365(6452). <https://doi.org/10.1126/science.aau9923>
- Benabdelmouna, A., & Ledu, C. (2016). The mass mortality of blue mussels (*Mytilus* spp.) from the Atlantic coast of France is associated with heavy genomic abnormalities as evidenced by flow cytometry. *Journal of Invertebrate Pathology*, 138, 30–38. <https://doi.org/10.1016/j.jip.2016.06.001>
- Benabdelmouna, A., Saunier, A., Ledu, C., Travers, M. A., & Morga, B. (2018). Genomic abnormalities affecting mussels (*Mytilus edulis-galloprovincialis*) in France are related to ongoing neoplastic processes, evidenced by dual flow cytometry and cell monolayer analyses. *Journal of Invertebrate Pathology*, 157, 45–52. <https://doi.org/10.1016/j.jip.2018.08.003>
- Bierne, N., Bonhomme, F., Boudry, P., Szulkin, M., & David, P. (2006). Fitness landscapes support the dominance theory of post-zygotic isolation in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Proceedings of the Royal Society B January*, 273(NUM), 1253–1260. <https://doi.org/10.1098/rspb.20053440>
- Breton, S., Burger, G., Stewart, D. T., & Blier, P. U. (2006). Comparative analysis of gender-associated complete mitochondrial genomes in marine mussels (*Mytilus* spp.). *Genetics*, 172(2), 1107–1119. <https://doi.org/10.1534/genetics.105.047159>
- Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., ... Bustamante, C. D. (2012). Pcadmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Human Biology*, 84(4), 343–364. <https://doi.org/10.3378/027.084.0401>
- Burioli, E. A. V., Trancart, S., Simon, A., Bernard, I., Charles, M., Oden, E., ... Houssin, M. (2019). Implementation of various approaches to study the prevalence, incidence and progression of disseminated neoplasia in mussel stocks. *Journal of Invertebrate Pathology*, 168(April), 107271. <https://doi.org/10.1016/j.jip.2019.107271>
- Campalans, M., Gonzalez, M., & Rojas, P. (Escuela de C. del M. U. C. de V. C. 1020, V. (Chile)). (1998). Neoplasia in *Mytilus chilensis* cultivated in Chiloe Island (Chile). *Bulletin of the European Association of Fish Pathologists (United Kingdom)*. Retrieved from <https://agris.fao.org/agris-search/search.do?recordID=GB1997047632>
- Carballal, M. J., Barber, B. J., Iglesias, D., & Villalba, A. (2015). Neoplastic diseases of marine bivalves. *Journal of Invertebrate Pathology*, 131, 83–106. <https://doi.org/10.1016/J.JIP.2015.06.004>
- Carella, F., Figueras, A., Novoa, B., & De Vico, G. (2013). Cytomorphology and PCNA expression pattern in bivalves *Mytilus galloprovincialis* and *Cerastoderma edule* with haemic neoplasia. *Diseases of Aquatic Organisms*, 105(1), 81–87. <https://doi.org/10.3354/dao02612>



Carrasco, N., Arzul, I., Berthe, F., Fernández-Tejedor, M., Durfort, M., & Furones, M. (2008). Delta de l'Ebre is a natural bay model for *Marteilia* spp. (Paramyxea) dynamics and life-cycle studies. *Diseases of Aquatic Organisms*, 79(1), 65–73. <https://doi.org/10.3354/dao01844>

Charles, M., Bernard, I., Villalba, A., Oden, E., Burioli, E. A. V., Allain, G., ... Houssin, M. (2020). High mortality of mussels in northern Brittany – Evaluation of the involvement of pathogens, pathological conditions and pollutants. *Journal of Invertebrate Pathology*, 170, 107308. <https://doi.org/10.1016/j.jip.2019.107308>

Coustau, Christine, Renaud, F., Maillard, C., Pasteur, N., & Delay, B. (1991a). Differential susceptibility to a trematode parasite among genotypes of the *Mytilus edulis/galloprovincialis* complex. *Genetical Research*, 57(3), 207–212. <https://doi.org/10.1017/S0016672300029359>

Coustau, C., Renaud, F., & Delay, B. (1991b). Genetic characterization of the hybridization between *Mytilus edulis* and *M. galloprovincialis* on the Atlantic coast of France. *Marine Biology*, 111(1), 87–93. <https://doi.org/10.1007/BF01986350>

Crego-Prieto, V., Ardura, A., Juanes, F., Roca, A., Taylor, J. S., & Garcia-Vazquez, E. (2015). Aquaculture and the spread of introduced mussel genes in British Columbia. *Biological Invasions*, 17(7), 2011–2026. <https://doi.org/10.1007/s10530-015-0853-z>

Cremonte, F., Puebla, C., Tillería, J., & Videla, V. (2015). Histopathological survey of the mussel *Mytilus chilensis* (Mytilidae) and the clam *Gari solida* (Psammobiidae) from southern Chile. *Latin American Journal of Aquatic Research*, 43(1), 248–254. <https://doi.org/10.3856/vol43-issue1-fulltext-21>

Cuenca, J., Aleza, P., Navarro, L., & Ollitrault, P. (2013). Assignment of SNP allelic configuration in polyploids using competitive allele-specific PCR: application to citrus triploid progeny. *Annals of Botany*, 111(4), 731–742. <https://doi.org/10.1093/aob/mct032>

Cuppen, E. (2007). Genotyping by Allele-Specific Amplification (KASPar). *CSH Protocols*, 2007, pdb.prot4841. <https://doi.org/10.1101/PDB.PROT4841>

Der Sarkissian, C., Möller, P., Hofman, C. A., Ilsøe, P., Rick, T. C., Schiøtte, T., ... Orlando, L. (2020). Unveiling the Ecological Applications of Ancient DNA From Mollusk Shells. *Frontiers in Ecology and Evolution*, 8, 37. <https://doi.org/10.3389/fevo.2020.00037>

Díaz, S., Iglesias, D., Villalba, A., & Carballal, M. J. (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*, 39(9), 1027–1042. <https://doi.org/10.1111/jfd.12436>

Elston, R. A., Kent, M. L., & Drum, A. S. (1988). *Progression, lethality and remission of hemic neoplasia in the bay mussel Mytilus edulis* (Vol. 4). Retrieved from <https://www.intres.com/articles/dao/4/d004p135.pdf>

Elston, R. A., Moore, J. D., & Brooks, K. (1992). Disseminated neoplasia of bivalve molluscs. *Rev. Aquat. Sci.*, 6(5–6), 405–466. Retrieved from <https://www.scopus.com/record/display.uri?eid=2-s2.0-0001141213&origin=inward&txGid=1704476f390ea78657746d8485ae4e7e>

Farley, C. A. (1969). Sarcomatoid proliferative disease in a wild population of blue mussels (*mytilus edulis*). *Journal of the National Cancer Institute*, 43(2), 509–516. <https://doi.org/10.1093/jnci/43.2.509>

Figueras, A. J., Jardon, C. F., & Caldas, J. R. (1991). Diseases and parasites of rafted mussels (*Mytilus galloprovincialis* Lmk): preliminary results. *Aquaculture*, 99(1–2), 17–33. [https://doi.org/10.1016/0044-8486\(91\)90285-F](https://doi.org/10.1016/0044-8486(91)90285-F)

Fraïsse, C., Belkhir, K., Welch, J. J., & Bierne, N. (2016). Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Molecular Ecology*, 25(1), 269–286. <https://doi.org/10.1111/mec.13299>

Fuentes, J., López, J., Mosquera, E., Vázquez, J., Villalba, A., & Álvarez, G. (2002). Growth, mortality, pathological conditions and protein expression of *Mytilus edulis* and *M. galloprovincialis* crosses cultured in the Ría de Arousa (NW of Spain). *Aquaculture*, 213(1–4), 233–251. [https://doi.org/10.1016/S0044-8486\(02\)00046-7](https://doi.org/10.1016/S0044-8486(02)00046-7)

Galimany, E., & Sunila, I. (2008). Several Cases of Disseminated Neoplasia in Mussels *Mytilus edulis* (L.) in Western Long Island Sound. *Journal of Shellfish Research*, 27(5), 1201–1207. <https://doi.org/10.2983/0730-8000-27.5.1201>

Garcia-Souto, D., Bruzos, A. L., Diaz, S., Rocha, S., Pequeño, A., Roman-Lewis, C. F., ... Tubio, J. M. C. (2021). Mitochondrial genome sequencing of marine leukemias reveals cancer contagion between clam species in the Seas of Southern Europe. *BioRxiv*. <https://doi.org/10.1101/2021.03.10.434714>

Gardner, J. P. A. (1994). The *Mytilus edulis* species complex in Southwest England: Multi-locus heterozygosity, background genotype and a fitness correlate. *Biochemical Systematics and Ecology*, 22(1), 1–11. [https://doi.org/10.1016/0305-1978\(94\)90109-0](https://doi.org/10.1016/0305-1978(94)90109-0)

Gombač, M., Sitar, R., Pogačnik, M., Arzul, I., & Jenčič, V. (2013). Haemocytic neoplasia in Mediterranean mussels (*Mytilus galloprovincialis*) in the Slovene Adriatic Sea. *Marine and Freshwater Behaviour and Physiology*, 46(2), 135–143. <https://doi.org/10.1080/10236244.2013.782736>

Grant, B. J., Rodrigues, A. P. C., Elsayy, K. M., Mccammon, J. A., & Caves, L. S. D. (2006). Bio3d: an R package for the comparative analysis of protein structures, 22(21), 2695–2696. <https://doi.org/10.1093/bioinformatics/btl461>

Green, M., & Alderman, D. J. (1983). Neoplasia in *Mytilus edulis* L. from United Kingdom waters. *Aquaculture*, 30(1–4), 1–10. [https://doi.org/10.1016/0044-8486\(83\)90146-1](https://doi.org/10.1016/0044-8486(83)90146-1)

Hilbish, T. J., Bayne, B. L., & Day, A. (1994). Genetics of physiological differentiation within the marine mussel genus *Mytilus*. *Evolution*, 48(2), 267–286. <https://doi.org/10.1111/j.1558-5646.1994.tb01311.x>

Hillman RE. (1993). Relationship of environmental contaminants to occurrence of neoplasia in *Mytilus edulis* populations from east and west coast mussel-watch sites. *Journal of Shellfish*, 12:109.

Howard, D. W., Lewis, E. J., Keller, B. J., & Smith, C. S. (2004). *Histological Techniques for Marine Bivalve Mollusks and Crustaceans - Google Livres*. National Oceanic and Atmospheric Administration Technical Memorandum, National Ocean Service, National Center for Coastal Ocean Science, 5, Oxford, MD., 218 pp.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>

Kumar, S., Stecher, G., Tamura, K., & Dudley, J. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets Downloaded from. *Mol. Biol. Evol*, 33(7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>

Kwon, Y. M., Gori, K., Park, N., Potts, N., Swiftid, K., Wang, J., ... Murchisonid, E. P. (2020). Evolution and lineage dynamics of a transmissible cancer in Tasmanian devils. <https://doi.org/10.1371/journal.pbio.3000926>

Laakkonen, H. M., Hardman, M., Strelkov, P., & Väinölä, R. (2021). Cycles of trans-Arctic dispersal and vicariance, and diversification of the amphi-boreal marine fauna. In *Journal of Evolutionary Biology* (Vol. 34, pp. 73–96). Blackwell Publishing Ltd. <https://doi.org/10.1111/jeb.13674>

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

Layton, K. K. S., Martel, A. L., & Hebert, P. D. N. (2014). Patterns of DNA barcode variation in canadian marine molluscs. *PLoS ONE*, 9(4), e95003. <https://doi.org/10.1371/journal.pone.0095003>

Leathlobhair, M. N., Gulland, F. M. D., & Murchison, E. P. (2017). No evidence for clonal transmission of urogenital carcinoma in California sea lions (*Zalophus californianus*). *Wellcome Open Research*, 2, 46. <https://doi.org/10.12688/wellcomeopenres.11483.1>

Leathlobhair, M. N., Perri, A. R., Irving-Pease, E. K., Witt, K. E., Linderholm, A., Haile, J., ... Frantz, L. A. F. (2018). The evolutionary history of dogs in the Americas. *Science*, 361(6397), 81–85. <https://doi.org/10.1126/science.aao4776>

Lesser, M. P., Bailey, M. A., Merselis, D. G., & Morrison, J. R. (2010). Physiological response of the blue mussel *Mytilus edulis* to differences in food and temperature in the Gulf of Maine. *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, 156(4), 541–551. <https://doi.org/10.1016/j.cbpa.2010.04.012>

Li, D., Luo, R., Liu, C. M., Leung, C. M., Ting, H. F., Sadakane, K., ... Lam, T. W. (2016). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*. Academic Press Inc. <https://doi.org/10.1016/j.ymeth.2016.02.020>

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>

Lohrmann, K. B., Bustos, E., Rojas, R., Navarrete, F., Robotham, H., & Bignell, J. (2019). Histopathological assessment of the health status of *Mytilus chilensis* (Hupé 1854) in southern Chile. *Aquaculture*, 503, 40–50. <https://doi.org/10.1016/j.aquaculture.2018.12.080>

Lowe, D. M., & Moore, M. N. (1978). Cytology and Quantitative Cytochemistry of a Proliferative Atypical Hemocytic Condition in *Mytilus edulis* (Bivalvia, Mollusca). *Journal of the National Cancer Institute*, 60(6), 1455–1459. <https://doi.org/10.1093/jnci/60.6.1455>

Marko, P. B., Hoffman, J. M., Emme, S. A., McGovern, T. M., Keever, C. C., & Nicole Cox, L. (2010). The “expansion-Contraction” model of Pleistocene biogeography: Rocky shores suffer a sea change? *Molecular Ecology*, 19(1), 146–169. <https://doi.org/10.1111/j.1365-294X.2009.04417.x>

Matozzo, V., Ercolini, C., Serracca, L., Battistini, R., Rossini, I., Granato, G., ... Quaglio, F. (2018). Assessing the health status of farmed mussels (*Mytilus galloprovincialis*) through histological, microbiological and biomarker analyses. *Journal of Invertebrate Pathology*, 153, 165–179. <https://doi.org/10.1016/j.jip.2018.02.018>

McClelland, S. E. (2017). Role of chromosomal instability in cancer progression. *Endocrine-Related Cancer*. Bioscientifica Ltd. <https://doi.org/10.1530/ERC-17-0187>

Md, V., Misra, S., Li, H., & Aluru, S. (2019). Efficient architecture-aware acceleration of BWAMEM for multicore systems. In *Proceedings - 2019 IEEE 33rd International Parallel and Distributed Processing Symposium, IPDPS 2019* (pp. 314–324). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/IPDPS.2019.00041>

Metzger, M. J., Reinisch, C., Sherry, J., & Goff, S. P. (2015). Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams. *Cell*, 161(2), 255–263. <https://doi.org/10.1016/j.cell.2015.02.042>

Metzger, M. J., Villalba, A., Carballal, M. J., Iglesias, D., Sherry, J., Reinisch, C., ... Goff, S. P. (2016). Widespread transmission of independent cancer lineages within

multiple bivalve species. *Nature*, 534(7609), 705–709. <https://doi.org/10.1038/nature18599>

Miroshnychenko, D., Baratchart, E., Ferrall-Fairbanks, M. C., Velde, R. Vander, Laurie, M. A., Bui, M. M., ... Marusyk, A. (2021). Spontaneous cell fusions as a mechanism of parasexual recombination in tumour cell populations. *Nature Ecology and Evolution*, 1–13. <https://doi.org/10.1038/s41559-020-01367-y>

MIX, M. C. (1983). Haemic neoplasms of bay mussels, *Mytilus edulis* L., from Oregon: occurrence, prevalence, seasonality and histopathological progression. *Journal of Fish Diseases*, 6(3), 239–248. <https://doi.org/10.1111/j.1365-2761.1983.tb00072.x>

Moore, J. D., Elston, R. A., Drum, A. S., & Wilkinson, M. T. (1991). Alternate pathogenesis of systemic neoplasia in the bivalve mollusc *Mytilus*. *Journal of Invertebrate Pathology*, 58(2), 231–243. [https://doi.org/10.1016/0022-2011\(91\)90067-Z](https://doi.org/10.1016/0022-2011(91)90067-Z)

Moreira, R., Balseiro, P., Forn-Cuní, G., Milan, M., Bargelloni, L., Novoa, B., & Figueras, A. (2018). Bivalve transcriptomics reveal pathogen sequences and a powerful immune response of the Mediterranean mussel (*Mytilus galloprovincialis*). *Marine Biology*, 165(4), 1–20. <https://doi.org/10.1007/s00227-018-3308-0>

Murgia, C., Pritchard, J. K., Kim, S. Y., Fassati, A., & Weiss, R. A. (2006). Clonal Origin and Evolution of a Transmissible Cancer. *Cell*, 126(3), 477–487. <https://doi.org/10.1016/j.cell.2006.05.051>

Muttray, A. F., & Vassilenko, K. (2018). Mollusca: Disseminated neoplasia in bivalves and the p53 protein family. In *Advances in Comparative Immunology* (pp. 953–979). Springer International Publishing. [https://doi.org/10.1007/978-3-319-76768-0\\_28](https://doi.org/10.1007/978-3-319-76768-0_28)

Nichols, C. A., Gibson, W. J., Brown, M. S., Kosmicki, J. A., Busanovich, J. P., Wei, H., ... Beroukhim, R. (2020). Loss of heterozygosity of essential genes represents a widespread class of potential cancer vulnerabilities. *Nature Communications*, 11(1), 1–14. <https://doi.org/10.1038/s41467-020-16399-y>

Noubissi, F. K., & Ogle, B. M. (2016). Cancer cell fusion: Mechanisms slowly unravel. *International Journal of Molecular Sciences*. MDPI AG. <https://doi.org/10.3390/ijms17091587>

Paradis, E., & Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>

Patton, A. H., Lawrance, M. F., Margres, M. J., Kozakiewicz, C. P., Hamede, R., Ruiz-Aravena, M., ... Storfer, A. (2020). A transmissible cancer shifts from emergence to endemism in Tasmanian devils. *Science*, 370(6522), eabb9772. <https://doi.org/10.1126/science.abb9772>

Pearse, A. M., & Swift, K. (2006). Allograft theory: Transmission of devil facial-tumour disease. *Nature*, 439(7076), 549. <https://doi.org/10.1038/439549a>

Pienta, K. J., Hammarlund, E. U., Axelrod, R., Brown, J. S., & Amend, S. R. (2020). Polyaneuploid cancer cells promote evolvability, generating lethal cancer. *Evolutionary Applications*, 13(7), 1626–1634. <https://doi.org/10.1111/eva.12929>

Popovic, I., Matias, A. M. A., Bierne, N., & Riginos, C. (2020). Twin introductions by independent invader mussel lineages are both associated with recent admixture with a native congener in Australia. *Evolutionary Applications*, 13(3), 515–532. <https://doi.org/10.1111/eva.12857>

Pye, R. J., Pemberton, D., Tovar, C., Tubio, J. M. C., Dun, K. A., Fox, S., ... Woods, G. M. (2016). A second transmissible cancer in Tasmanian devils. *Proceedings of the*

*National Academy of Sciences*, 113(2), 374–379.  
<https://doi.org/10.1073/pnas.1519691113>

R Core Team. (2020). R: A language and environment for statistical computing. Retrieved from <http://www.r-project.org/index.html>

Rasmussen, L. P. D. (1986). Occurrence, prevalence and seasonality of neoplasia in the marine mussel *Mytilus edulis* from three sites in Denmark. *Marine Biology: International Journal on Life in Oceans and Coastal Waters*, 92(1), 59–64.  
<https://doi.org/10.1007/BF00392746>

Rebeck, C. A., Thomas, R., Breen, M., Leroi, A. M., & Burt, A. (2009). ORIGINS AND EVOLUTION OF A TRANSMISSIBLE CANCER. *Evolution*, 63(9), 2340–2349.  
<https://doi.org/10.1111/j.1558-5646.2009.00724.x>

Riquet, F., Simon, A., & Bierne, N. (2017). Weird genotypes? Don't discard them, transmissible cancer could be an explanation. *Evolutionary Applications*, 10(2), 140–145. <https://doi.org/10.1111/eva.12439>

Seed, R. (1969). The incidence of the Pea crab, *Pinnotheres pisum* in the two types of *Mytilus* (Mollusca: Bivalvia) from Padstow, south-west England. *Journal of Zoology*, 158(4), 413–420. <https://doi.org/10.1111/j.1469-7998.1969.tb02158.x>

Semagn, K., Babu, R., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. *Molecular Breeding*. Springer. <https://doi.org/10.1007/s11032-013-9917-x>

Simon, A., Arbiol, C., Nielsen, E. E., Couteau, J., Sussarellu, R., Burgeot, T., ... Bierne, N. (2020). Replicated anthropogenic hybridisations reveal parallel patterns of admixture in marine mussels. *Evolutionary Applications*, 13(3), 575–599. <https://doi.org/10.1111/eva.12879>

Simon, A., Fraïsse, C., El Ayari, T., Liautard-Haag, C., Strelkov, P., Welch, J. J., & Bierne, N. (2021). How do species barriers decay? Concordance and local introgression in mosaic hybrid zones of mussels. *Journal of Evolutionary Biology*, 34(1), 208–223. <https://doi.org/10.1111/jeb.13709>

Skazina, M., Odintsova, N., Maiorova, M., Ivanova, A., Väinölä, R., & Strelkov, P. (2021). First description of a widespread *Mytilus trossulus*-derived bivalve transmissible cancer lineage in *M. trossulus* itself. *Scientific Reports*, 11(5809). <https://doi.org/10.1038/s41598-021-85098-5>

Skibinski, D. O. F., Beardmore, J. A., & Cross, T. F. (1983). Aspects of the population genetics of *Mytilus* (Mytilidae; Mollusca) in the British Isles. *Biological Journal of the Linnean Society*, 19(2), 137–183. <https://doi.org/10.1111/j.1095-8312.1983.tb00782.x>

Śmietanka, B., & Burzyński, A. (2017). Disruption of doubly uniparental inheritance of mitochondrial DNA associated with hybridization area of European *Mytilus edulis* and *Mytilus trossulus* in Norway. *Marine Biology*, 164(11), 209. <https://doi.org/10.1007/s00227-017-3235-5>

Strakova, A., Leathlobhair, M. N., Wang, G. D., Yin, T. T., Airikkala-Otter, I., Allen, J. L., ... Murchison, E. P. (2016). Mitochondrial genetic diversity, selection and recombination in a canine transmissible cancer. *eLife*, 5(MAY2016), 1–25. <https://doi.org/10.7554/eLife.14552>

Strakova, A., Nicholls, T. J., Baez-Ortega, A., Ní Leathlobhair, M., Sampson, A. T., Hughes, K., ... Murchison, E. P. (2020). Recurrent horizontal transfer identifies mitochondrial positive selection in a transmissible cancer. *Nature Communications*, 11(1), 1–11. <https://doi.org/10.1038/s41467-020-16765-w>

Sunila, I. (1987). Histopathology of mussels ( *Mytilus edulis* L.) from the Tvarminne area, the Gulf of Finland ( Baltic Sea). *Ann. Zool. Fennici*, 24, 55–69. Retrieved from <https://www.researchgate.net/publication/279565817>

Ujvari, B., Gatenby, R. A., & Thomas, F. (2016). The evolutionary ecology of transmissible cancers. *Infection, Genetics and Evolution*, 39, 293–303. <https://doi.org/10.1016/J.MEEGID.2016.02.005>

Usheva, L. N., & Frolova, L. T. (2000). A connective tissue tumor in the mussel *Mytilus trossulus* from a polluted region of Nakhodka Bay, the Sea of Japan. *Ontogenez*, 31(1), 63–70.

Väinölä, R., & Strelkov, P. (2011). *Mytilus trossulus* in Northern Europe. *Marine Biology*, 158(4), 817–833. <https://doi.org/10.1007/s00227-010-1609-z>

Vassilenko, E., & Baldwin, S. A. (2014). Using flow cytometry to detect haemic neoplasia in mussels (*Mytilus trossulus*) from the Pacific Coast of Southern British Columbia, Canada. <https://doi.org/10.1016/j.jip.2014.02.002>

Villalba, A., Mourelle, S., Carballal, M., & López, C. (1997). Symbionts and diseases of farmed mussels *Mytilus galloprovincialis* throughout the culture process in the Rías of Galicia (NW Spain). *Diseases of Aquatic Organisms*, 31(2), 127–139. <https://doi.org/10.3354/dao031127>

Yonemitsu, M. A., Giersch, R. M., Polo-Prieto, M., Hammel, M., Simon, A., Cremonte, F., ... Metzger, M. J. (2019). A single clonal lineage of transmissible cancer identified in two marine mussel species in South America and Europe. *ELife*, 8. <https://doi.org/10.7554/eLife.47788>

Zbawicka, M., Wenne, R., & Burzyński, A. (2014). Mitogenomics of recombinant mitochondrial genomes of Baltic Sea *Mytilus* mussels. *Molecular Genetics and Genomics*, 3(6), 1275–1287. <https://doi.org/10.1007/s00438-014-0888-3>

## Author Contributions

M.H., A.S., M.J.M and N.B. designed the project. All authors contributed to sampling, sample preparation and data curation. M.H., A.S., C.A. conducted the molecular work with guidance from G.C., D.D-G and N.B., and A.V., E.A.V.B. and A.B. performed histology, cytology and flow cytometry. M.H., A.S., J.J.W. and N.B. analyzed the data. M.H. and N.B. wrote the initial manuscript. All authors contributed insights about interpretation of results, and critically reviewed and edited the manuscript.

## Data accessibility

Original genotype data from LGC (fluorescence coordinate and calling) and VAFF are available on DRYAD (doi:10.5061/dryad.sj3tx9652). Accession numbers of sequences used in Figure 6 are available in Table S4. More generally, all data generated by this study are available in Supplementary Datasets and Supplementary Information.

## Supplementary tables

**Table S1: Description of the sampled locations of *M. trossulus* references (PopRef) and *M. edulis* and *M. galloprovincialis* individuals (PopEu).** Pop: population name; N: number of individuals sampled; Lat: latitude; Long: longitude; MtrBTN1: number of individuals with MtrBTN1 tumor; MtrBTN2: number of individuals with MtrBTN2, the tumor detected in this study; DNB: number of individuals with disseminated neoplasia type B (non-transmissible cancer).

**Table S2: Bi-allelic SNPs informations.** Freq\_tross, freq\_edu, freq\_gallo correspond to allele frequency in references populations; Allele: allele the most frequent in *M. trossulus* references populations; MtFF: 10 diagnostics SNPs used to estimate *M. trossulus* fluorescence fraction; MtFF2: 17 diagnostics SNPs used to estimate the proportion of cancer cells in the 6 MtrBTN2 tumors from *M. edulis* hosts; Sequence: sequence around the SNP targeted in KASP experiment, alleles are represented [alleleX/alleleY].

**Table S3: Results of additional diagnosis methods.** Pop: population. N: number of individuals. Unchim: diagnosed as non-chimeric. NMt-Chim: diagnosed as non-*trossulus* chimeric (not due to cancer). Mt-Chim: diagnosed as *trossulus* chimeric (transmissible cancer, MtrBTN2). MtFF: *M. trossulus* Fluorescence Fraction value. DN-type: cell type described in results section. NP: not performed.

**Table S4: *Mytilus trossulus* and MtrBTN COI sequences used for MP phylogenetic tree.**

**Table S5: 2TD, PUF and MtFF values for the 5907 European mussels screened.** For mussels sampled at hemolymph and mantle tissues, both associated statistics values are reported (see "Sample type" and "Tissue type" columns).

**Table S6: Results of real-time PCR of EF1 $\alpha$ -H (Yonemitsu et al. 2019).** Double-sampled: mussels sampled on hemolymph and mantle tissue; Single-sampled: mussels samples on hemolymph or gill tissue; MtFF: *M. trossulus* Fluorescence Fraction value; -: *trossulus* chimerism: positive to almost two indexes (MtFF and PUF or 2TD); +: specific amplification of EF1 $\alpha$ -H on real-time PCR; NP: not performed.

**Table S7: Inferred genotypes of the 6 MtrBTN2 tumors (hemolymph and mantle samples).** Call\_LGC: calling performed by LGC; Call\_infered: inferred genotypes (see also FigureS1).

## Supplementary figures

**Figure S1: Graphical inference of the 6 MtrBTN2 tumors genotypes at 76 SNPs. (A)** KASP fluorescence data plot where genotypes are indicated by colored

points. **(B)** Correlation plot between  $y_i$  fluorescence signal and MtFF2 (*M. trossulus* fluorescence fraction 2, proxy of the proportion of cancerous cell in the sample). *M. edulis* samples are represented in red (MtFF2 value close to 0) and *M. trossulus* samples are represented in yellow (MtFF2 value close to 1). In both panels, the 6 tumors are represented in black: triangle filled correspond to hemolymph samples and open triangle to mantle/gill samples (samples from the same individual are related by a pointed line).

**Figure S2: Proportion of unbalanced allelic Fluorescence (PUF) and two-tissue genetic**

**distance (2TD) used for genetic diagnosis. (A)** PUF (left) and 2TD (right) values for individuals sampled at two tissues. Colors correspond to genetic diagnosis: red for *trossulus* chimeric individuals (positive to MtFF, PUF and mostly 2TD), green for non-*trossulus* chimeric individuals (positive to PUF and 2TD indexes) and grey for non-chimeric individuals (negative to all indexes). The symbol corresponds to phenotypic diagnosis: triangle pointing upwards for neoplastic individuals, triangle pointing downwards for healthy individuals and a dot for not performed. Black triangles pointing upwards represent neoplastic individuals of type B which are not chimeric (negative to all indexes). **(B)** PUF values for single-sampled individuals. Colors correspond to genetic diagnosis: red for *trossulus* chimerism (positive to MtFF) and grey for non-chimeric individuals (negative to MtFF). The symbol form corresponds to additional diagnosis methods (histology, cytology or flow cytometry): triangle pointing upwards for neoplastic individuals, triangle pointing downwards for healthy individuals and a dot for not performed. **(A, B)** Note that all grey points correspond to the baseline of 2TD and PUF statistics in healthy mussels.

**Figure S3: Neighbor-joining Tree estimated on all SNPs fluorescence values ( $y_i$ ).** Color corresponds to genetic diagnosis: red for *trossulus* chimerism (MtrBTN2), green for non-*trossulus* chimerism, grey for non-chimeric individuals and black for non-chimeric individuals with neoplastic type B cells. Orange branch corresponds to in silico chimera generated by mixing one sample (“cancer”) with 9 others samples (“hosts”) at 40%, 60% or 80% (3 of each).

**Figure S4: PUF and 2TD estimation of in silico chimera.** Panels (A) and (B) correspond to PUF and 2TD values (respectively) estimated for in silico chimera in *M. edulis* « hosts », while panels (C) and (D) correspond to PUF and 2TD values (respectively) estimated for in-silico chimera in *M. galloprovincialis* « hosts ». Colored circles correspond to intraspecific (blue and orange circles) and interspecific (green and red circles) mixed genotypes, black circles corresponds to non-chimeric mussels performed of this study and typeB cancerous samples are reported by black triangle. Curves indicate the distribution of values and grey lines correspond to the maximum (continuous line) and percentile 95% (dotted line) of PUF and 2TD in non-chimeric individuals.

**Figure S5: Mussels classification as *M. edulis*, *M. galloprovincialis* or hybrid.** PCA on all SNP  $y_i$  values (VAFF). The first axis discriminates *M. edulis* (blue) from *M. galloprovincialis* (orange). We use it to infer a hybrid index and assign the European mussels to *M. edulis* (HI > 0.85), *M. galloprovincialis* (HI < 0.4) or hybrids (0.4 < HI < 0.85) genetic backgrounds.

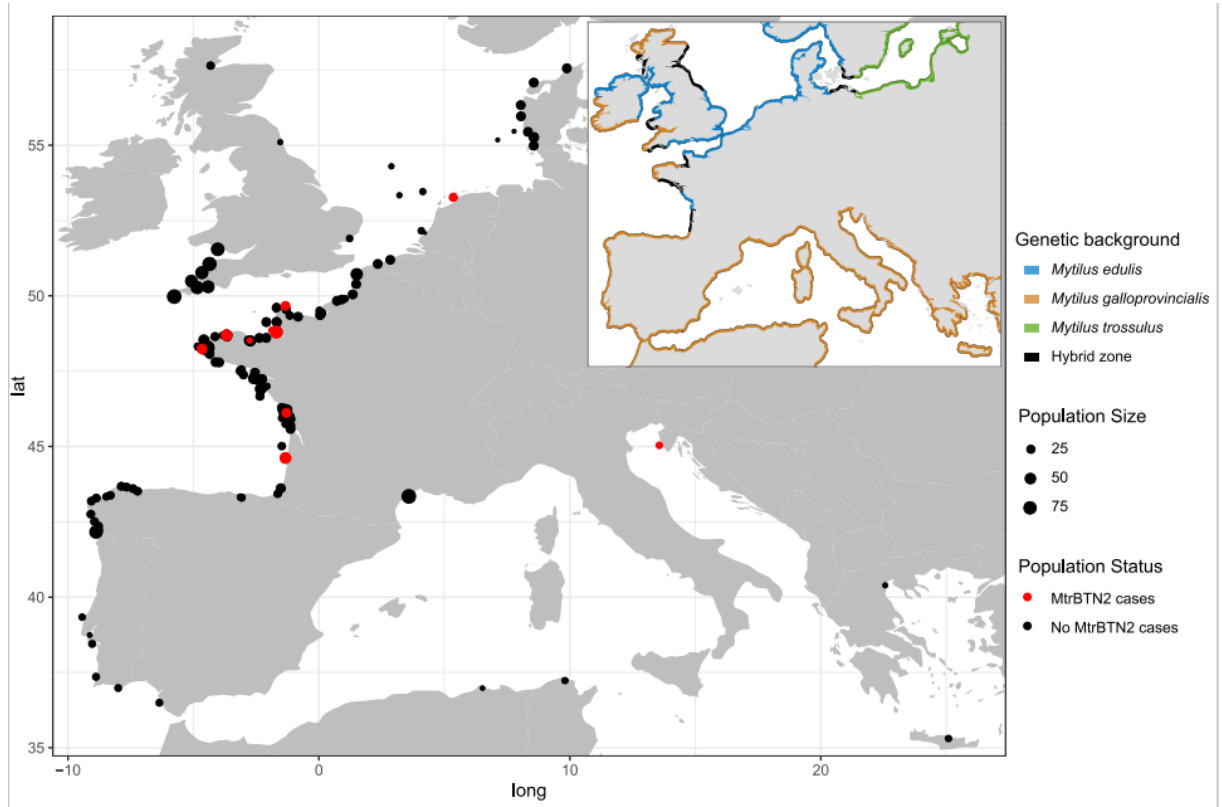
**Figure S6: Plot of genotypes for all mitochondrial and nuclear SNPs** in column and individuals in rows. Red: homozygote of *M. edulis/galloprovincialis*-state allele; yellow: homozygote of *M. trossulus*-state allele; orange: heterozygote; white: missing data. The first fiftyfour columns correspond to nuclear SNPs fixed in the 6 MtrBTN2 tumors, the next eight to SNPs fixed between the two sublineages, and the final 12 to



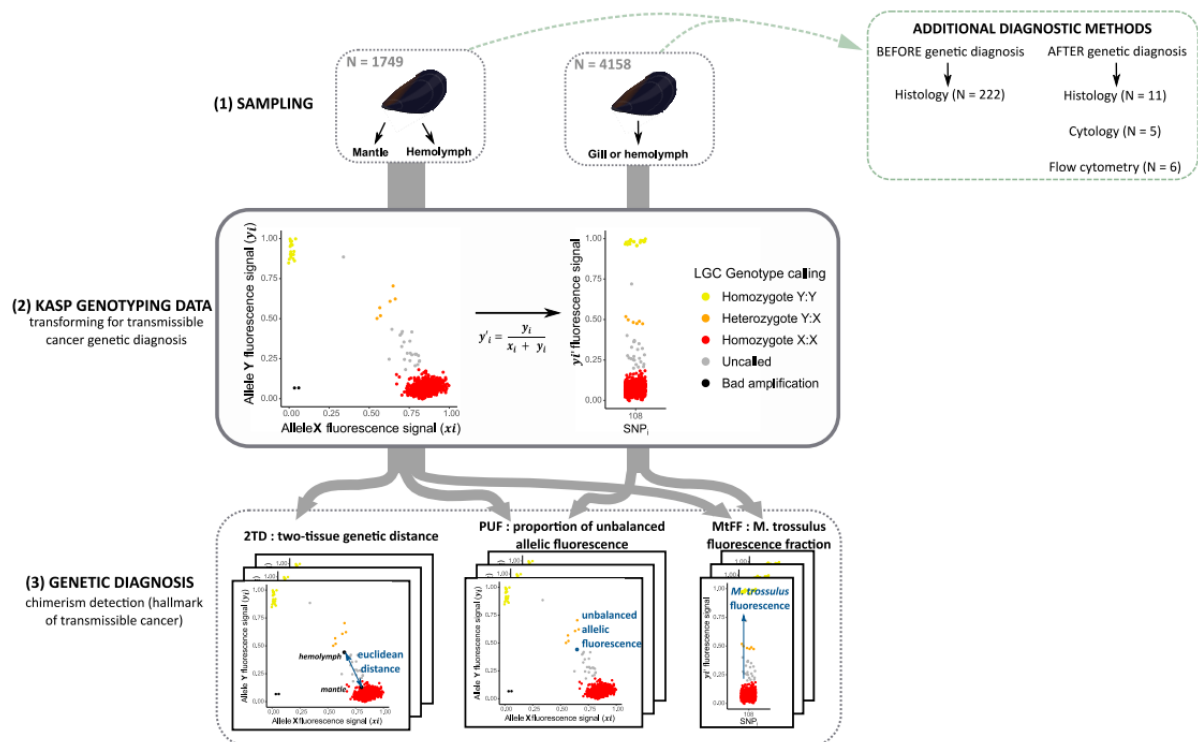
SNPs revealing polymorphism within each sublineage. Black stars highlight the five SNPs for which *M.edulis/galloprovincialis*-state alleles in tumors are not explained by the null allele hypothesis. Note that SNPs which appears homozygote for the *M. trossulus*-state allele (yellow) in both *M. edulis* and *M. trossulus* healthy mussels correspond to SNPs nearly fixed for the *M. edulis/galloprovincialis*-state allele in *M. galloprovincialis* populations (not shown here, see allele frequencies in Table S2).

**Figure S7: Fluorescence signal intensity (index of amplification,  $a_{i,j}$ ) among genotypes of tumors and healthy mussels.** Genotypes of the 20 polymorphic SNPs (referring to Figure 6 and S6) are reported in three categories: “E:E” homozygote for *M. edulis/galloprovincialis*-state allele (red); “T:T” homozygote for *M. trossulus*-state allele homozygote (yellow); “E:T” heterozygote (orange). The 6 tumors fluorescence signal for each genotypes are reported on the right while healthy mussels once are on the left (healthy mussels correspond to *M. edulis* individuals shown in Figure S6). Distribution of the fluorescence signal do not show decrease in tumors compared to healthy mussels as expected under the null allele hypothesis.

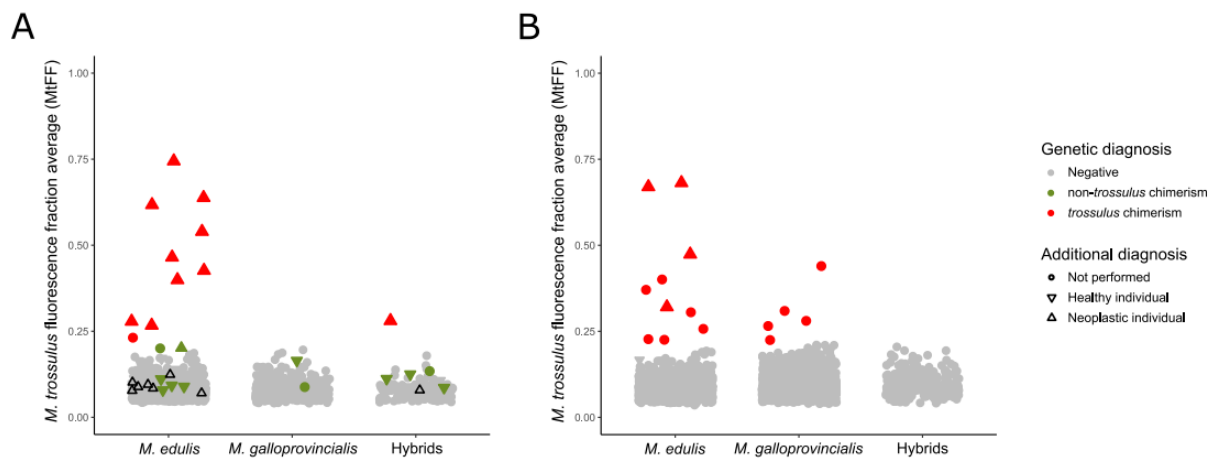
**Fig. 1: Locations of sampling sites and distribution of genetic lineages of *Mytilus* mussels in Europe.** Point size indicates sampling size and color indicates the absence (black) or presence (red) of MtrBTN2 cancer. Coastline colors represent the different genetic lineages: *M. edulis* (blue), *M. galloprovincialis* (orange), *M. trossulus* (green) and hybrid zone (black). All the information about sites is reported in Table S1.



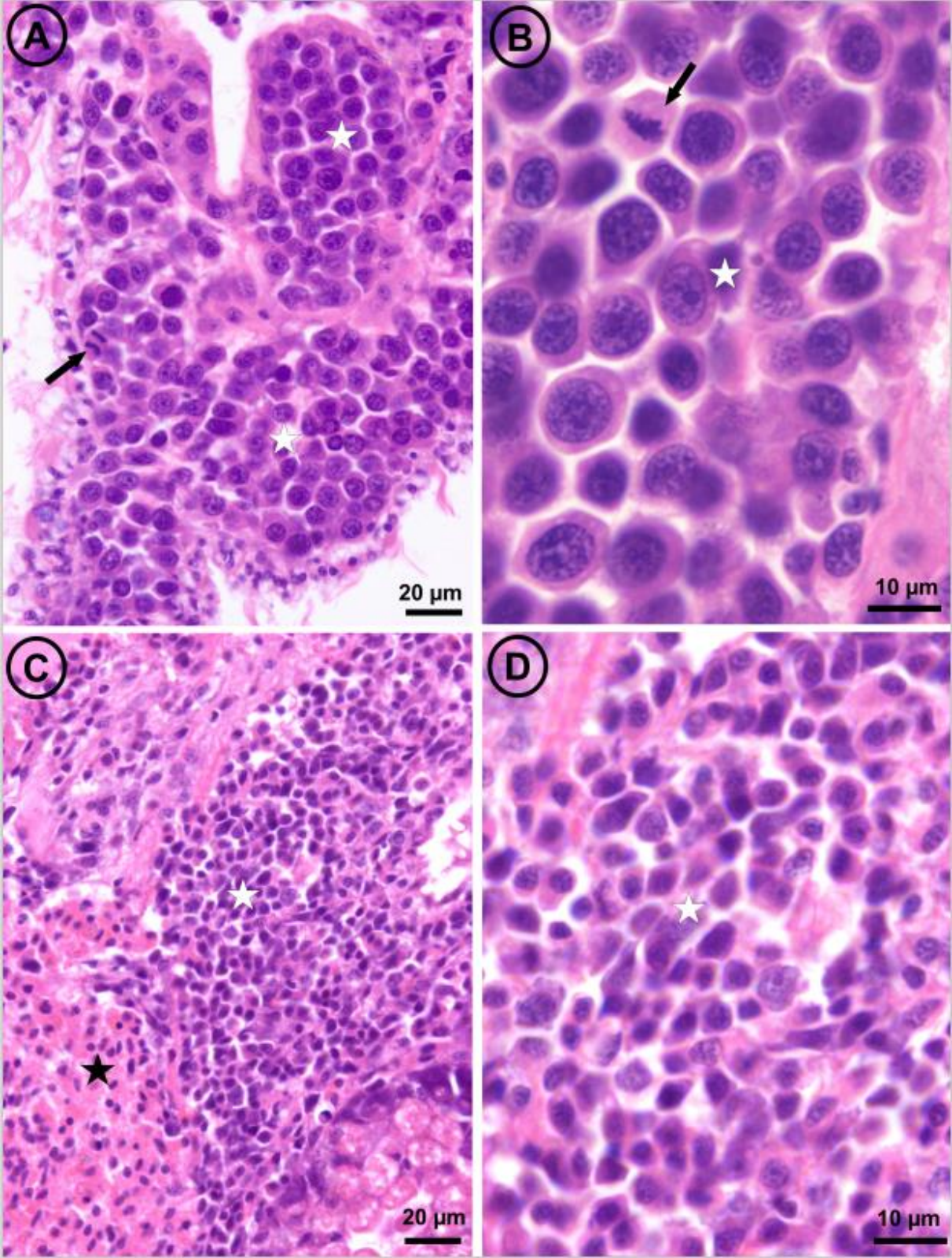
**Fig. 2: Experimental design diagram, from sampling to genetic diagnosis. (1)** Mussel sampling. 1749 individuals double sampled on hemolymph and mantle and 4158 individuals only sampled on gill or hemolymph. The additional diagnostic methods used are reported in the green box (see Table S3). **(2)** Transformation of KASP genotyping data for transmissible cancer detection. Example of KASP fluorescence data plot (left) and the transformed fluorescence value  $y_i'$  for the bi-allelic SNP “108” (right). This SNP is diagnostic and allele Y ( $y_i$ ) corresponds to the most frequent allele in *M. trossulus* reference samples (Table S2). Thus, homozygotes X:X (red) are *M. edulis*-*M. galloprovincialis* individuals while homozygotes Y:Y (yellow) are *M. trossulus* individuals. Heterozygotes Y:X (orange) are introgressed genotypes, healthy individuals with a *trossulus*-state allele at a locus due to shared polymorphism. Uncalled (grey) are *M. edulis*-*M. galloprovincialis* hosts with an excess of *M. trossulus* fluorescence likely due to *M. trossulus* transmissible cancer. Poor amplification (black) were removed from the analysis. **(3)** Genetic diagnosis, 3 indexes to detect genetic chimerism. Two-tissue genetic distance (2TD) calculated for double sampled mussels ( $n=1749$ ). Proportion of unbalanced allelic fluorescence (PUF). *M. trossulus* fluorescence fraction (MtFF) computed from 10 diagnostic SNPs.



**Fig. 3: *Mytilus trossulus* Fluorescence Fraction (MtFF) used for genetic diagnosis.** (A) MtFF values for individuals sampled at two tissues. Color correspond to genetic diagnosis: red for *trossulus* chimeric individuals (positive for MtFF, PUF and 2TD), green for non-*trossulus* chimeric individuals (positive for PUF and 2TD indexes) and grey for non-chimeric individuals (negative to all indexes). Black triangles pointing up represent neoplastic individuals of type B which are not chimeric (negative to all indexes). (B) Single-sampled individuals for which genetic diagnosis correspond to MtFF value: red for *trossulus* chimerism (positive to MtFF) and grey for nonchimeric individuals (negative to MtFF). (A, B) The symbol form corresponds to additional diagnosis methods (histology, cytology or flow cytometry): triangles pointing up for neoplastic individuals, triangles pointing down for healthy individuals and dot for not performed (see TableS3).

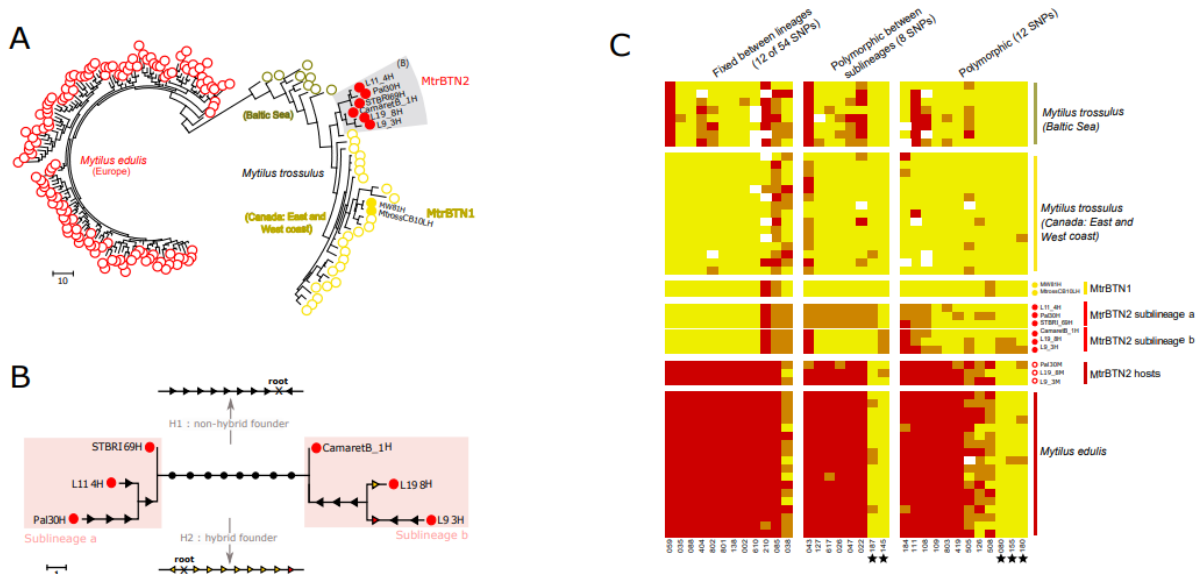


**Fig. 4: Light micrographs of histological sections through the visceral mass of two mussels, one affected by disseminated neoplasia type A (panels A and B) and the other by disseminated neoplasia type B (panels C and D). The arrows point out mitotic figures; white stars mark masses of neoplastic cells; the black star marks a mass of normal hemocytes.**





**Fig. 5: MtrBTN2 tumor genotypes.** (A) Maximum parsimony tree estimated from 74 nuclear SNP genotypes of MtrBTN2 tumors (red filled circles), *M. edulis* individuals (red open circles), MtrBTN1 tumors (yellow filled circles), *M. trossulus* individuals from British Columbia (yellow open circles) and Batic Sea (dark yellow open circles). (B) Unrooted subtree of the 6 MtrBTN2 tumors. Mutations are reported and represent gain of heterozygosity (black arrow), loss of heterozygosity (colored arrow, yellow for mutation toward a *trossulus*-state allele and red toward an *edulis*-state allele). The founder (common ancestor) genotype being unknown, the 8 mutations separating the two lines (black circles) cannot be oriented without assumptions. We present two hypotheses: a non-hybrid founder (H1, all mutations are gain of heterozygosity) and a hybrid founder (H2, all mutations are loss of heterozygosity but mostly to *trossulus*-state allele). (C) Plot of genotypes for a subset of SNPs in columns and individuals in rows. Red: homozygote of *edulis*-state allele; yellow: homozygote of *trossulus*-state allele; orange: heterozygote. Individuals' origins are reported by circle type and color (see (A)) and only 17 of the *M. edulis* individuals are represented here. The first two columns correspond to mitochondrial SNPs, the twelve next to a subset of nuclear SNPs fixed in the 6 MtrBTN2 tumors, then eight SNPs fixed between the two sublineages and then 12 SNPs revealing polymorphism within each sublineages. Red arrowhead pointed the five SNPs with *M. edulis*-state allele excess in the 6 tumors. Black star highlights the five SNPs for which *M.edulis*-state allele in tumors are not explained by the null allele hypothesis. All individuals and SNPs are detailed in Figure S5.



**Fig 6: Phylogenetic analysis of mtCOI sequences.** **(A)** Maximum parsimony tree on mtCOI 603bp aligned sequences of our 5 MtrBTN2 tumor samples (red filled circles, L9\_3, CamaretB\_1, L19\_8, L11\_4, STBRI69) and all available *M. trossulus* (GenBank) including three MtrBTN1 tumor samples (yellow filled circles), seven MtrBTN2 tumor samples of which four are from *M. trossulus* hosts (Skazina et al. 2020) and three are from *M. chilensis* hosts (red filled circles), a 100 000 years old (Tx101A, blue open circles) and contemporary *M. trossulus* mussels. **(B)** Subtree of MtrBTN2 mtCOI haplotypes with the two closest relatives from healthy mussels. Synonymous and non-synonymous mutations are reported with slashed traits and stars respectively. Host species and geographical origins are indicated on the right side of the tree.

