**Blue mussels Mytilus edulis L. and M. trossulus Gould in sympatry: assessment of ecological niche divergence using species distribution modeling**

V. M. Khaitov, A. A. Zaychikova, P. Y. Safonov, M. V. Katolikova, M. V. Ivanov, P. P. Strelkov

## **Abstract**

## Было 246 слов, стало 247 слов.

Species distribution models (SDMs) describing the relationship between species occurrence and environmental parameters can be used to assess the ecological niche of a species. Usually applied to morphologically distinct species, SDMs are also a promising tool for describing niche partitioning in coexisting cryptic species. An example of the latter in the marine realm are blue mussels *Mytilus edulis* (*ME*) and *M. trossulus* (*MT*). Despite considerable research effort, little is known about how they share space and resources in sympatry anywhere except in the Baltic Sea. Salinity, substrate, surf and proximity to harbors have been suggested as candidate factors but no conclusion has been made. Here we assessed partial effects of these predictors on divergence of *ME* and *MT* in the White Sea littoral applying SDMs to 570 mussel samples with known taxonomic structure. We found that each of the predictors influenced spatial segregation. The most expected habitat of *ME* was a bottom substrate in a wind-exposed location with a “normal” salinity (24 ppt) away from ports and large rivers, while for *MT* it was an algal substrate in a wind-protected area with a lower salinity close to ports and large rivers. The degree of segregation was density-dependent, increasing with increasing *ME* abundance. This indicates that *ME* outcompetes *MT* on bottom substrates. We discuss whether the predictors used in our study can drive the segregation of these species in other habitats than the littoral fucoid belt, elsewhere than in the White Sea and outside the Kola contact zone.

## **Introduction** 1500 слов но без не многих референсов

Было: 1608 со ссылками. Стало: примерно 1553.

Species distribution models (SDMs) are a numerical tool describing the relationship between species occurrence and environmental parameters. Using SDMs, it is possible to predict distribution patterns of species in space and time and to assess their ecological niche in a formal way (Elith, Leathwick, 2009). Joint application of SDMs to several coexisting species, i.e. a community, allows one to describe the partitioning of ecological niches between them. This is referred to as Joint Species Distribution Modeling (JSDM) (Ovaskainen, Abreg, 2020). In other words, SDMs may describe the axes in ecological space along which coexisting species occupying separate niches are segregated. SDMs can be built using various approaches, from regular multiple regressions to advanced machine learning (Elith et al., 2006).

SDMs are usually applied to “good”, i.e. morphologically distinct species (e.g. Reiss et al., 2011; Lindegren et al., 2022), which can be easily involved in routine studies requiring numerous samples. However, there is increasing evidence about coexistence of cryptic species (Bickford et al. 2007; Geller et al. 2010; Struck et al. 2017) and infraspecific taxa (Dufrenses et al. 2023). It is unlikely that any coexisting taxa have identical ecological phenotypes. The question how such species share space and resources in sympatry can be answered using SDM/JSDM (DeMarche et al. 2019?; REF). In marine ecology, this approach has already been successfully used, though only in a few studies (Lowen et al 2019; Dennis, Hellberg, 2010; Hu et al. 2021?).

The longest-known and best-studied cryptic species in the marine realm are those of the blue mussel (*Mytilus*) complex (Knowlton 1993; Gosling 2022). It is represented by several species that are easier to distinguish genetically than morphologically and that hybridize in sympatry (Koehn 1991; Gardner et al. 2021). Blue mussels, powerful ecosystem engineers in temperate and subpolar seas, play an major role in coastal communities (Bushbaum et al, 2009; REF) and are important aquaculture objects (Gosling 2021). In North Atlantic, the dominant species of blue mussels are *M. edulis* (*ME*) and *M. trossulus* (*MT*), which form zones of sympatry (hereafter, contact zones) in different parts of the region (Wenne et al. 2020 and references therein). *ME* and *MT* are fairly old species dating back to the Pliocene. They evolved in allopatry in the Atlantic and the Pacific Ocean, respectively, and their contact zones are thought to have formed as a result of repeated *MT* invasions from the Pacific Ocean to the Atlantic as well as from one part of the Atlantic into another (REF).

On the biogeographic scale, the distribution of *ME* and *MT* is thought to be regulated mostly by temperature and its correlates (Hayhurst, Rawson, 2009; Wenne et al., 2020). *MT*, which appears to be a more stenothermic and cold-loving species, is not distributed as far south into temperate seas as *ME* (Wenne et al., 2020). In the Arctic these two species co-occur. In contact zones, *ME*, *MT* and their hybrids are often found in the same mussel settlements (VS 11; REF), hereafter referred to as “mixed” settlements.

There is a general agreement that sympatric *ME* and *MT* are ecologically distinct (RC 05; Katolikova et al. 2016; REF) and have different economic value in aquaculture (Beaumont et al. 2008; Penney et al. 2002). However, information on the factors of their ecological segregation is fragmentary and contradictory. The greatest progress in comparative ecological studies of *ME* and *MT* in sympatry has been made in the contact zones in the Baltic Sea, in the waters of the Kola Peninsula (White and Barents Sea) and in West Atlantic (mostly Gulf of Maine and New Scotland). In the Baltic Sea, the brackish areas of its inner part are inhabited by *MT*, while the saltier areas closer to the North Sea are inhabited by *ME*. In the contact zone between these parts, mixed settlements can be dominated by hybrids, and *MT* gene frequency gradually increases towards the inner Baltic (VS 11, Zbawicka et al. 2014, Stuckas et al. 2017). As a result, the distribution of *ME* and *MT* is strongly correlated with salinity, while the role of other factors is negligible (Kijevsky; REF).

Another situation is observed in the contact zones of the Kola Peninsula and West Atlantic. Hybrids are always in the minority in mixed settlements there, and spatial distribution of *ME* and *MT* is mosaic both at the regional (i.e. dozens to hundreds of kilometers) and at the local scale. The relationship between the distribution of these species and salinity is not obvious anywhere in these contact zones (RC 05, Katolikova et al. 2016; Wenne et al. 2020, Marchenko et al. 2023), though several other factors of ecological segregation have been proposed (see below).

In the White and the Barents Sea, the frequency of *MT* is greater in port areas, possibly because this species has been introduced into the region with ship traffic in historic times (VS11, Katolikova et al. 2016). The only segregation factor explicitly tested in the White Sea is the substrate to which littoral mussels attach (Katolikova et al. 2016). It has been shown that *MT* is more common on fucoid algae while *ME* mostly lives directly on the bottom, on substrates such as mud, sand, stones and gravel. However, segregation across substrates cannot fully explain the local-scale mosaic in the distribution (Katolikova et al. 2016).

In the Barents Sea, no correlation with substrate has been found. However, these species have different depth preferences there (Marchenko et al. 2023). *ME* appears to be a more sublittoral species and *MT* a more littoral one (Marchenko et al. 2023). In West Atlantic, depth, anthropogenic pollution levels and surf effects have been considered as possible factors affecting the segregation of *ME* and *MT* (++++), but no definite conclusions have been made (RC 05, наши обзоры)

To sum up, no simple “single-factor” pattern of species distribution has been revealed in the contact zones of *MT* and *ME* outside the Baltic. Moreover, some of the factors may potentially be collinear and confound the analysis. Ports are often located in storm-protected areas, usually close to river mouths, so that the effects of shipping (and other anthropogenic factors), surf and salinity are difficult to distinguish. The effects of depth and substrate may obscure each other since fucoids, common in the littoral, are rare in the sublittoral, where they are replaced by kelps (REF).

This lack of conclusive evidence is partly due to the fact that until recently scientists could identify cryptic species of blue mussels only with the help of labor-intensive genotyping methods and therefore could not handle large amounts of material (Khaitov et al., 2021). In addition, there were no reliable statistical methods for modeling the distribution of sympatric taxa in the space of multiple factors, i.e. an SDM approach could not be implemented. To our knowledge, this approach has been applied only twice in the history of *ME* and *MT* studies: by Kijewski et al.(+++) and by Wenne et al. (+++). In both studies the machine learning techniques were used to model the macro-geographic distribution of species (technically, of allele frequencies at taxonomically informative genes) in the space of multiple climatic and oceanographic characteristics available from public databases. The authors concluded that temperature and salinity were important factors influencing the geographical distribution of these two species, with *MT* tolerating lower salinities and temperatures than *ME* (Kijewski et al. ???; Wenne et al. ???, see also above).

In our previous studies we found a simple semi-diagnostic trait for *ME* and *MT*, namely, the presence or absence of an uninterrupted strip of prismatic layer under the ligament on the inner side of the shell (Zolotarev ???, Katolikova et al. 2016). Using this finding, one could make reliable interpretations of the taxonomic structure of mixed settlements on the basis of morphotype frequencies in samples, i.e. without genotyping. This procedure was referred to as the “morphotype test” (Khaitov et al., 2021). In the White Sea 74% of *MT* but only 4% of *ME* have the strip (Katolikova et al., 2016), and the proportion of *MT* in samples (thereafter *Ptros*) is linearly dependent on the ratio of morphotypes (Khaitov et al., 2021). To note, hybrids are not considered as a separate category within this approach.

The aim of this study was to estimate the divergence of ecological niches between *ME* and *MT* in the White Sea littoral along environmental gradients such as substrate, salinity, surf level, and distance from ports. All these factors have been suggested as potentially influencing segregation of these two species in sympatry (REF). Another candidate factor, depth (REF, Marchenko et al. 2023), was not examined in our study but was controlled by sampling at the same littoral level. To achieve our aim, we examined the variability of the environmental predictors mentioned above and the taxonomic structure of mussel settlements using an extensive material (95 study sites, 570 mussel samples, 55,529 mussels) and assessed the partial influence of the predictors on the distribution of proportion of *MT* using SDMs. Since all predictors were included in one model, collinearity could be controlled. The predictive power of the model was evaluated using testing datasets from the White and the Barents Sea. In addition, we checked whether the pattern of species segregation by substrate was density-dependent, i.e. whether there was competition between the two species.

## **Materials and methods** 1530 слов но без некоторых ссылок **НЕ ПРАВИЛА**

Study area

The study area was the Kandalaksha Bay, where all previous *ME* and *MT* studies in the White Sea have been conducted (REF). The 185 km long Bay is funnel-shaped with numerous islands and skerries and a highly indented coastline (Fig. 1). Climate is continental subarctic with 4-5 months of ice cover and the average monthly sea surface temperature in August of 13.8°C. Mean tidal range is about 2 m. Summer surface salinity is 24 ppt throughout most of the Bay (“normal” salinity for most of the White Sea), but is lower in the estuarine areas (Berger ???). Two hydropower plant’s canals and 24 rivers with a catchment area of 141 - 12830 km2 (Median 240 km2; see Stable ++ ) flow into the Bay, with the largest river, the Niva, entering the Bay at its very top. Due to the complex geometry of the shoreline and numerous rivers, local surf and salinity gradients are pronounced (Filatov et al., 2005).

Historically (through the 20th century), six ports were functioning in the area operating oceanic vessels (Fig. 1). Two of these ports, both at the top of the Bay, are functioning while the remaining are abandoned (Sailing directions of the White Sea, 1932; Krasavtsev, 2011) but still visited by small ships, according to common knowledge.

Mussels are omnipresent in the shallow waters of the Bay. They are particularly abundant in the littoral fucoid belt (mainly *Fucus vesiculosus* L. and *Ascophyllum nodosum* L.), which is continuous 0.5-1.0 m above mean spring tide depth (Berger et al. 2001). According to data from 2002–2013, both mussel species were almost ubiquitous in the Bay, but their ratio in settlements varied greatly and *ME* was generally dominant (Katolikova et al. 2016).

## Modelling data set

Mussel sampling and processing

Between 2011 and 2018, mussels were sampled at 95 sites within the littoral fucoid belt (Fig. 1). Data for 17 of these sites were included in the study by Katolikova et al. 2016, the remaining data are new. Sites were chosen to describe littoral populations of the Bay in as much detail as possible and to account for the heterogeneity of their habitat by substrate type, surf level, and distance from rivers and ports. All samples were taken within the fucoid belt to minimize differences in depth. At each site, three samples from fucoid thalluses (hereinafter, algal samples) and three samples from bottom substrates (bottom samples) were collected a few meters from each other using 0.25 m2 and 0.025 m2 frames respectively. When sampling, the frames were not placed randomly, but in such a way as to capture the dense mussel aggregations.

Because shell morphotypes can be reliably identified in mussels larger than 10 mm (Khaitov et al. 2021), only such mussels were selected from samples and used in further analyses. The procedure for selection of mussels from samples taken from different substrates was different. From the bottom samples, all mussels were selected. From algal samples, one bundle of algae, selected so that to contain at least a few dozens of mussels, and all other algae were weighed separately, along with attached mussels. Only mussels from the bundle were used in further analysis. To calculate the total number of mussels in the sample, the weight of the bundle with the known number of mussels and the rest of the algae was compared. For 12 sites the information on total number of mussels in algal samples was lacking, and these sites were excluded from those analyses which required data on mussel abundance (Model 2, see below).

Shell morphotypes (E-morphotype, characteristic to *ME*, and T-morphotype, characteristic to *MT*) were identified for all selected mussels as in Khaitov et al. (2021). For further, the proportion of morphotypes was converted to the proportion of *MT* (*Ptros*) in each sample, in pooled samples from each substrate from each site (*Ptros*Algae and *Ptros*Bottom) and in pooled samples from each site (*Ptros*Site), using equation

, where *PT* - proportion of T-morphotype.

This equation, derived from the 24 genotyped samples (in total 1105 multilocus mussel genotypes) from the Kandalaksha Bay, reliably predicts *Ptros* over the entire salinity range in the White Sea (i.e., up to 24 ppt), but may overestimate *Ptros* at higher salinities, as it was observed in the Barents Sea (Khaitov et al. 2021).

Environmental parameters assessment

In total, we used seven parameters describing possible influence of rivers, ports, surf and type of substrate on mussels. The list of parameters and their description is provided in **Table 1**. We used three different proxies of salinity (*RiverSize*, *DistRiver* and *Salinity*), believing that a singular estimate of salinity at low tide could be insufficient to characterize overall salinity, and river influence *per se* at the sampling sites. *Salinity* was measured directly with accuracy of 1 ppt using an “Atago S/Mill-E” refractometer. To classify rivers by size (*RiverSize*), the data from the ESM +++ was used. To calculate *Fetch*, the R-package “windfetch” (Seers, 2022) was applied to regional geographic map shape-files.

Table 1. Environmental parameters studied

|  |  |  |  |
| --- | --- | --- | --- |
| Environmental parameter/ model predictor | Type | Explanation | Range (median) in the data |
| *Influence of substrate* | | | |
| *Substrate* | Categorical | Algal and bottom samples for each site are treated separately | Algae vs bottom |
| *Influence of rivers* | | | |
| *Salinity* | Continuous | Surface salinity (ppt) at the time of sampling, i.e. at low tide. | 2-30 (19) |
| *DistRiver* | Continuous | The straight line distance (km) between the site and the nearest river mouth by map. Log-transformed values were used. | 0-18.5 (4.9) |
| *RiverSize* | Categorical | Rivers are categorized according to whether their catchment area is larger or smaller than the median area for all rivers in the region. | Small vs large |
| *Influence of ports* | | | |
| *DistPort* | Continuous | The straight line distance (km) between the site and the nearest port by map. Log-transformed values were used. | 0.1-82.2 (18.7) |
| *PortStatus* | Categorical | Ports are categorized according whether they are active or abandoned | Active vs abandoned |
| *Influence of surf* | | | |
| *Fetch* | Continuous | Unobstructed length of water surface (km) over which wind from a certain direction can blow over. Log-transformed values of average fetch for the four cardinal directions were used. | 0.2-28.8 (3.3) |

Testing datasets

Three datasets were used as testing ones. The “Kanalaksha littoral” included 23 samples from 12 littoral sites in Kandalaksha Bay. At four sites only algal samples were taken, at four - only bottomones, and at the rest four - from both substrates (STable ++, SFig. 1 B). Environmental parameters were assessed in the same way as for the modeling dataset.

The second and third testing datasets were extracted from the published data of Marchenko et al. 2023. These authors mapped in detail the distribution of *Ptros* in mussel settlements of the Tyuva Inlet in the Kola Bay of the Barents Sea (**Fig. 1**) in 2009-2010 and provided for each sampling site a number of environmental characteristics including depth, *Salinity*, cover of macrophytes in rank scale, and dominant algal species (usually, kelps in the sublittoral and fucoids on the littoral). From these data we extracted two datasets, “Tyuva littoral” and “Tyuva sublittoral”. The first set included samples from all 23 studied littoral sites from the depth range corresponding to the fucoid belt (0.5-1.5 m above mean spring tidal depth, Marchenko et al. 2023; note that the position of fucoid belt in the Barents Sea differs from that in the White Sea due to the different tidal amplitude). The second included samples from all 15 studied sublittoral sites (depth range from -0.5 to -3.5 m). Since the substrate of mussel fouling was not registered during sampling, we classified samples into bottom and algalones by the algal cover in studied sites (ranks 1-3 and 4-5, correspondingly). The remaining environmental parameters were assessed as for the modeling dataset, with the nearest port in Ekaterininskaya Gavan Bight considered as Active and the river Tyuva flowing into the inlet as a large one.

*Statistical analysis*

All processing was performed using the statistical programming language R 4.05 (R core Team, 202++)

***Dependency of Ptros on environmental parameters in modeling dataset (Model 1)***

We used GAM (generalized additive model, Wood, 2017) as a modeling technique, which works well for SDM construction (Elith et al., 2006). Importantly, it assumes that the relationship between the dependent variable (in our case *Ptros*) and continuous predictors may not necessarily be linear, as in ordinary regression analysis, but curvilinear (Austin, 2002). The GAM fitted (thereafter *Model 1*) was based on beta-binomial residuals distribution and the restricted maximum likelihood method for parameters estimation. Smoothers for all continuous predictors were fitted using cubic basic splines. Categorical predictors were included as parametric terms in the model. *Site* was considered as a random factor. The function gam() from the package “mgcv” (Wood 2017) was used to fit the model.

To check for the predictors’ collinearity in the model we calculated the variance inflation factor (VIF, Fox & Monette, 1992). Additionally we calculated Pearson correlation between continuous predictors. To verify that the Model 1 meet the assumptions of sampling independence, we examined the presence of residuals’ spatial autocorrelation by means of spline correlogram construction (Bjornstad, Falck, 2001) with the function spline.correlog() from the package “ncf” (Bjornstad, 2022) and founded no evidence of spatial autocorrelation.

#### *Association between Ptros, substrate and mussel abundance*

The ultimate goal of the analysis was to answer the question of how the segregation of *ME* and *MT* between algal and bottom substrates depends on the abundance of each species on each substrate. For each site we calculated the difference between proportion of *MT* in algal and bottomsamples: *Dif* = *PtrosAlgae* - *PtrosBottom*. The obtained *Dif* values were used as a dependent variable in the *Model 2* which was constructed as GAM with gaussian residuals’ distribution.

To assess the dependence of *Diff* on *PtrosSite* and mussel abundances, we could not directly operate on *ME* and *MT* densities because they cannot be calculated other than through *Ptros*; this will inevitably lead to collinearity of the predictors. Therefore, we performed principal component analysis for the abundance matrix of T- and E-morphotypes on algal and bottom substrates and used PC1 and PC2 values as independent variables, along with *PtrosSite*, in the Model 2. Although we are here forced to equate morphotypes with species, this should not crucially bias the results of the analysis, given the proportional relationship between *PT* and *Ptros* in mussel settlements from the studied area (Khaitov et al. 2021). We used VIF to control for the level of collinearity of the final set of predictors considering the value less than 3.5 as acceptable (Quinn & Keugh REF).

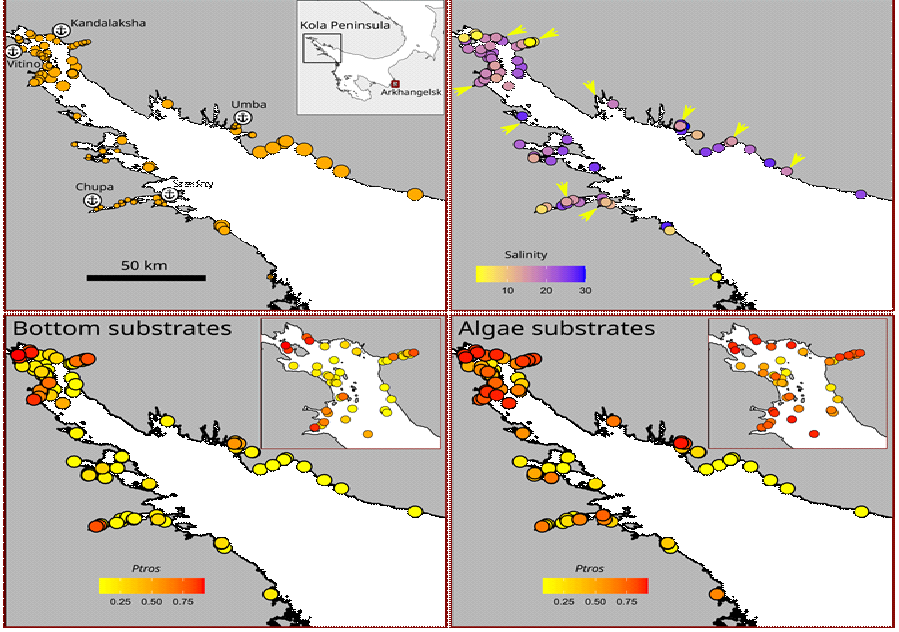
#### *Assessment of the Model 1 predictive power*

We sought to answer the question of whether Model 1 could be used to predict which species dominates bottom and algal samples in a site with known environmental parameters, *MT* (*Ptros*>0.5) or *ME* (*Ptros*<0.5). To do so, using all the model parameters we predicted *PtrosAlgae* and *PtrosBottom* for each site both within the modeling and each of the three testing datasets and categorized them into those greater than or less than 0.5. The predicted values were considered as classifiers for detecting *MT*- or *ME*-dominated samples. The receiver operating characteristics (ROC) followed by analyzing the area under the curve (AUC, Fielding, Bell,1997; Fawcett, 2006) was used to evaluate models performance. The function roc() from the package “pROC” (Xavier et al. 2011) was used.

## **Results**

147

Ranges and median values of the continuous predictors are summarized in Table 1. While the distribution of *Fetch* and *Salinity* values was highly mosaic, the most wind-exposed sites were located on the southeastern coast of the Bay and on open shores of the islands in its top (Fig. 1 A) while the most desalinated areas were located in the very top of the Bay (Fig. 1 B). Expectedly, *Salinity* tended to decrease towards river mouths (SFig. ++ D) and was lower closer to large rivers than to small ones (SFig. ++ D). Sites close to ports tended to have lower *Fetch* (SFig. +++ F), but no association between *DistPort* and *Salinity* was observed (SFig. +++ E). All correlations between *Salinity,* *DistRiver,* *DistPort* and *Fetch* were rather low (STable +++), the largest being that between *Fetch* and *DistPort* (r = ++). 140

~~~~

**Figure 1**. Taxonomic structure of mussel settlements and their habitat characteristics. (A) Map of Northern Europe. Box shows position of Kandalaksha Bay, arrow show position of Tyuva Inlet. B) Surf conditions. Point size is proportional to *Fetch*. Anchors with names mark ports; asterisks mark abandoned ports. (C) Salinity conditions. Point filling is proportional to *Salinity*. Arrows mark mouths of large rivers. (D-G) Proportion of *MT* in bottom (*PtrosBottom*, D-E) and algal (*PtrosAlgae*, F-G) samples. Point filling is proportional to *Ptros*. E and G show the Bay’s top in higher resolution. Было 104; стало 90.

Было 94.

When one examines separate maps of *Ptros* distribution across algal and bottom substrates, the universally elevated proportion of *MT* on the former is striking (**Fig. 2 D-G**). While spatial distribution of *Ptros* was highly mosaic, its maximum values on both substrates were observed in the Bay’s top and in some deep inlets, while its minimum values were observed along the open part of the southeastern coast (**Fig. 2 D-G**). Associations between *Ptros* and environmental predictors other than substrate could not be discerned on the maps (**Fig. 2**). Стало 87.

Relationship of *Ptros* and environmental parameters evaluated by Model 1

45

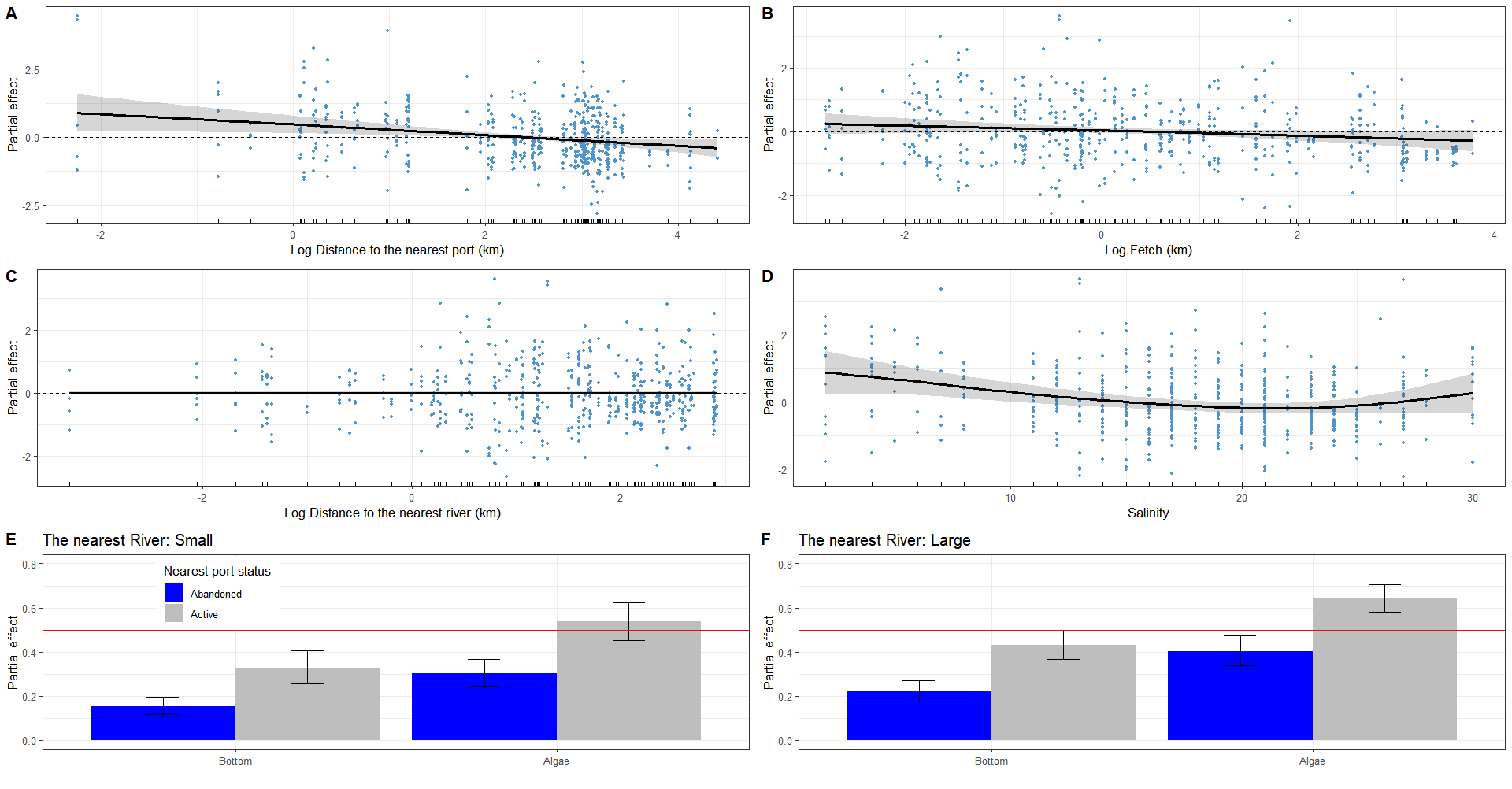
Although some non-zero pairwise correlations between environmental factors were found (see above), VIF values calculated for the predictors were generally low (maximal VIF being that for *Fetch*, 1.76). In our opinion, this result means that the collinearity between the predictors was negligible, i.e. they did not mask each other’s influence. 50

56

Model 1 explained 77% of the total deviance. It revealed a significant dependency of *Ptros* on all predictors except *DistRiver*. Effective degrees of freedom for *DistPort* and *Fetch* were close to one, indicating the linear dependence of *Ptros* on them. On the contrary, the dependence on the third continuous predictor, *Salinity*, was curvilinear (Table 2). 55

Table 2. Parameters of smoothers and coefficients of parametric terms for Model 1 describing dependency of proportion of *M. trossulus* in mixed settlements (*Ptros*) on environmental predictors. Edf – effective degrees of freedom.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Smoother terms** | **edf** | **ref.edf** | **Chi.sq** | **p-value** |
| s(Salinity) | 2.4 | 9 | 396.7 | 0.0033 |
| s(DistRiver) | 0.0 | 9 | 0.0 | 0.6724 |
| s(Fetch) | 0.9 | 9 | 88.2 | 0.0417 |
| s(DistPort) | 1.0 | 9 | 276.2 | 0.0016 |
| Random effect s(Site) | 74.4 | 92 | 453.6 | 0.0000 |
| **Parametric terms** | **Parameter estimate** | **SE** | **z-statistic** | **p-value** |
| (Intercept) | -1.7 | 0.1 | -11.8 | 0.0000 |
| Substrate(Algae) | 0.9 | 0.1 | 14.6 | 0.0000 |
| RiverSize(Large) | 0.4 | 0.2 | 2.6 | 0.0091 |
| PortStatus(Active) | 1.0 | 0.2 | 5.7 | 0.0000 |

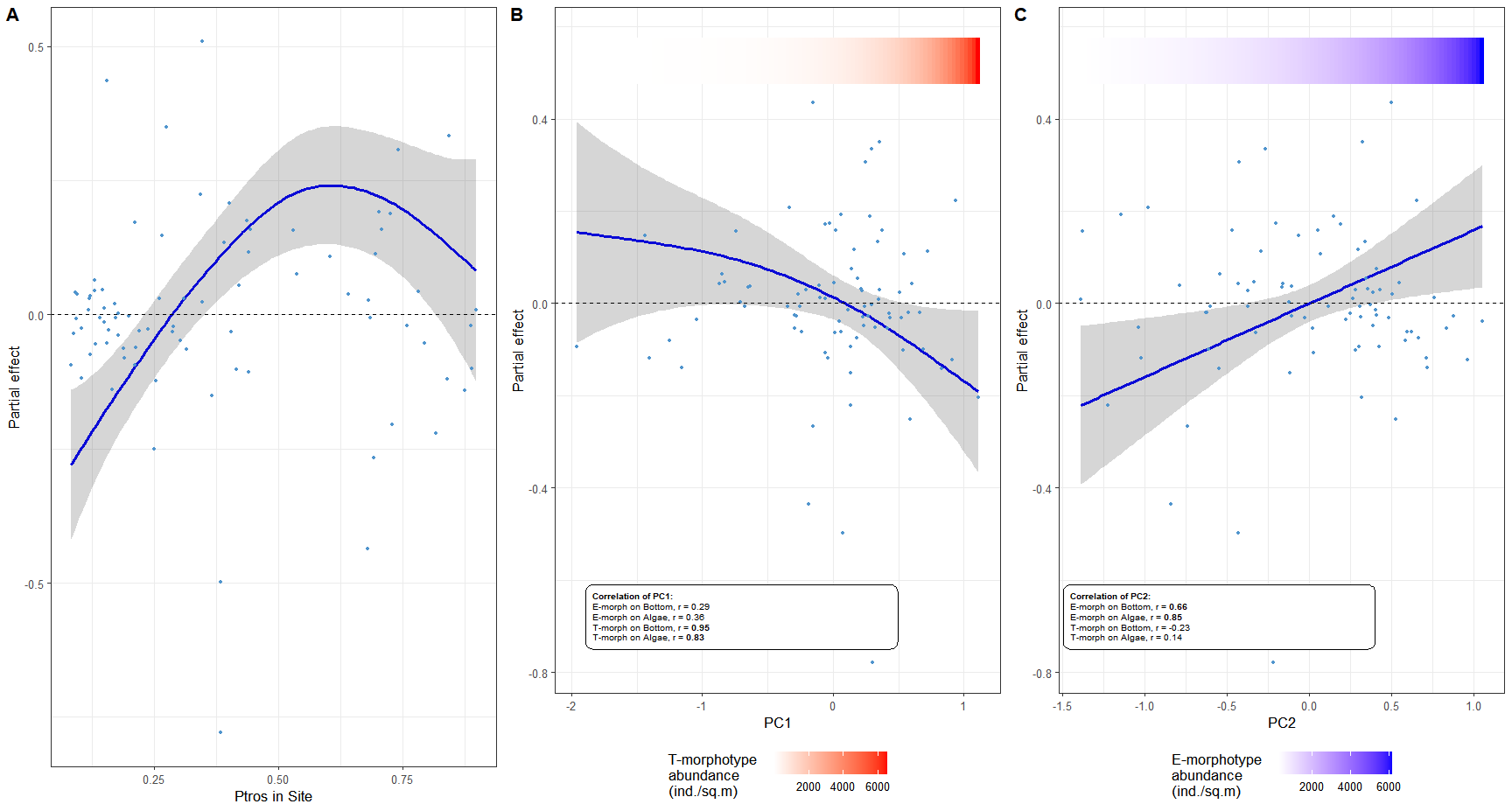


**Figure 2.** Partial effects of environmental parameters on proportion of *M. trossulus* in mixed settlements (*Ptros*) evaluated by the GAM fitted (Model 1). A-C. Dependency of *Ptros* on distance to the nearest port (*DistPort*, A), wind exposure (*Fetch*, B), distance to the nearest river (*DistRiver*, C) and salinity at low tide (*Salinity*, D). Gray ribbons represent 95% confidence intervals. Dotted horizontal lines indicating zero partial effect are given to show the wiggling of the fitted curves. Points on panels A-D show partial residuals, not raw data. E-F. Dependency of *Ptros* on combinations of categorical predictors. Partial effects of substrate (bottom vs algae) and status of the nearest port (active vs abandoned) when the nearest river is small (E) or large (F). Whiskers represent 95% confidence intervals. Red solid lines indicating a partial effect of 0.5 are provided to facilitate visual comparison of panels E and F.

151

According to the model, *Ptros* decreased both with *DistPort* (Fig. 3 A) and with *Fetch* (Fig. 3, B). This means that the proportion of *MT* was higher near ports and in surf-protected areas. *PortStatus* also had a significant effect: predicted *Ptros* was higher near active ports than near abandoned ones (Fig. 3 E, F). The curvilinear dependence of *Ptros* on salinity can be described as follows: predicted *Ptros* decreases with salinity in the range from low to “normal” salinity (24 ppt in the White Sea) and increases again at higher salinities (up to 30 ppt) (Fig. 3, D). Besides, predicted *Ptros* was higher near large rivers than near small ones. Finally, *Ptros* was higher on algal substrates than on bottom ones (Fig. 1 C, D; Fig. 3 E, F). As mentioned above, distance to the nearest river did not affect *Ptros* (Fig 3 C). 144

Dependency of *Ptros* on substrate and mussel abundance evaluated by Model 2



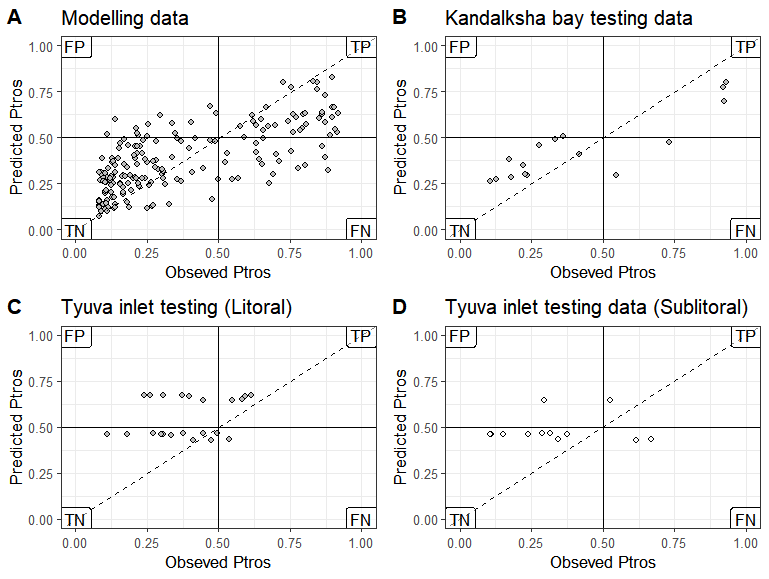
**Figure 3.** The dependence of difference between proportion of *MT* on algal and bottom substrates (*Diff*) on proportion of *MT* in a site (*PtrosSite*) (A) and estimations of total abundance of *MT* (B) and *ME* (D). Principal components from the matrix of T- and E-morphotypes abundances on different substrates are considered as proxies for *MT* and *ME* abundances (PC1 and PC2 respectively). Points reflect partial residuals, not raw data. Colored gradient bars at the top of the figures reflect linear associations between PC1 and T-morphotype (B) and PC2 and E-morphotype abundance (C).

92

In the principal component analysis of the abundance matrix of T- and E- morphotypes on different substrates, PC1 and PC2 explained 62% and 20% of variation, respectively. A high positive correlation of PC1 with abundances of T-morphotypes and of PC2 with abundances of E-morphotypes was found on both substrates (Fig. 4 B, C). Thus, the abundance of conspecific morphotypes varied consistently on different substrates (see also **Fig. 1C, D)**. Therefore, PC1 and PC2 can be considered as proxies of *MT* and *ME* abundance, respectively. 84

Parameters of Model 2, which explained 31% of the deviance, are provided in ESM (STable ++). **Figure 4** demonstrates how the difference between *MT* proportion on algal (*PtrosAlgae*) and bottom (*PtrosBottom*) substrates (*Dif*) depends on *MT* prevalence at the site (*PtrosSite*) and mussel abundances in terms of PCs according to the model. The dependence of *Diff* on *PtrosSite* was significant (p < 0.001, STable ++) and, expectedly, bell-shaped, with minimal values at sites absolutely dominated by *ME* or *MT* (*Ptros* close to 0 or 1) and maximal at sites with equal presence of both species (Fig.4 A). Dependence of *Diff* on PC1 was marginally significant (p = 0.087) and tended to decrease with increasing PC1 (Fig.4 B). The dependence of *Diff* on PC2 was significantly positive (p = 0.011, STable ++) (Fig.4 C). This means that the species were strongly segregated by substrates at sites with a high *ME* abundance but not at sites with a high *MT* abundance.

Assessment of predictive power of Model 1



**Figure 4.** Ability of SDM (Model 1) to predict proportion of *MT* (*Ptros*) in mussel samples from the modeling (A) and the testing data sets (B-D). Each plot compares empirical *Ptros* in samples from algal and bottom substrates and *Ptros* predicted by the model within the particular data set. If the empirical and the predicted values were the same, the points would lie on the diagonal (dashed line). Solid lines delineate *MT*- and *ME*-dominated samples on each axis. Labels mark the quadrants with false positive (FP), true positive (TP), true negative (TN) and false negative (FN) predictions in the analysis of the ability of the model to classify samples into *ME*- and *MT*-dominated ones. Dataset names are shown in chart headers.

95

The ability of Model 1 to classify samples into *ME*- and *MT*-dominated ones was good for the “Kandalaksha littoral” testing dataset (AUC=0.84 vs AUC=0.85 for modeling dataset), with only a few false negatives (i.e. sites unpredictably dominated by *MT*) (Fig. 5 A, B). Predictive value of the model for the two testing sets from the Barents Sea was lower, although not fatally so: AUC = 0.71 for “Tyuva littoral” and AUC=0.69 for “Tyuva sublittoral”. Unlike the “Kandalaksha littoral” testing dataset, most false results were positive, i.e., the model overestimated *Ptros* more often. 92

## Discussion

## 2574 слов, что многовато, но вместе со всем в 8000 слов вроде влезаем

Стало примерно 2540 слов.

85

Having applied the SDM approach to an unprecedentedly extensive material, we demonstrated that each of the environmental predictors considered in our study—namely, surf level, distance to the port, status of the port (active vs abandoned), salinity at low tide, size of the nearest river and fouling substrate (fucoid algae vs bottom substrates)—influenced the distribution of *Mytilus edulis* (*ME*) and *M. trossulus* (*MT*) in the White Sea. The differences in the distribution, evident at scales ranging from meters to tens of kilometers, reflected the partial divergence of ecological niches of these two species. 94

77

Below we discuss the species adaptations possibly underlying the patterns of *ME* and *MT* distribution against different predictors. Then we consider the possible role of competition in segregation by substrates. Further, we discuss whether the same set of predictors can drive segregation of these species in other habitats than the littoral fucoid belt, elsewhere than in the White Sea and outside the Kola contact zone. Finally, we review the strengths and weaknesses of our approach to assessing ecological niche partitioning of sympatric mussels. 83

Ecological niche partitioning between *MT* and *ME* in the Kola contact zone

96

Our results show that the most expected habitat for *ME* in the White Sea littoral is a bottom substrate in a surf-exposed location with a “normal” surface salinity for the White Sea (24 ppt) situated away from ports and large rivers. The most expected habitat for *MT* is an algal substrate in a wind-protected location with a lower-than-“normal” salinity situated close to active ports and large rivers. Only the differences related to ports and substrates have been previously noted in the White Sea (VS 11, Katolikova et al. 2016). 89

335

*Segregation by salinity*. In the Baltic *MT* is adapted to an extremely low salinity (REF +++), as also confirmed by ecophysiological data (Knöbel et al. 2021). Comparative ecophysiological data on *MT* and *ME* elsewhere are inconclusive (Gardner & Thompson 2001; Qiu et al., 2002; Sokolova et al. 2024). Before our study, there has been no convincing evidence of segregation of these species by salinity in contact zones outside the Baltic, in particular, in the Kola zone (Moreau et al. 2005; REF; Katolikova et al. 2016; Marchenko et al. 2023). For the White Sea, this lack of evidence could be due to at least three reasons. Firstly, the role of salinity in species segregation may be masked by other important factors. Secondly, the range of salinity in mussel habitats in \*\*\* is relatively narrow as compared to the Baltic Sea. These two reasons might also explain the vague relationship between species segregation and salinity in other contact zones.

The third reason is the curvilinear dependence of the proportion of *MT* in mixed settlements (*Ptros*) on salinity: *Ptros* increases not only when the salinity is reduced but also when it is extremely high for the White Sea (up to 30 ppt, **Fig. 2D**). This nonlinearity, which may prevent the dependence from being detected, can be explained in two ways. On the one hand, local summer surface salinity above 24 ppt in the Kandalaksha Bay, supposedly associated with irregular episodes of upwelling (Dale & Prego 2003), may be a nonspecific stress for littoral animals adapted to lower salinity, while *MT* can tolerate it better (Katolikova et al. 2016, see also below). On the other hand, as shown in detailed studies at the Barents Sea (Khaitov et al. 2021, Marchenko et al. 2023), the method of predicting *Ptros* (“morphotype test”) used in our study may slightly overestimate it at salinities close to 30 ppt. Therefore, we cannot rule out the possibility that the increased *Ptros* at sites with a high salinity is an artifact. 330

141

*Non-random distribution depending on distance to ports.* It has been suggested that the confinement of *MT* to harbors in the White and the Barents Sea is associated its invasion into the region with maritime transport from the western Atlantic in the 20th century (Vainola & Strelkov 2011). This hypothesis agrees with the available genetic data (Simon, Wenne, Laakkonen ...). It has also been hypothesized that *MT* is more resistant to anthropogenic pollution and is in general a more opportunistic species, better adapted to disturbed habitats, than *ME* (Katolikova et al. 2016). Our observation that *MT* frequency is lower near abandoned ports than near active ones is consistent with this hypothesis. However, the propagule pressure of *MT* may have decreased near abandoned harbors in recent decades, which may have affected the size of its populations. 134

110*Segregation by surf level*. The fact that *ME* and *MT* are segregated by surf levels may be due to the well-known differences in the mechanical properties of their shells and the ability to form dense aggregations. *ME* has less flexible, thicker and heavier shells (Beaumont et al. 2008, Michalek et al, 2020), and is more inclined to form tight clumps (Liu et al. 2011). These features may be adaptive on exposed coasts. Unfortunately, there are no comparative data on the differences between *ME* and *MT* in byssus secretion and attachment strength, which theoretically might also affect their distribution by surf level as well as across substrates. 106

105*Segregation by substrate.* The differences in shell structure and aggregation behavior possibly explaining segregation by surf may also explain that by substrate. An ability to form dense aggregations is an adaptation to life on bottom, not on algae. Other things being equal, *MT*, with its thinner shells, should be lighter than *ME* (Michalek et al, 2020) and thus better adapted to life on algae. Further, fucoid thalli may serve as shock absorbers for fragile *MT* (Katolikova 2016) and shelter them from starfish selectively preying on *MT* in mixed settlements (Khaitov et al. 2019, 2023). 94

116

*Competition for substrate.* Whatever physiological, morphological, behavioral and other features influence the segregation of *MT* and *ME* by the environmental factors, interspecific competition may also be involved. Assessing the role of mussel abundance in the degree of species segregation across substrates, we found that while *MT* abundance did not significantly affect it, *ME* abundance did: as the latter increased, the degree of segregation increased, too (Fig. 4 B,C). In our opinion, this pattern results from the divergence of the realized species niches. *ME* outcompetes MT on bottom substrates displacing it to algal thalli, which appear to be a less suitable substrate for *ME* (see above). 105

84

Spatial segregation of sympatric mussels by substrates, which is apparently density-dependent, is evident at the level of tens of centimeters (Katolikova et al. 2016). Direct analogies for segregation at such a small scale can be found in other attached organisms, terrestrial plants (Raventos et al., 2010). A “biologically generated spatial pattern” model, relating inter-specific segregation with the intra-specific clustering in competing species, has been suggested (Paccala and Levin, 1997; Amarasekare, 2003). Our findings suggest that this model can also be applied to mussels. 83

107

*Predictive power of SDM.* The ability of our model to classify sites into *ME*- and *MT*- dominated ones in an independent testing dataset from the White Sea was high (AUC = 0.84). Therefore, we assume that the predictors included in the model explain most of the variation in species distribution within the studied habitat, the littoral fucoid belt. The model also showed a satisfactory performance on independent data from the Tyuva inlet in the Barents Sea (AUC ≈ 0.7), including sublittoral data. This result highlights the versatility of this set of predictors as regulators of *ME* and *MT* distribution in the Kola contact zone. 104182

The worst predictive value of the model for the Barents Sea data may be due to the following reasons. Firstly, considering that distribution of *ME* and *MT*  in the Tyuva Inlet by depth is non-random (Marchenko et al. 2023), it may be associated with a large depth range of the sampling sites. The second reason may be a coarser categorization of the Barents Sea samples into algal and bottom ones. Since fouling substrate was not taken into account during sampling, we predicted it based on the projective cover of algae at the sampling site. Thirdly, we do not know whether the two species are non-randomly distributed across bottom and algal substrates in the sublittoral, where fucoids are replaced by kelps. Finally, the fact that SDM tended to overestimate *Ptros* in the Barents Sea data (false positive predictions) is consistent with the observation that the proportion of *MT* has been declining in the study area in the 2010s under seemingly stable environmental conditions in terms of predictors included in our model (Marchenko et al. 2023). This observation suggests the presence of some yet unknown factors regulating taxonomic structure. 187

Ecological niche partitioning between *MT* and *ME* in the Kola contact zone as compared to other zones

229

Blue mussels are a challenging model for studying ecological niche partitioning between “cryptic” species in sympatry due to their wide distribution, biogeographic history and hybridization. *ME* and *MT* play similar ecological roles in their native oceans, Atlantic and Pacific, respectively (REF) and therefore may inherently have strongly overlapping fundamental ecological niches. Contact zones between these species in the Atlantic can be considered as ecological (and evolutionary) “experiments”, set in strikingly different environments (from Baltic to Spitsbergen) at different times (from late post-glacial to the historical period, VS 11, Wenne et al. 2020 and references therein). The “design” of these “experiments” was possibly different too, because that in some zones the original settler could be *ME* and in others, *MT*. In addition, competition (“ecological character displacement”, Pfennig and Pfennig 2020), hybridization (“reinforcement of prezygotic reproductive isolation”, Lukhtanov 2011) and introgression (“adaptive introgression” Herdick 2013) could influence the divergence of their ecological phenotypes differently in different zones. These considerations suggest that the zones should differ, and this hypothesis has been a recurrent theme in genetic research on blue mussel contact zones (RC, Bierne et al. 2011, Fraisse et al. 2016). Nevertheless, we believe that the differences between these two species are more fundamental and thus that conspecific ecological phenotypes (“niches”) in different zones should be similar, producing comparable patterns in species distributions. Some results of this study support this hypothesis. 228

123

The observation that *MT* frequency is elevated in low-salinity habitats not only in the Baltic but also in the White Sea seems to resolve the old conundrum about seemingly contrasting salinity adaptations of the Balticand other Atlantic *MT* populations (e.g. RC 05, Katolikova, VS11, see also above). Further, an increased *MT* frequency has been repeatedly observed in calm and freshened waters e.g. in the tops of fjords near Bergen in Norway (Ridgway, Nævdal 2004) and Uummannaq in Greenland (Wenne et al. 2016) and in Loch Etive in Scotland (Beaumont et al. 2008), which is hardly a coincidence. Our observations indicate that this combination of weak surf and low salinity is also favorable for *MT* in the White Sea. 119 146No non-random relationship between the distribution of *ME* and *MT* and any of the predictors significant in the White Sea has been convincingly demonstrated in other contact zones, with the exception of salinity in the Baltic contact zone. Data on surf are inconsistent (compare Bates, Innes 1995; Comesaña et al., 1999; Tam and Scrosati, 2013 and this study), while data on fouling substrates are, as far as we know, completely absent. If our assumption is correct and the diverging preferences of *ME* and *MT* for sites differing as to surf and substrates are associated with the differences in their morphology and behavior (see above), then these differences should be manifested universally. The causes of an increased resilience of *MT* to stress, particularly anthropogenic pollution, remain unclear (see Brooks et al. 2015 and Beyer et al. 2017 for discussion). 138

44

The classical review on the divergence of ecological niches of *ME* and *MT* in different contact zones (RC 05) is already 20 years old. The time is obviously ripe for a new survey, and our observations from the Kola zone may prove useful. 43

Strengths and weaknesses of our approaches to the study of sympatric mussels.

150

The methods of taxa identification, environment parameters assessment and modeling used in our study have certain limitations. We identified the mussels using the “morphotype test” allowing the assessment of the taxonomic structure of mussel settlements without genotyping. This test works well in habitats with salinity below 25 ppt in the Kola contact zone (Khaitov et al. 2021) but does not allow a direct assessment of species abundances or an identification of hybrids as a separate category. The former limitation makes it difficult to account for the role of inter-species competition, which, judging from our experience with different substrates, is important. The latter limitation is alleviated by the fact that hybrids are relatively scarce in the Kola zone. However, this is not the case in other contact zones (Väinölä, Strelkov 2011; Wenne et al. 2020), where hybrids may be important ecological actors (e.g. Schwartz et al. 2024). 146

82

Although *ME* and *MT* differ universally in morphotype frequencies, the magnitude of the differences varies between contact zones and between habitats with different salinities in the Arctic (Khaitov et al. 2021). This means that the “morphotype test” must be additionally calibrated before use (see Khaitov et al. 2021 for recommendations). Multilocus genotyping, while still too costly for processing dozens of thousands of specimens needed for SDM, remains the gold standard of taxonomic assessment in blue mussels. 76 80It should also be noted that we did not account for all potential predictors affecting species segregation (e.g., depth, Marchenko et al. 2023 or predators, Khaitov et al. 2019; 2023). Moreover, some of our predictors could have been estimated more carefully (for example, salinity at high water could be more informative for littoral mussels than salinity at low water when they are not submerged). However, since most of our predictors were shown to be significant, they should not be ignored in future studies. 84

73The correlative approach used in our study does not allow a direct assessment of either relative or absolute “importance” of factors. For instance, we cannot say whether salinity or substrate is more crucial. However, the take-home message from our research is that there is no single “leading” factor determining distribution of *ME* and *MT*, contrary to the idea that has dominated the field since the pioneering studies in the Baltic (RC 05; ???). 7361

The limitations discussed above do not detract from the fact that, as shown in our pioneering study, SDMs are a promising tool for the study of distribution of *ME* and *MT* in sympatry. Their obvious benefits include the possibility to analyze the distribution of the species in the space of multiple predictors simultaneously, the possibility to control the collinearity of the predictors and the lack of necessity to treat dependencies as linear. The development of new approaches to SDM construction may help unlock their full potential. 86

91

Promising directions of further research on niche partitioning in sympatric mussel species are, in our opinion, as follows. Firstly, a parallel study in different contact zones would reveal common and zone-specific patterns. Secondly, the use of taxonomic methods allowing direct assessment of abundances of species and their hybrids would elucidate the nature of competition between them all. Incorporation of additional environmental factors, including biotic ones, into SDMs might yield surprising results. Finally, it would be worthwhile to have a closer look at different spatial scales, down to the smallest one, in the segregation of these two mussel species. 98