

# Factors regulating distribution of *Mytilus trossulus* and *M. edulis* in the Kandalaksha Bay of the White Sea

## Introduction

Blue mussels (genera *Mytilus*) being powerful ecosystem engineers (Günther, 1996; Commito & Rusignuolo, 2000; Arribas et al., 2014; Khalaman et al. 2021) play an important role in structuring coastal communities in the seas of the Northern Hemisphere (Buschbaum et al. 2009). Dense assemblages of these mollusks create habitats for a number of animals, plants and microorganisms inhabiting the biogenic environment created by mussels (Tsuchia & Nishihira, 1985, 1986; Dittman 1990; Khaitov 2013; Khalaman et al 2021 +++). On the other hand, mussels play a crucial role as a food for numerous consumers such as dog whelks (Hunt, Sheibling 1998), crabs (Chakravarti, Cotton, 2013; Lowen et al, 2013; Wójcik et al. 2015 ), starfish (Nauen, 1978; Dare 1982; Kautsky et al., 1990; Seed, Suchanek, 1992; Gaymer et al., 2001; Saier 2001 ), fish (Lappalainen et al, 2005), birds (Norton-Griffiths, 1967; Beukema et al., 2010; Ens, Goss-Custardt, 1984), and humans, both ancient (Colonese et al., 2011) and modern (Dias et al., 2011).

In recent decades one of the significant challenges in the investigation of the ecological role of mussels is their taxonomic heterogeneity. Thus, three closely related, cryptic, species *M. galloprovincialis*, *M. edulis* (thereafter ME) and *M. trossulus* (MT) can form mixed settlements in North Atlantic seas (Varvio et al., 1988; Hilbish et al., 2000; Mathiesen et al. ++++; Brooks, Farnen, 2013; Riginos, Cunningham, 2005; Katolikova et al. 2016; Vainola, Strelkov 2011; Wenne et al. 2020). The genetic effects occurring in such contact zones are well described (Bierne et al., 2002; Bierne et al., 2003; Simon et al., 2019). These cryptic species keepin their relative genetic isolation (hybrides are rare) but are involved in limited introgressive hybridization (Michalek et al 2016; Simon et al, 2019

23 and references therein). Introgressive hybridization leads to a redistribution of alleles between species  
24 genofonds (+++) and as a result to shifts of some morphological traits which could be important from  
25 an ecological point of view (e.g. shell thickness, see Michalek et al., 2021).

26 At the same time, it is unclear whether the ecological role of cryptic mussel species differs when  
27 they form mixed populations. In some cases, ecologists turn a blind eye to the presence of several  
28 mussel species, continuing to believe that a single species is represented in their region (Khalaman et  
29 al. 2021), although genetic investigations revealed two different species (Katolikova et al., 2016).  
30 Another approach is to designate the cryptic species complex as “*Mytilus* spp.” (Leopold et al 2019).  
31 Both approaches have the potential to coarsen results and lose some ecologically important patterns  
32 (Beyer et al. 2017) because different mussel species interact differently with both natural abiotic factors  
33 (see below) and pollutants (Lobel et al., 1990; Beyer et al., 2017). Clear different patterns in the  
34 interactions of coexisting mussel species with biotic components of ecosystems were revealed. For  
35 example, predators have been shown to distinguish between MT and ME in mixed settlements,  
36 favoring one of them (Khaitov et al. 2018). Different mussel species, in turn, show a different arsenal  
37 of defenses against predators (Lowen et al., 2013). The most striking difference in the relationship of  
38 cryptic mussel species with high-level consumers is their different commercial value when mussels as  
39 aquaculture target (Mallet et al., 1995; Penney et al. 2002; Beaumont et al., 2008; Dias et al., 2011).

40 In light of the above, it seems important to develop approaches to the analysis of the ecological  
41 role of mussels, taking into account their taxonomic heterogeneity. The primary task, if moving in this  
42 direction, becomes to understand what factors regulate the distribution of species: in which  
43 environment one species should be in abundance and in which the other.

In this paper, we will consider the case where ME and MT come into contact, forming mixed populations. At least seven such contact zones are currently known in the North Atlantic: West Greenland, the coast of North America from the Gulf of Maine to Hudson Bay, northeastern Scotland, the western Baltic Sea, western Norway (Vainola, Strelkov, 2011; Wenne et al. 2016; Wenne et al. 2020), the coast of the Kola Peninsula in the Barents Sea (Vainola and Strelkov, 2011; Khaitov et al. 2021), and the coast of the Kandalaksha Bay of the White Sea (Katolikova et al. 2016; Khaitov et al, 2021).

A common feature of all these contact zones is the spatial segregation of MT and ME. The proportion of MT:ME in the contact zones can change along some key gradients of the environmental parameters (Väinölä and Hvilsum 1991; Riginos and Cunningham 2005; Väinölä and Strelkov 2011) and the spatial structure of this gradient can change over time (Strelkov et al., 2017). However, some cases of mosaic distribution of MT and ME have also been described, when spots of higher concentrations of one or other species are present within the region (Moreau et al., 2006; Katolikova et al., 2016; Wenne et al. 2020). The spatial scale of this patchiness can range from several kilometers (Moreau et al., 2005; Moreau et al., 2006 report) to a few meters (Katolikova et al., 2016).

On a biogeographic scale, the distribution of *Mytilus* species is thought to be regulated by regional temperature and its correlates (Popovic & Riginos, 2019; Wenne et al., 2020). MT is believed to be a more northerly species, more tolerant to cold waters inhabiting areas that once existed in glaciation zones. ME potentially can distribute in cold waters but tend to form settlements in more temperate latitudes (Wenne et al., 2020). At the same time there is no clear latitudinal gradient in the distribution of MT and ME: in the more southern regions, the Baltic Sea, MT-dominated settlements are observed (Vainola and Strelkov, 2011; Kiewsky et al. 2019), and in the high Arctic areas, Svalbard, mussel settlements formed by ME (Wenne et al., 2020).

67 Salinity is often considered as a key factor regulating the spatial segregation of species in local  
68 ME and MT contact zones. MT are thought to gravitate toward areas of more freshened water, and ME  
69 toward areas with higher salinity (Riginos & Cunningham, 2005; Ridgway & Naevdal, 2004; Kijewski  
70 et al., 2019; Ridgway & Naevdal, 2004). For example, MTs form virtually single-species settlements in  
71 the heavily freshened Baltic Sea (Gosling, 1992; Vainola Strelkov, 2011). A change in the dominant  
72 species occurs when crossing the Danish straits toward the North Sea with normal salinity (Kijewski et  
73 al., 2019). In other contact areas (Gulf of St. Lawrence, Moreau et al. 2005, Bergen area, Ridgway &  
74 Naevdal, 2004) a similar pattern is observed: MT settlements are represented in the most desalinated  
75 areas .

76 Wave action is also considered as an important factor separating species in space. MT has been  
77 shown to gravitate toward more turbulent habitats and ME toward quieter areas (Tam, Scrosati, 2014;  
78 Comesaña et al., 1999; Bates, Innes, 1995; Innes and Bates, 1999 Проверить в каком Innes что  
79 говорится). The pattern of species segregation by substrate has also been attempted to be linked to surf  
80 exposure (Katolikova et al, 2016): MT tended toward fuchcoid thalli, and ME toward the surface of  
81 bottom stony substrates. The authors (Katolikova et al, 2016) suggested that fucoids act as shock  
82 absorbers, mitigating the impact of surf damaging MT which possess thinner and more fragile shells  
83 (Beaumont et al, 2008; Michalek et al, 2021). Importantly, in the last case, an inverse relationship is  
84 assumed: MT avoids surf-exposed areas. However, in some studies, when the material collection  
85 design was specifically adapted to study of the surf effects, the role of this factor in the regulation of  
86 the MT:ME ratio has not been proven (Moreau et al., 2005).

87 Another explanation of the spatial segregation of MT and ME can be related to the history of  
88 formation of the contact zone. For example, it is believed that in the Barents and White Seas, where  
89 initially only ME were represented, the MT settlements appeared as a result of the introduction of this

species from the coast of North America by ships performing transatlantic traffic during the World War II (Vainola, Strelkov, 2011). According to the authors, the increased concentration of MT is observed in the port areas (Vainola and Strelkov, 2011; Katolikova et al. 2016), where this species was originally introduced.

We can see a large spectrum of factors (salinity, wave action, proximity to harbor, and substrate type) which were considered as possible agents regulating the spatial segregation of ME and MT. Moreover, some correlations that seem obvious may have a different interpretation. For example, the relationship of high abundance of MT with reduced salinity can be interpreted by another way: the most desalinated areas (river mouths) are often also the quietest areas with minimal wave impact. Port structures, similarly, are often represented in the most sheltered and frequently desalinated areas. Thus, there is no strict evidence for the role of one or another factor in the regulation of species distribution.

The only way to solve this problem is to include all predictors potentially affecting the composition of mixed settlements in one regression model. Such studies have been attempted before (Kijewski et al., 2019; Wenne et al., 2020) and allowed to show a clear relationship of the MT:ME ratio primarily with salinity and temperature. However, in the mentioned works, the populations involved in the analysis were scattered over a very wide geographic range, with distances between them reaching hundreds and thousands of kilometers. On the other hand, the gradients of the key factors can vary on much smaller spatial scales. For example, salinity, as a prime candidate regulating the MT:ME ratio, can vary significantly on scales of hundreds of meters or a few kilometers in small river estuaries (+++). Does this mean that the MT:ME ratio will show a relationship with salinity at these scales as well? Importantly, other candidate factors (connection to ports, surf level, substrate character) may vary at even smaller spatial scales.

112       The contact zone of MT and ME found in the relatively small Kandalaksha Bay of the White Sea  
113 (Katolikova et al., 2016) is extremely convenient for investigation of the role of one or another factor in  
114 the regulation of the MT:ME ratio. Firstly, salinity gradients are well expressed here. The salinity  
115 variation is determined by the inflow of several large rivers into the sea (+++++ Katolikova et al.,  
116 2016). Secondly, the coast of the gulf possesses numerous sheltered areas (numerous bays and skerries)  
117 and areas open to the surf. Third, there are several active and several abandoned ports (Fig. +). Fourth,  
118 the fucoid belt is almost ubiquitous on the intertidal zone of the bay, mussels settle both on algae and  
119 on the ground surface throughout the bay (Katolikova et al., 2016), which allows us to assess the role  
120 of substrate in the regulation of microscale species segregation. Fifth, the small size of the Kandalaksha  
121 Bay (Fig. 1) ensures that there are no clear temperature gradients in both warm and cold seasons  
122 (++++). Therefore, the influence of temperature, which probably regulates the distribution of *Mytilus*  
123 species on a broader geographic scale (Wenne et al. 2020) or in areas where clear water temperature  
124 gradients exist (Hayhurst & Rawson, 2009), can be excluded from the discussion.

125       Finally, it was shown that in the Kandalaksha Bay genetic manipulations aimed at species  
126 identification can be replaced by a simple “morphotype test” based on the study of an easy to observe  
127 conchological trait (Khaitov et al. 2021). The results of this test allow identifying mussel species with  
128 sufficient accuracy (Khaitov et al. 2021). This makes it possible to process large material amount  
129 (hundreds of samples and thousands of mussels), which is necessary for revealing ecological patterns  
130 and which could not be processed using genetic technologies of species identification, the collections  
131 for which are limited as a rule to units of sites with only a few dozens of individuals involved into  
132 processing (Moreau et al., 2005).

133        In this paper, based on the large material collected in the Kandalaksha Bay of the White Sea, we  
134        will conduct a regression analysis including all the above candidate factors into one model and estimate  
135        the strength of their influence on the ratio of abundance of the two mussel species.

## 136        *Material and methods*

### 137        Field collections

138        The material was sampled on the intertidal of the mainland and the islands of the Kandalaksha  
139        Bay of the White Sea (Fig. ++). Collections were carried out in July-August in 2011-2018 at 95 sites.  
140        At each site, three bundles of fucoids were randomly collected inside of the fucoid belt and considered  
141        as three different samples, which will further be denoted as “Algae” samples. In addition three samples  
142        from the bottom surface (thereafter “Bottom” samples) were collected. For the “Bottom” sampling we  
143        chose mussels patches on the stones surface or on the muddy-sand bottom within the very same fucoid  
144        belt. A 1/40-square-meter square frame was placed on these patches, and all mollusks were sampled  
145        from the frame area. The distance between sampling points (both “Algae” and “Bottom” ) within one  
146        site did not exceed 20 m. Geographic coordinates were determined for each site (i.e., the same  
147        coordinates were assigned to each sample within the same site). Each sample was placed in a plastic  
148        bag and transported to the laboratory.

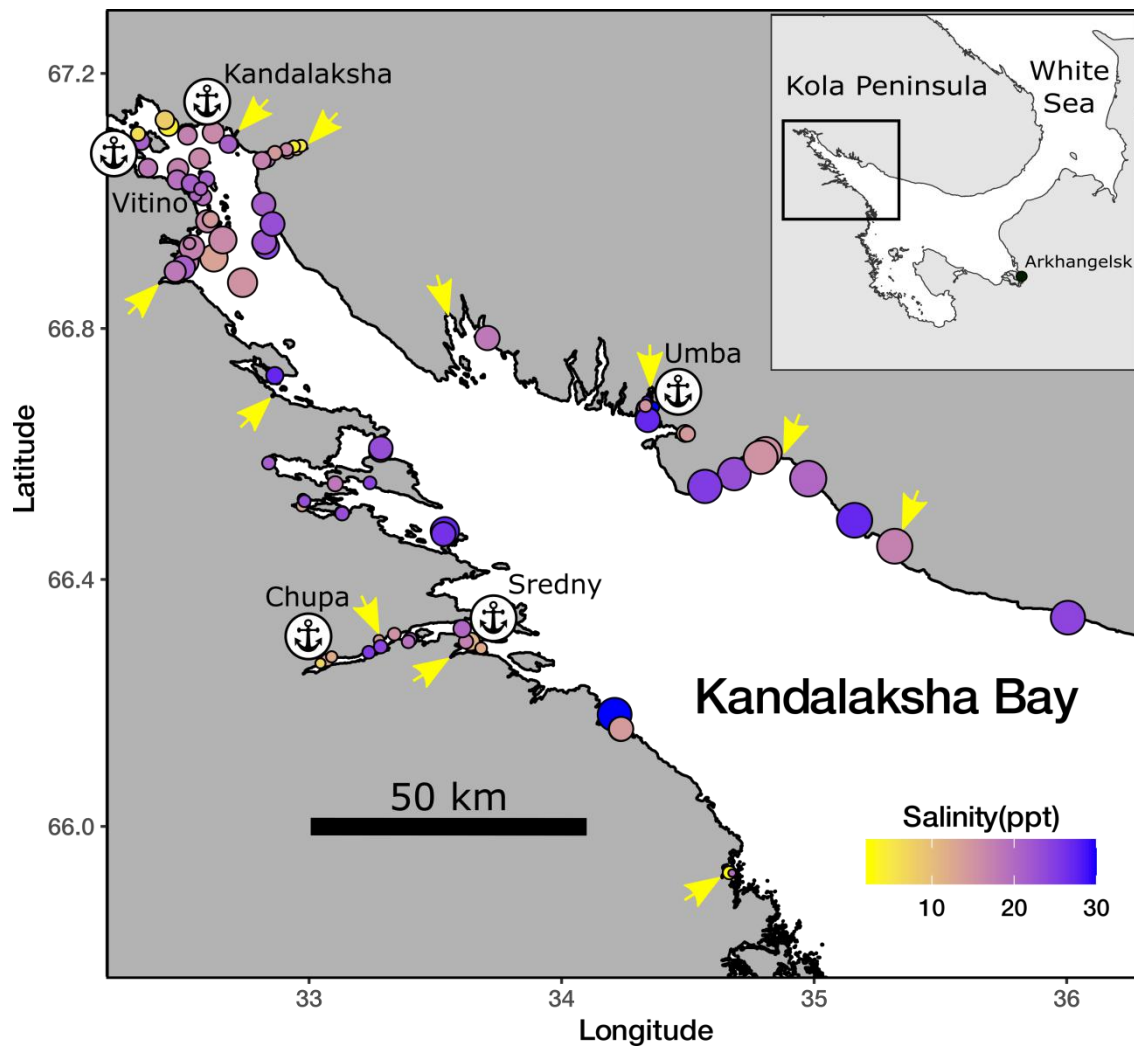


Fig. 1 Sampling sites. The size of points is proportional to average wind fetch. The fill of points represent salinity

The mussels were separated from the fucoids or washed from the sediment (sieve with a mesh diameter of 1 mm). Only mussels with a shell length greater than 10 mm were involved in further processing. These mussels were boiled, soft tissues were removed, and the shells were dried.

Analysis of the inner surface of the shell allowed us to assign each mussel to one of two morphotypes (Khaitov et al., 2021). Mussels in which a band of the prismatic layer was visible below the ligament nympha (Zolotarev & Shurova, 1997; Kepel & Ozolinsh, 1992; Katolikova et al., 2016;



158 Khaitov et al., 2018; Khaitov et al., 2021) were assigned to the T-morphotype. If the nacreous layer  
159 closely adjoined the ligament nymph, such mussels were assigned to the E-morphotype (for  
160 morphotype images, see Khaitov et al., 2018; Khaitov et al., 2021). When determining the morphotype,  
161 the right valve was analyzed (in cases of doubt, both valves were inspected). According to previously  
162 published data in the study region, 74% of mussels with multilocus genotypes dominated by *M.*  
163 *trossulus* genes, possess a T-morphotype, and 96% of mussels dominated by *M. edulis* genes exhibit an  
164 E-morphotype (Katolikova et al., 2016).

165 The number of T- and E-morphotype mussels in each sample was counted and proportion of T-  
166 morphotype mussels (PT) was calculated. Additionally this proportion was assessed for each site (the  
167 total number of T-morphotype mussels collected at this site divided by total number specimens  
168 collected). The proportion of MTs (Ptros) in mixed populations in the White Sea is strongly correlated  
169 with PT. The Ptros value can be recalculated from PT using the following formula (Khaitov et al.,  
170 2021):

$$Ptros = \frac{e^{-2.4+5.4PT}}{1 + e^{-2.4+5.4PT}}$$

171 Thereafter we assume the PT as a proxy for MT:ME ratio.

172 In total we processed material from 570 samples, from which 55500 specimens were inspected  
173 using morphotype-test. The mean number of mussels collected per sample was 97 (SD = 97, minimal  
174 mussel number 2, maximal amount 785).

175 Assessment of environmental parameters

176       The focus of this work was to examine the effects of substrate, salinity, surf level, and proximity  
177 to port. Some of these parameters were measured directly but for others several proxies were  
178 considered.

179       The nature of the substrate (Algae vs Bottom) was characterized by where the corresponding  
180 sample was taken from. The degree of proximity to port was estimated as the straight line distance (not  
181 accounting shoreline bends) from the sampling site to the nearest of the 5 large ports presented in the  
182 area (Fig++, Table ++). Additionally each port was assigned to one of two categories: “Active” if at the  
183 time of sampling the port was serving heavy-tonnage vessels plying out of the White Sea, or  
184 “Abandoned” if the port served only small vessels cruising within the White Sea, or was completely  
185 abandoned at the time of sampling. Earlier all abandoned ports (at least during World War II, before it  
186 or immediately after it) served heavy-tonnage vessels plying between the White Sea and the Barents  
187 Sea.

188       Table +. Ports presented in the Kandalaksha bay

Port	Status	Latitude	Longitude
Kandalaksha	Active	67.13728	32.40799
Vitino	Active	67.07657	32.33363
Umba	Abandoned	66.67797	34.35766
Chupa	Abandoned	66.26996	33.06953
Sredny	Abandoned	66.29418	33.64066

189

190       *Wind fetch as a proxy for wave impact.* When selecting sampling sites, we assumed that  
191 collections should characterize both surf and sheltered habitats. Each site was visually assigned to one  
192 of two categories: “Exposed” (sites located on open shores exposed to waves coming from the open

193 part of the bay) or “Sheltered” (sites located in bays or skerries). For a more formal assessment, the  
194 value of wind fetch was calculated for each site (Seers, 2018). It is calculated as the unobstructed  
195 length of water surface (in km) over which wind from a certain direction can blow over. These values  
196 can be obtained from the analysis of geographical maps of the region. To calculate the wind fetch value,  
197 the package “fetchR” was used (Seers, 2018). The higher the fetch from a certain direction, the more  
198 impact could potentially be brought by waves. Wind fetch was measured for each site from all  
199 directions (North, East, South, and West, Fig. ++ ) but as a value for the further analysis we used an  
200 averaged fetch value.

201 *Salinity.* Water samples were taken at each site when mussels were collected, i.e. in the time of  
202 low tide. The salinity was determined using a hand-held refractometer (++++, with 1 ppm accuracy).  
203 However, salinity measured directly at the sampling site poorly reflects the salinity regime in the site  
204 due to its fluctuations during the tidal cycle and due to weather conditions (waving, precipitation etc.).  
205 Therefore, as an additional estimate of the desalination level in a given site, we used the characteristics  
206 of fresh water discharging sources closest to the sampling site (rivers, cooling channels of hydroelectric  
207 power plants).

208 In total, 24 rivers are present in the region (Table +), information about which is present in open  
209 sources ( Государственный водный реестр : [арх. 15 октября 2013] / Минприроды России. —  
210 2009.; [http://www.mnr.gov.ru/files/part/0306\\_perechen.rar](http://www.mnr.gov.ru/files/part/0306_perechen.rar)), in addition, 2 powerful freshwater  
211 dischargers (HPP canals) were included in the analysis. For each river, we assessed the catchment basin  
212 area and coordinates of its mouth. For HPP canals and one of the rivers, this information was obtained  
213 by visual analysis of geographical maps. Each of the rivers was assigned to one of two categories:  
214 “Large” (rivers with a catchment area of at least 240 square kilometers) and “Small” (catchment area of

less than 240 square kilometers). The median catchment area of all rivers in the study region was used as the boundary between the categories (Table ++).

We calculated the distance (straight line, not accounting shoreline bends) from each sampling site to the mouth of the nearest river. In further analysis, this value was used as a continuous predictor. We considered the categorical value corresponding to the size of the nearest river as another predictor characterizing the desalinization level at the sampling site. Thus salinity was characterized by three values: directly measured salinity, distance to the nearest river and the nearest river size.

Table +. Parameters of river and other fresh-water discharging source associated with Kandalaksha bay.

River	Drainage Area sq. km	River Size	Mouth Latitude	Mouth Longitude	Number of sampling sites near the river
Chernaya	141.0	Small	66.51896	32.94966	6
Gridina	540.0	Large	65.92462	34.65846	2
Keret	3360.0	Large	66.28172	33.58034	4
Knyazhegubskaya HPP channel		Large (map analysis)	66.87178	32.40742	10
Kolvitsa	1260.0	Large	67.08320	32.99110	14
Kovda	26100.0	Large	66.69549	32.87103	1
Kuzreka	255.0	Large	66.59999	34.81303	4
Kyatka	204.0	Small	65.88838	34.68294	0
Lupche-Savino	146.0	Small	67.16291	32.34259	0
Luvenga	175.0	Small	67.10037	32.69313	11
Nilma	167.0	Small	66.50577	33.13507	8
Niva	12800.0	Large	67.13027	32.41825	9
Nivskaya HPP channel		Large (map analysis)	67.15676	32.37713	0
Olenitsa	403.0	Large	66.45276	35.30060	2
Pila	313.0	Large	66.78056	34.15745	0
Plavezhma		Small (map analysis)	66.25417	33.00065	3
Porya	240.0	Large	66.81954	33.55116	1
Pulonga	630.0	Large	66.30251	33.27729	6
Ryasanka	102.0	Small	66.73710	33.44663	0
Salnitsa	121.0	Small	66.37576	35.68487	1
Shoguy	103.0	Small	66.74761	34.02558	0
Shushpanikha	65.2	Small	66.85175	33.60705	0
Sonreka	167.0	Small	66.15655	34.22415	4
Umba	6250.0	Large	66.66461	34.29766	6
Varsuga	9840.0	Large	66.26487	36.94030	0
Vypcha	161.0	Small	66.63304	34.49202	3

## Statistical analysis

All statistical processing was performed using the statistical programming language R 4.05 (R core Team, 2021). The aim of the analysis was to determine which of the candidate-factors have an influence on the MT:ME ratio (Ptros). We used PT (a proxy for Ptros, see above) as a dependent variable in further analyses. The analysis was aimed to answer two questions. First, which predictors are statistically significantly related to the response variable (the answer to this question was treated as Step 1 of our analysis). Second, in what ratio are the strengths of the influence of each of the factors (Step 2).

The solution in Step 1 is achieved by constructing a generalized linear mixed model (GLMM, Bolker et al., 2009). We constructed a random intercept logistic GLMM with beta-binomial residuals distribution where “Site” was considered as a random factor (**Model 1**). The fixed part of the model included predictors as follows:

- Substrate (categorical predictor with two levels “Bottom” vs “Algae”);
- Salinity (continuous predictor, characterising the directly measured salinity);
- MinDistRiver (continuous predictor, characterising distance to the nearest river);
- RiverSize (categorical predictor with two levels “Small” vs “Large” );
- MinDistPort (continuous predictor, characterising distance to the nearest port);
- PortStatus (categorical predictor with two levels “Abandoned” vs “Active”).

For all categorical predictors above, the first mentioned category was used as base level. The function `glmmadmb()` from the package “glmmADMB” (Fournier et al., 2012; Skaug et al., 2016 ) was used to build this GLMM.

245 Step 2 of the analysis is aimed to estimate the contribution of each predictor to the overall  
246 variation. The total variation explained by a model is usually estimated using the coefficient of  
247 determination  $R^2$  (Quinn & Keugh, 2002; Nakagawa & Schielzeth, 2013). In the case of GLMM, the  
248 fraction of total variability explained by the fixed part of the model is estimated using marginal  $R^2$  ( see  
249 Jagger et al., 2017; 2019 (read!!!!) Nakagawa & Schielzeth, 2013). The latter can be decomposed into  
250 semi-partial coefficients of determination, also known as semi-part  $R^2$ , that describe the contribution of  
251 each individual predictor to the overall coefficient of determination (Stoffel et al., 2021). The “partR2”  
252 package (Stoffel et al., 2021) was used for this analysis. Unfortunately this package for now is adopted  
253 to use GLMM with binomial residual distribution only (Stoffel et al., 2021). That is why we  
254 constructed the additional model, **Model 2**, fixed part of which was analogous to that in **Model 1**.  
255 However, to take into account the overdispersion in **Model 2** we included in addition to “Site”, as a  
256 random factor, the observation level random effect (OLRE). OLRE is a random factor which adsorbs  
257 overdispersion in GLMM (Harrison, 2015; Harrison, 2014). The **Model 2** was constructed using glmer()  
258 function from the lme4 package (Bates et al., 2015)

## 259 *Results*

### 260 Distribution of environmental parameters

261 There are 5 ports presented in Kandalaksha Bay (Table +++, Fig ++1). Most sites (34 sites) had  
262 “Vitino” as the closest harbor, which came into service in the 1990s. The remaining ports, in operation  
263 since the beginning of the 20th century, were attributed as closest to 14-17 sites. The distance from  
264 sampling sites to the nearest ports varied in broad range (0.11 - 82 km) but on average the distance  
265 from port to the corresponding sites ranges in more narrow limits (17 - 26 km, Fig ++, A).

266 Most of the sites visually evaluated during sampling as “Sheltered” have wind fetch values lower  
267 than those assigned to the “Exposed” group (Fig. ++, B). For sites in the latter group, the most open  
268 direction is the southeastern one (Fig. ++, B), which coincides with the wind direction that prevails in  
269 the region during the warm seasons (Lebedev, 2016). The most open sites (maximum wind fetch) were  
270 located on the northwestern coast and on the open areas of the islands closer to the top of the  
271 Kandalaksha bay (Fig. ++1).

272 Salinity measured at the time of sampling ranged from 2 to 30 ppm (Median = 19 ppm). The most  
273 desalinated areas were at the top of both the Kandalaksha Bay itself and at the tops of smaller inlets  
274 deeping into the shore, where rivers often flow into (Fig. ++1). The closest to the sampling sites were  
275 the mouths of 18 rivers, of which 11 were classified as “Large” (Table ++, Fig. ++). Salinity at sites  
276 where the nearest river was classified as “Large” was markedly lower than at sites closer to small rivers  
277 (Fig. ++, C). Salinity tended to increase as the distance to the nearest river increased (Fig. ++ D).

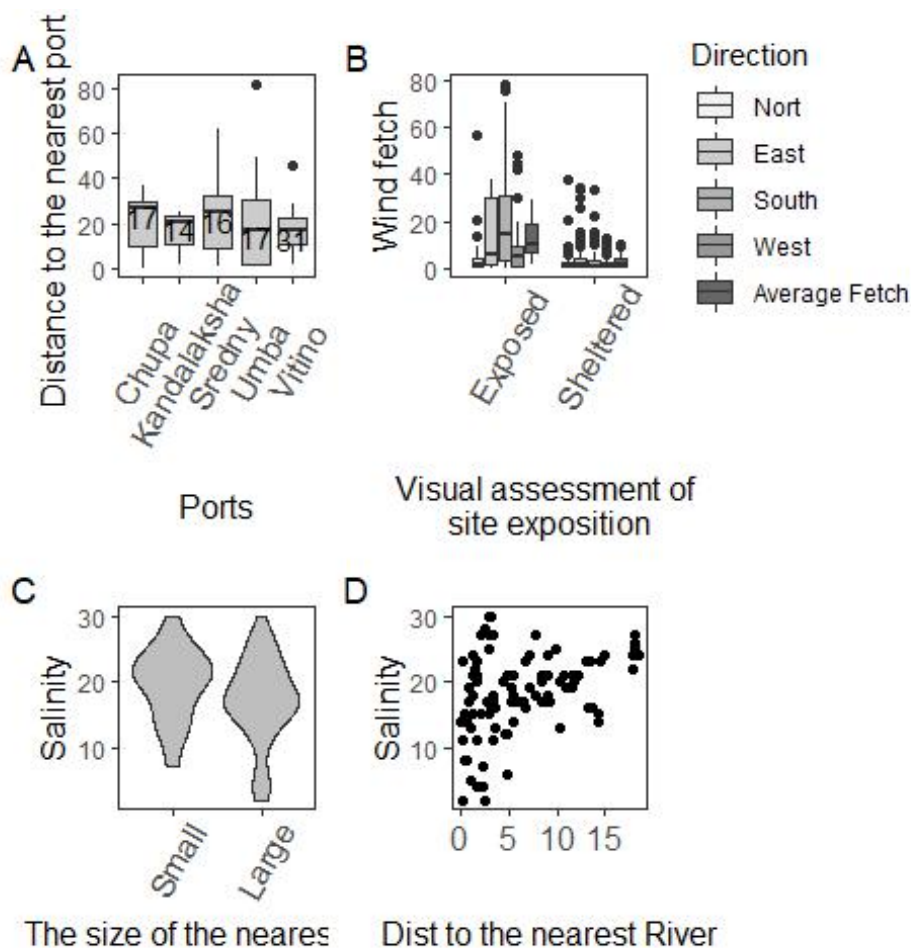


Figure ++. The distribution of values used for environmental factors assessment

### *PT distribution*

The values of PT ranged from 0 to 0.85 at different sites. The frequency distribution of PT values displayed a bimodal structure both in the case of “Bottom” and “Algae” samples (Fig. ++). Thus two types of populations can be recognized in the studied area: populations dominated by E-morphotype (conditional boundary  $PT < 0.5$ ) and samples in which mussels of T-morphotype dominated ( $PT > 0.5$ ). In collections from the “Bottom” substrate, samples belonging to the first group were more frequent (the left peak in the distribution predominated). In collections from fucoids, both groups of specimens were represented in more or less equal proportions.



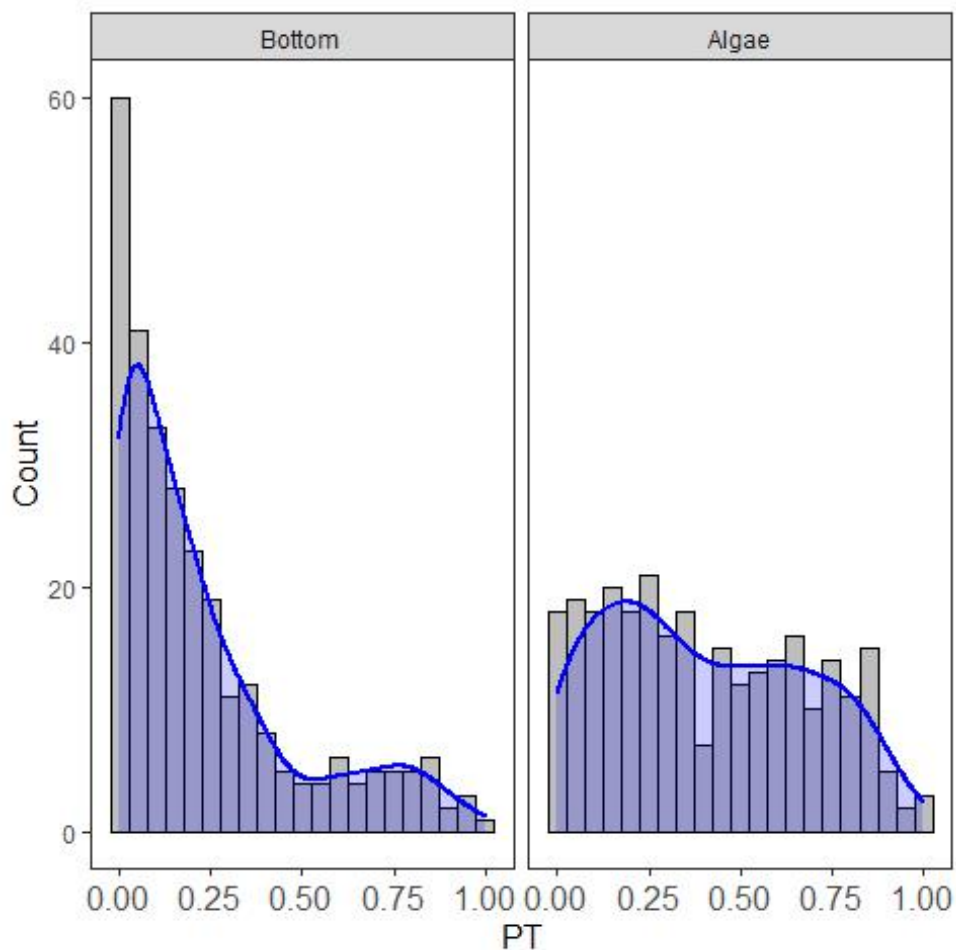


Fig ++. The frequency distributon of PT values at different substrates

The maximum proportion of the T-morphotype was observed in the top of the Kandalaksha Bay itself and in the small bays deeping into the mainland (Fig. ++). At the same time, sites with a lower proportion of the T-morphotype were located in the immediate vicinity of sites where settlements with high PT values were represented (Fig. ++), reflecting the high level of patchiness in PT spatial distribution. Settlements with the lowest frequency of T-morphotype were represented in the open areas of the Northwest Coast.

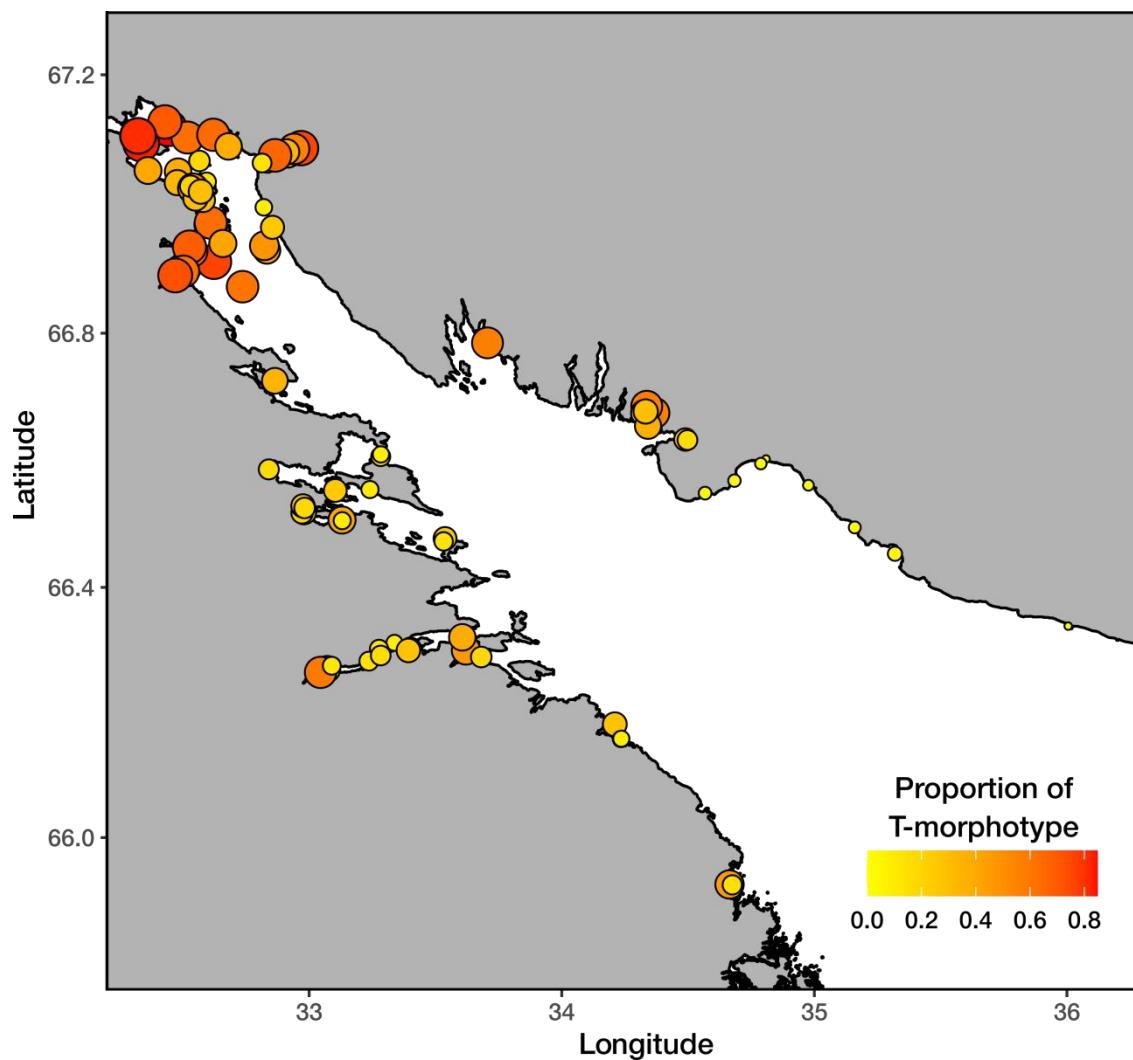
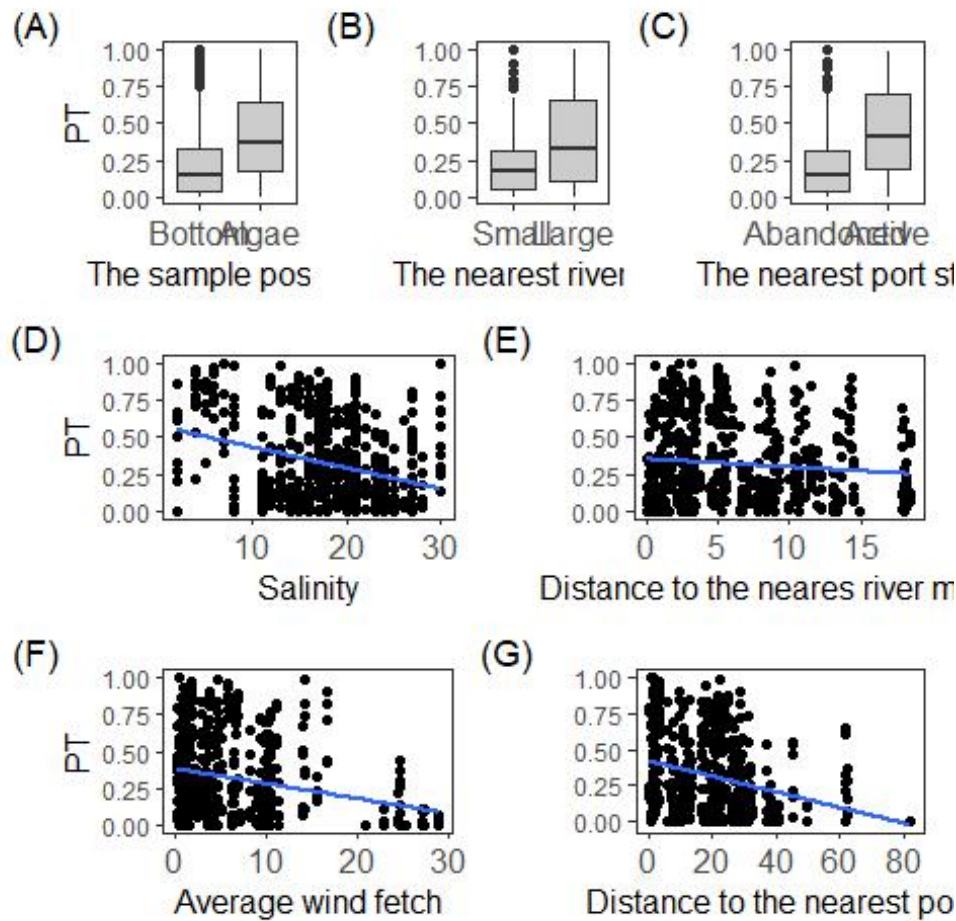


Fig. + Proportion of mussels with T-morphotype on sampling sites. The fill of points represents PT and size of proportional to PT.

#### *Relationship of PT to predictors.*

The proportion of T-morphotype mussels was, on average, higher on fucoids than on bottom substrates (Fig. ++ A). This value was slightly higher at sites for which the nearest river was a large river than at sites in the vicinity of small ones (Fig. ++ B). PT at sites for which the nearest port was active was, on average, higher than at sites closer to abandoned ports (Fig. ++ C). The studied value

304 tended to decrease as Salinity, Average wind fetch, and Distance to the nearest port increased (Figs. ++  
 305 D, F, G, respectively). No clear association of PT with distance to the nearest river mouth was revealed  
 306 (Fig. ++, E).



307  
 308 Figure 1. Proportion of mussels with T-morphotype in samples from different conditions. (A) PT  
 309 on different substrate types; (B) PT at sites adjacent to rivers of different sizes. (C) PT in areas around  
 310 abandoned and active ports. (D-G) PT as a function of Salinity, Distance to mouth of nearest rivers,  
 311 average wind fetch and distance to the nearest port. Blue line represent ordinary least square regression  
 312 fitted, which represent only tendency of association in raw data. This line should not be considered as

313 evidence of correlatin since regression line was fitted without taking into account of presence of other  
 314 predictors (this is the reason why confidential intervals are not presented at figures).

315 According to Model 1 (Table ++) significant relationships were found between PT and the  
 316 substrate type, average wind fetch, distance to the nearest port, and the status of the nearest port.  
 317 According to positive coefficients (Table ++) PT is higher on “Algae” in comparison with “Bottom”  
 318 substrates and at sites for which the nearest port belongs to “Active” group in contrast to “Abandoned”  
 319 one. The negative slope for “MinDistPort” and “WindFetch” (Table ++) indicates that PT decreased  
 320 when distance to the nearest port increased and potential wave impacts were enforced.

321 Model 2 which was used for assessment of predictors influencing power shows the similar  
 322 pattern of association between PT and predictors as was revealed in Model 1 (estimated coefficients for  
 323 Model 2, Table ++, are close to those for Model 1). Partial  $R^2$  for Model 1 was quite low (part  $R^2 =$   
 324 0.21, Table ++, Fig. ++) indicating that none of the predictors included in the model unambiguously  
 325 explained most of the variation.

326 Table +. Parameters of the GLMM fitted

effect	group	term	estimate	std.error	statistic	p.value
Model 1						
fixed		(Intercept)	-1.20	0.385	-3.121	0.0018
fixed		Substrate <sub>Algae</sub>	1.07	0.070	15.224	<0.0001
fixed		Salinity	-0.02	0.018	-1.044	0.2965
fixed		MinDistRiver	-0.03	0.023	-1.473	0.1409
fixed		RiverSize <sub>Large</sub>	0.45	0.205	2.199	0.0279
fixed		AverageFetch	-0.05	0.016	-3.035	0.0024
fixed		MinDistPort	-0.02	0.007	-2.630	0.0085
fixed		PortStatus <sub>Active</sub>	1.18	0.232	5.100	<0.0001
random	Site	SD <sub>(Intercept)</sub>	0.83			
Model 2						
fixed		(Intercept)	-1.40	0.456	-3.072	0.0021
fixed		Substrate <sub>Algae</sub>	1.24	0.081	15.361	<0.0001
fixed		Salinity	-0.02	0.021	-0.768	0.4426
fixed		MinDistRiver	-0.04	0.028	-1.460	0.1442
fixed		RiverSize <sub>Large</sub>	0.45	0.243	1.865	0.0622
fixed		AverageFetch	-0.07	0.018	-3.683	0.0002

fixed		<i>MinDistPort</i>	-0.02	0.008	-2.528	0.0115
fixed		<i>PortStatus<sub>Active</sub></i>	1.34	0.275	4.867	<0.0001
random	OLRE	<i>SD<sub>(Intercept)</sub></i>	0.83			
random	Site	<i>SD<sub>(Intercept)</sub></i>	0.99			

327

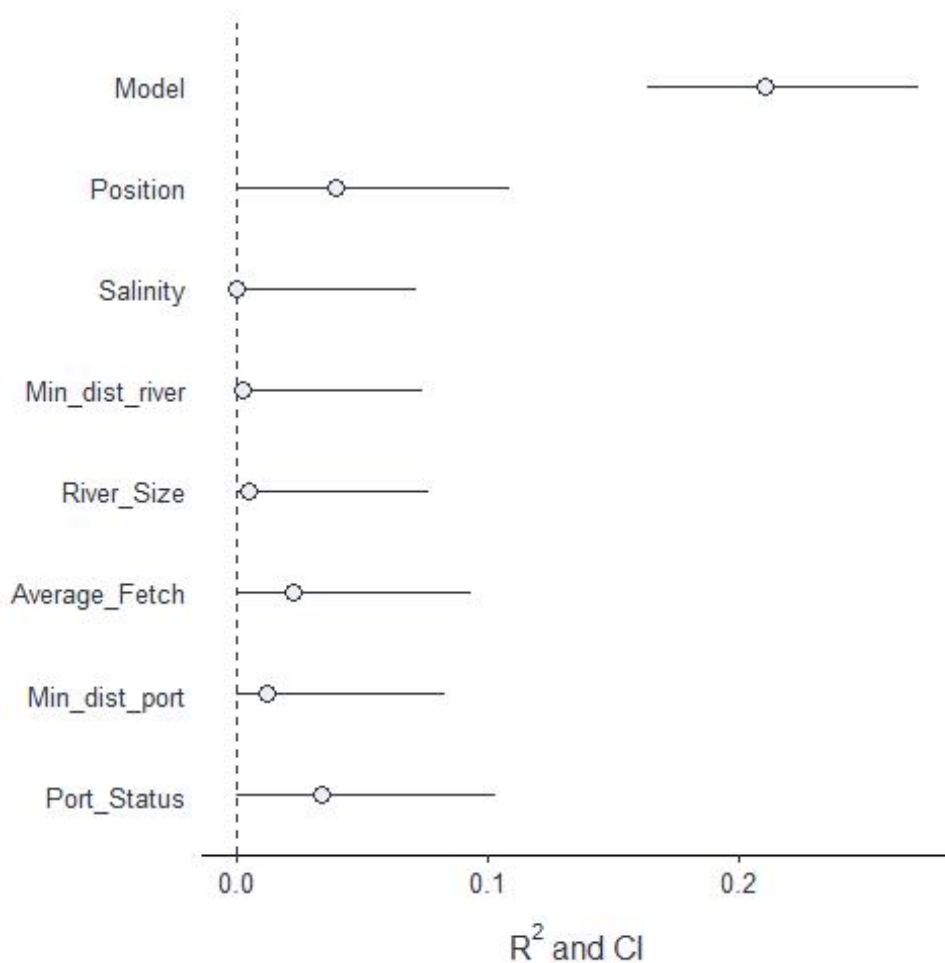
328

329

330           However, the semi-partial  $R^2$  indicates that the best predictors for PT is substrate type (semi-part  
331  $R^2 = 0.040$ ), followed by nearest port status (0.034), Average wind fetch (0.023), and distance to  
332 nearest port (0.013). The signal strength from all predictors characterizing salinity was in order of  
333 magnitude lower.

334           Table +. Partial  $R^2$  for the full model and semi-partial  $R^2$  with bootstrap CI for predictors from  
335 the Model 1.

Term	Estimate	Bootstap_CI	df
<i>FullModel</i>	0.211	0.164 : 0.271	8
<i>Substrate</i>	0.040	0 : 0.109	7
<i>Salinity</i>	0.000	0 : 0.072	7
<i>MinDistRiver</i>	0.003	0 : 0.075	7
<i>RiverSize</i>	0.005	0 : 0.077	7
<i>AverageFetch</i>	0.023	0 : 0.093	7
<i>MinDistPort</i>	0.013	0 : 0.084	7
<i>PortStatus</i>	0.034	0 : 0.104	7



336

337

338

Figure ++. The importance of predictors as drivers of MT distribution. Partial  $R^2$  reflecting the proportion of total variation explained.

339

### *Discussion*

340

341

342

343

344

The application of the morphotype test (Khaitov et al., 2021), on which this study is based, leaves room for doubt that our results reflect the real MT:ME ratio. However, on the other hand, the high correlation between PtroS and PT (Khaitov et al., 2021), as well as the very large volume of material (numerous sites and a large number of individuals collected on them) allow us to hope that the obtained pattern correlates with the real pattern of species distribution.

345 An intriguing result of our study is the very weak signal of the influence of salinity on the  
346 MT:ME ratio in mixed settlements. GLMM fitted do not reveal a statistically significant relationship  
347 between PT and salinity, per se, measured at the time of sampling. The influence power of this  
348 predictor was extremely low. We attempted to estimate the degree of desalination indirectly by  
349 introducing into the model quantities that may alternatively characterize salinity at the sampling site:  
350 the distance to the nearest river and the size of the nearest river. Both of these quantities, like salinity,  
351 contributed negligibly to the total variance explained by the model. This result is in contradiction with  
352 the fact that many papers analyzing MT distribution in MT-ME contact zones indicate that MT setting  
353 populations to more desalinated habitats and MT gene frequencies decrease as salinity increases  
354 (Riginos and Cunningham, 2005; Katolikova et al., 2016; Stuckas et al., 2017; Kijewski et al 2019).

355 Surprisingly, in spite of the fact that salinity is considered as a leading factor determining the  
356 segregation of MT and ME (+++) all studies discussing this relationship are correlative. Only a few  
357 works have been devoted to comparison of ecophysiological responses of these two species to salinity  
358 (Gardner&Thompson 2001; Qiu et al., 2002; Knobel et al. 2021), although this is the very kind of data  
359 that should be considered as evidence of causal relations.

360 Moreover, direct investigations tended to influence salinity analysis, did not reveal unequivocal  
361 evidence that adult MT stages are better adapted to reduced salinity than adult ME  
362 (Gardner&Thompson 2001; Qiu et al., 2002). It has only been shown that larval stages of MT survive  
363 lower salinity better than the same ME stages (Qiu et al., 2002). That is, MT larvae are better  
364 performed in low salinity corresponding to those conditions in which adults exist (Knobel et al. 2021).

365 Looking at the raw data presented in our work, as in previous studies, one gets the impression  
366 that salinity is an important environmental parameter (Fig. ++ shows a clear negative correlation of PT

367 with salinity). However more accurate analysis, where salinity influence is considered against the  
368 influence of other predictors included in a common model, is rather consistent with those  
369 ecophysiological works in which no explicit correlation between physiological indices of MT and ME  
370 and salinity was found. Thus, the correlation between salinity and the MT:ME ratio, noted on raw data,  
371 at least in our work, allows us to consider the above relationship as a spurious correlation. At the same  
372 time, it is apparently too early to completely abandon the regulatory role of salinity. A model based on  
373 beta-binomial distribution (**Model 1**) allowed us to capture weak but significant signal about the role of  
374 salinity: if we express salinity in terms of the size of the nearest river then PT turns out to be higher on  
375 average if the nearest river is large.

376 We admit that our methods of salinity determination, the way of sampling, the use of morphotype  
377 test instead of direct genotyping, all this does not allow us to reveal the fine relationship between the  
378 MT:ME ratio and salinity. However, if such clear links with salinity, which have been revealed in other  
379 regions, exist also in the Kandalaksha Bay of the White Sea, where salinity gradients are very  
380 pronounced, then these patterns should have been reflected in the distribution of morphotypes (which  
381 are very tightly associated with the species status of mussels) as well. Perhaps we should look for the  
382 answer in other influences that shift the observed pattern. For example, it has been shown that only  
383 larvae (Qiu et al., 20++) show a clear dependence on salinity. If the critical stage for the formation of  
384 the MT:ME ratio relationship with salinity is larvae, then the ratio of larval numbers of the two species  
385 in the plankton will play a crucial role. We currently have no data on this ratio in the White Sea.  
386 However, it is logical to assume that the number of larvae will be proportional to the abundance of  
387 adults living on the bottom. If we consider only bottom substrates, the relative frequency of ME-  
388 dominated settlements (low PT values, Fig. ++) greatly exceeds the frequency of MT-dominated  
389 settlements (high PT values, Fig. ++). In the case of fucoid settlements, this ratio is close to equal.



390 Mussel settlements existing on the bottom substrates in the White Sea are very extensive and are  
391 represented by numerous extensive littoral and sublittoral mussel beds located on gravel or muddy-sand  
392 substrates (Lukanin +++; Khaitov, +++; Khaitov&Lentsman, 2013). The furoid belt is only a narrow  
393 band along the lower part of the intertidal zone (+++). These allow us to expect that in the plankton the  
394 absolute advantage should be with the ME. It is possible that with such ratios of species abundance in  
395 plankton and with a relatively small area of the Kandalaksha Bay, the salinity driven selection  
396 discussed in previous studies (Knobel et al., 2021) cannot manifest itself.

397         Against the background of the unexpected failure in identifying the relationship between the  
398 MT:ME ratio and salinity, we found other, previously poorly studied, correlations. We found that the  
399 proportion of MT is clearly governed by three factors: the nature of the substrate, the proximity of the  
400 mixed settlement to active port systems, and the degree to which the coast is open to wave action.  
401 According to the data we obtained, the highest MT concentration should be expected on the furoid  
402 tally, in sheltered areas close to active ports.

403         It is believed that MT was introduced into the Barents and White Seas due to ship traffic during  
404 World War II (Vainola & Strelkov, 2011). This hypothesis was based primarily on the fact that high  
405 concentrations of MT were found in the vicinity of currently active or abandoned ports in the Barents  
406 and White Seas (Vainola & Strelkov, 2011; Katolikova et al. 2016). Our study confirmed this  
407 hypothesis. The proportion of T-morphotype (and hence MT) in the studied area decreases with the  
408 distance from the port (negative coefficients with the predictor MinDistPort in Models 1 and 2, Table  
409 ++).

410         Apparently, the degree of port activity also plays an important role (the predictor “PortStatus” is  
411 among the leaders by the strength of influence on the explained variance, Table ++, Fig. ++).

412 According to our data, if the port nearest to the mixed settlement is active, i.e. it serves the ships  
413 coming from outside the White Sea (in our case it is “Kandlaksha” and “Vitino” ports), then the PT  
414 value in such settlements is higher than in the settlements adjacent to the abandoned harbors. This can  
415 be explained in several ways.

416 It is possible that new MTs are constantly brought to the ports of the Kandalaksha Bay from the  
417 Barents Sea (primarily from the ports of Murmansk in the Kola Bay, where the abundance of MT is  
418 very high, Vainola&Strelrov, 2013). In the abandoned ports this constant renewal is absent and  
419 during the time of port inactivity MTs have been displaced by MEs.

420 On the other hand, active ports tend to have a higher degree of anthropogenic pollution, leading  
421 to less stress tolerance of mussels (Hellou & Law, 2003). Different mussel species are known to  
422 interact differently with pollutants (Beyer et al., 2017) and in the Halifax area, for example, MT are  
423 more abundant in the most polluted habitats than ME, but show less stress tolerance than ME (Hellou  
424 & Law, 2003). However, studies of the immune systems of MT and ME in the Canadian contact zone  
425 suggest a greater suppression of the MT immune system in areas prone to pollution (Coray et al., 2006),  
426 making them more sensitive to stress than ME. The last fact contradicts the hypothesis of better  
427 performance of MT to pollutant influence allowing them to set populations in active port harbors.

428 A third explanation is also possible. Transmissible cancer is known to circulate in MT  
429 populations (Skazina et al., 2021). This disease can cross species barriers (Metzger et al., 2016): MTs  
430 can infect other mussel species with their cancer cells, causing disseminated neoplasia in them.  
431 Although it has not yet been proven that transmissible cancer coming from MT can cause a spike in  
432 clam mortality, it can be assumed that it is very likely, since cancer-like diseases are one of the cause of  
433 mussels mortality (Benabdelmouna & Ledu, 2016; Burioli et al 2019), mussels infected by MT-

434 originated neoplasia die during several weeks, but cases of remission is known as well (Burioli et al  
435 2019). If propose that MT are more resistant to MT-originated transmissible cancer (Yonemitsu et al.,  
436 2019, проверить) then it is expected that in port harbors where MT were introduced to, the abundance  
437 and sequently the proportion of ME should decrease due to their higher mortality.

438 In any case, our data allow us to confirm the earlier stated hypothesis that in the case of the White  
439 Sea, at least, MT is an invasive species (Vainola&Strelkov 2011). Unfortunately, only few data could  
440 be provided to check this hypothesis for other european MT populations since the closeness to the port  
441 harbors was never included into species distribution models developed for other areas. In some cases it  
442 is probably an unrealistic aim due to absence of predictor variability. For example in the case of MT  
443 populations in the Baltic Sea, where there is extremely active ship traffic and vessels visiting almost all  
444 parts of the Baltic, it is not possible to identify a correlation between the distribution of MTs and the  
445 presence of ports. Unless, of course, we consider the whole Baltic Sea as a huge port area where MTs  
446 were once introduced some ages ago. This is contradicted by the significant genetic difference of the  
447 Baltic MTs from the conspecifics represented in other regions of the Atlantic (Wenne et al., 2020). At  
448 the same time, if we look at the map of MT distribution (Vainola & Strelkov, 2011), the confinement of  
449 MT to large European ports out of the Baltic Sea (e.g. Bergen, Murmansk), rather speaks in favor of the  
450 hypothesis of MT introduction due to ship traffic. The unexpected increase in the number of MT in  
451 aquaculture plantations in the Scottish fjords (Beaumont 200+) is rather consistent with the hypothesis  
452 of the recent invasion as well.

453 Our analysis revealed that the higher the wind fetch value, the lower the PT i.e. MTs avoid open  
454 coasts potentially exposed to the surf. The lowest PT value (i.e., E-morphotype mussels dominate in the  
455 settlements) is observed on open shores of the northeast coast of Kandalaksha Bay (Fig. ++), exposed  
456 to wave action due to southeast and southwest winds which are frequent in Kandalaksha Bay (+++).

457 This result corresponds well with the data on the thickness of MT and ME shells. It is known that MT  
458 have thinner, more fragile shells (Beaumont et al., 2008; Michalek et al, 2021). It is logical to expect  
459 that mollusks with such characteristics should avoid places with high wave loading. In such habitats,  
460 MEs with stronger shells are expected to benefit. A similar pattern can be seen in other contact zones.  
461 For example, in Greenland, an increased frequency of MT was noted at the sheltered top of the fjord  
462 (GLS and GLD sites, Fig. 1 in Wenne et al. 2016), while the population at a more open site located  
463 near fjord mouth (GLL site in Wenne et al. 2016) is dominated by ME.

464 Finally the nature of the substrate is the most powerful predictor that has an influence on PT  
465 variation (Table ++, Figure ++). The proportion of T-morphotype mussels appears to be significantly  
466 higher on fucoids than on bottom substrates (silty-sandy sediment, pebbles, boulders and rock surfaces)  
467 located in close proximity (not more than 20 m, within the same fucoid belt). This pattern was first  
468 detected in the analysis of a significantly smaller volume of material (Katolikova et al., 2016). Bringing  
469 in new data collected in a wider range of conditions confirmed the presence of this pattern. It can be  
470 assumed that the effect of substrate type on PT has the same mechanism that underlies the negative  
471 correlation of PT with wind fetch. Fucoids may work as shock absorbers, mitigating, to a certain extent,  
472 the impact of the surf. Another explanation could also be given. Perhaps MTs are capable of attaching  
473 to underwater substrates more strongly than MEs do. This should cause less strongly attached molluscs  
474 to break off and remain on the bottom under constant fluctuations in the thallus of the fucoids. This  
475 hypothesis is contradicted by the fact that a comparison of attachment strength in MTs and MEs from  
476 the contact zone in Canada, showed no significant differences between the species when they were kept  
477 under laboratory conditions (Lowen et al., 200+). Additional investigations of attachment strength of  
478 both species are needed.

## *Conclusion*

Our analysis did not reveal the large-scale clinal patterns in the distribution of MT and ME associated with salinity gradients (temperature gradient was out of our interest since temperature does not show variation in the studied region) discussed in previous works. We have shown that in Kandalaksha Bay MTs avoiding exposed areas are concentrated near port harbors, where they were probably originally introduced. Apparently, there are some factors preventing MT dispersal from these refugia. We have previously suggested that such a factor may be predators (Khaitov et al., 2018), for which port areas may be unfavorable due to their anthropogenic contamination. Our current study revealed the segregation of MT and ME at even smaller spatial scales. We showed that these two species segregate within the same habitat (the local fucoid belt), inhabiting different substrate types. This is the pattern to be expected if we have deal with the two competing species occurring in common habitat. This suggests that, in addition to the influence of predators, competitive relations with ME may also constrain MT dispersal.

If this hypothesis is correct, then we can expect three temporal patterns in changes of MT:ME ratio. Firstly, one of the species can complitelly outcompete the other. Secondly, MT and ME can form a stable system with each species occupying its own ecological niche. Or, more likely, taking into account the instability of mussel settlements (Khaitov, Lentsman, +++), these two species will alternately colonize certain habitats changing each other in time.

## *Literature cited*