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

Assessment of ecological niche divergence between blue mussels Mytilus edulis L. and M. trossulus Gould in the White Sea using species distribution modeling approach

Modeling of Mytilus edulis L. and M. trossulus Gould distribution in sympatric populations in the White Sea

Niche divergence of Mytilus edulis L. and M. trossulus Gould in the White Sea: towards to species distribution modeling

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

## **Abstract** - 246 слов из 250.

## Despite a long history of research on the widespread "cryptic" blue mussel species Mytilus edulis (ME) and M. trossulus (MT), little is known about how they share space and resources in sympatry anywhere but the Baltic Sea, where MT is a more estuarine species. Salinity, fouling substrate, surf level, and proximity to the harbors, among others, were considered as factors of the ecological segregation of species. Here we evaluate the partial effects of all listed predictors for ME and MT using a species distribution modeling approach and 570 mussel samples with known taxonomic structure from all over the Kandalaksha Bay in the White Sea. It was found that each of the predictors influence the spatial segregation of species. In the light of our findings, the most expected habitat for ME in the littoral is a bottom substrate in a location exposed to wind, with a “normal” for the White Sea salinity (24 ppt) and away from ports and big rivers, while for MT it is an algal substrate in an area protected from wind, with reduced salinity and close to ports and big rivers. We also attempted to answer the question whether the species segregation by substrate is density-dependent and found that the degree of segregation positively depends on ME abundance, an indication that ME outcompeting MT on the bottom substrates. We ask whether the differences between ME and MT revealed reflect differences in their niches and manifest in other areas of ME and MT coexistence.

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

Species distribution models (SDMs) being a numerical tool describing the relationship between species occurrence and environmental parameters allows predicting distribution patterns of species both in space and time. SDMs are also considered as a formal way of species’ ecological niche assessment (Elith, Leathwick, 2009). When the SDMs are jointly applied to a number of coexisting species, i.e. community, it allows to describe the ecological niches partitioning between species (an approach sometimes referred to as Joint Species Distribution Modelling, JSDM, Ovaskainen, Abreg, 2020). In other words, the SDM can answer the question along which axes in ecological space coexisting species are segregated, occupying different niches. The range of approaches to building SDM is very wide: from regular multiple regressions up to advanced machine learning methods (Elith et al., 2006).

SDMs are mostly applied for «good», morphologically distinct species (e.g. Reiss et al., 2011; Lindegren et al., 2022) that can be easily involved in routine studies requiring numerous samples. Increasing evidence for coexisting cryptic species (Bickford et al. 2007; Geller et al. 2010; Struck et al. 2017) as well as intraspecific taxa within polytypic species (i.e. subspecies, phylogenetic lineages) (Dufrenses et al. 2023) raises the question of how they share space and resources in sympatry. Since it is unlikely that any coexisting taxa have identical ecological phenotypes, there must be ecological niche partitioning between them, a question that can be answered using SDM/JSDM modeling (DeMarche et al. 2019?; REF). Strictly speaking, when applying modeling to such cases, we consider coexisting cryptic taxa as a community. In marine ecology, only a few studies in this direction have been conducted so far (Lowen et al 2019; Dennis, Hellberg, 2010; Hu et al. 2021?).

In the marine realm, the blue mussel *Mytilus edulis* complex is the longest scientifically known and best-studied complex of cryptic species (Knowlton 1993; Gosling 2022). It includes several species that are easier to distinguish genetically than morphologically and hybridize in sympatry (Koehn 1991; Gardner et al. 2021). Blue mussels are powerful ecosystem engineers in the temperate and subpolar seas playing an important role in coastal communities (Bushbaum et al, 2009; REF). They are also important objects of aquaculture (Gosling 2021). In the North Atlantic, the dominant species are *M. edulis* (thereafter *ME*) and *M. trossulus* *(MT*) which form zones of sympatry (thereafter contact zones) in different corners of the region from Scotland and the Gulf of Maine in the south to Greenland and Spitzbergen in the north (Wenne et al. 2020 and references therein). ME and MT are quite old species of Pliocene age, having evolved in allopatry in the Atlantic and Pacific oceans, respectively. The contact zones between them are thought to have formed as a result of repeated *MT* invasions from the Pacific Ocean as well as within the Atlantic (REF).

In contact zones, *ME*, *MT* and their hybrids are often found in the same samples from mussel settlements (VS 11; REF); such settlements are hereafter referred to as “*mixed*”. Scientists generally agree that *ME* and *MT* are ecologically distinct in sympatry (RC 05; Katolikova et al. 2016; REF) and have different economic value in aquaculture (Beaumont et al. 2008; Penney et al. 2002), but the data on the factors of their ecological segregation are fragmentary and contradictory.

On a biogeographic scale, the distribution of *Mytilus* species is thought to be regulated primarily by temperature and its correlates (Hayhurst, Rawson, 2009; Wenne et al., 2020). Both species occur in the Arctic but *MT* does not penetrate as far south into temperate seas as *ME*, appearing to be a more stenothermic, cold-loving species (Wenne et al., 2020).

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 Seas) and in the West Atlantic (mainly, Gulf of Maine and New Scotland). In the Baltic Sea, the brackish areas of the inner part of the sea are inhabited by *MT*, while the saltier areas closer to the North Sea are inhabited by *ME*. In the middle runs the contact zone, where mixed settlements could 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, species distribution is strongly correlated with salinity, against which the role of other factors is negligible (Kijevsky; REF).

In the Kola and West Atlantic zones the situation differs from that in the Baltic. Hybrids are always in the minority in mixed settlements. Species are distributed in space in a mosaic fashion, both in regional (i.e. dozens to hundreds of km) and local scales. The relationship between distribution of species and salinity is nowhere obvious (RC 05, Katolikova et al. 2016; Wenne et al. 2020, Marchenko et al. 2023), but a number of other factors of ecological segregation have been proposed.

In the White and Barents Seas, the frequency of *MT* is elevated in port areas, possibly due to introduction of this species into the region in historic times with ship traffic (VS11, Katolikova et al. 2016). The only one factor of species segregation explicitly tested in the White Sea was the substrate to which littoral mussels attach (Katolikova et al. 2016). It turned out that *MT* is more common on fucoid algae while *ME* mostly lives directly on the bottom (mud, sand, stones, gravel). The segregation of species across substrates, however, cannot explain the entire local-scale mosaic in their distribution (Katolikova et al. 2016). Studies in the Barents Sea did not consider mussel substrates, but confirmed the hypothesis of different depth preferences of species. It turned out that on littoral-sublittoral vertical transects, the proportion of *ME* increases with depth and, as a result, *MT* appears to be a more littoral species and *ME* a more sublittoral one (Marchenko et al. 2023). In the western Atlantic, depth, anthropogenic pollution levels and surf effects have been considered as possible factors in species segregation (++++), but no definite conclusions could be drawn in any case (RC 05, наши обзоры)

To sum up, in the non-Baltic *MT*-*ME* contact zones no simple, “single-factor” pattern of species distribution was revealed. We also draw attention that some candidate factors involved in the analysis can potentially be collinear, masking each other. For example, ports are often located in storm-protected areas, usually close to river mouths. This makes it difficult to isolate the effects of shipping (and other anthropogenic factors), surf and salinity. The same could be said about the effects of depth and fouling substrate since the littoral fucoids are rare in the sublittoral where they are replaced by kelps (REF).

This state of knowledge is not surprising given that for most of the time of blue mussel research, scientists have used labor-intensive genotyping methods to identify cryptic species and therefore could not handle the large amount of material (Khaitov et al., 2021). In addition, there were no reliable statistical methods available to model the distribution of sympatric taxa in the space of multiple factors, i.e. no SDM approach was realized. To our knowledge, in the history of *ME* and *MT* studies this approach has been applied only twice, by Kijewski et al.(+++) and by Wenne et al. (+++). Both times 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 conclusions of these studies are summarized by recognizing temperature and salinity as important factors influencing the geographical distribution with *MT* tolerating lower salinities and temperatures than *ME* (Kijewski et al. ???; Wenne et al. ???, see also above).

Finding a simple semi-diagnostic shell trait for *ME* and *MT* - the presence or absence of a uninterrupted strip of prismatic layer under the ligament on the inner side of the shell (Zolotarev ???, Katolikova et al. 2016), allowed us to reliably interpret the taxonomic structure of their mixed settlements without genotyping basing on morphotype frequencies in samples (a procedure referred to as “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 proportion of *MT* in samples (thereafter *Ptros*) is linearly dependent on the ratio of morphotypes (Khaitov et al., 2021). Note that hybrids are not considered as a separate category under this approach.

The goal of our study was to estimate the divergence of ecological niches between *ME* and *MT* in the White Sea littoral along such environmental gradients as substrate of fouling, salinity, surf level, and distance from ports. All these factors have been proposed in previous studies as candidate factors of segregation of species in sympatry (REF). The last of the previously suggested factors, depth (REF, Marchenko et al. 2023), was not studied by us, but was controlled by sampling at the same littoral level. To achieve this goal, we examined the variability of the aforementioned environmental predictors and the taxonomic structure of mussel settlements in an extensive material (95 study sites, 570 mussel samples, 55529 mussels) and applied the SDM approach to assess the partial influence of the predictors on the distribution of *Ptros*. Since all predictors were included in one model, this allowed us to control the collinearity between them. Ideally, a model trained on reliable data should allow the prediction of *Ptros* in independent data, and we evaluated its predictive power using testing datasets from the White and Barents Seas. We also attempted to answer the question whether the pattern of species segregation by substrate is density-dependent, i.e. whether the competition between species is invoked.

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

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).

There were 4 data sets used in the work: 1 modeling set and three testing sets

## Modelling data set

Mussel sampling and processing

Mussels were sampled at 95 sites within the littoral fucoid belt in 2011-2018 (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.

We used mussels with a shell length larger than 10 mm to identify the shell morphotypes reliably (Khaitov et al. 2021). In the bottom samples all mussels from a frame were used. In the algal samples the procedure was different. One bundle of algae, containing at least a few dozens of mussels, was chosen and weighed along with attached mussels. The rest of algae from a frame was weighted as well. Mussels from the bundle were counted and used for further analysis. The ratio between the counted number of mussels and to the boundle weight was applied to the total algal weight to reconstruct the total number of mussels in the sample. 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). Further, the proportion of morphotypes was converted to the proportion of *MT* (*Ptros*) in each sample, in pooled samples from each substrate at 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 “Tyuva littoral” and “Tyuva sublittoral” 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**) sampled in 2009-2010. They provided 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) for each sampling site. The “Tyuva littoral” 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 “Tyuva sublittoral” 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.

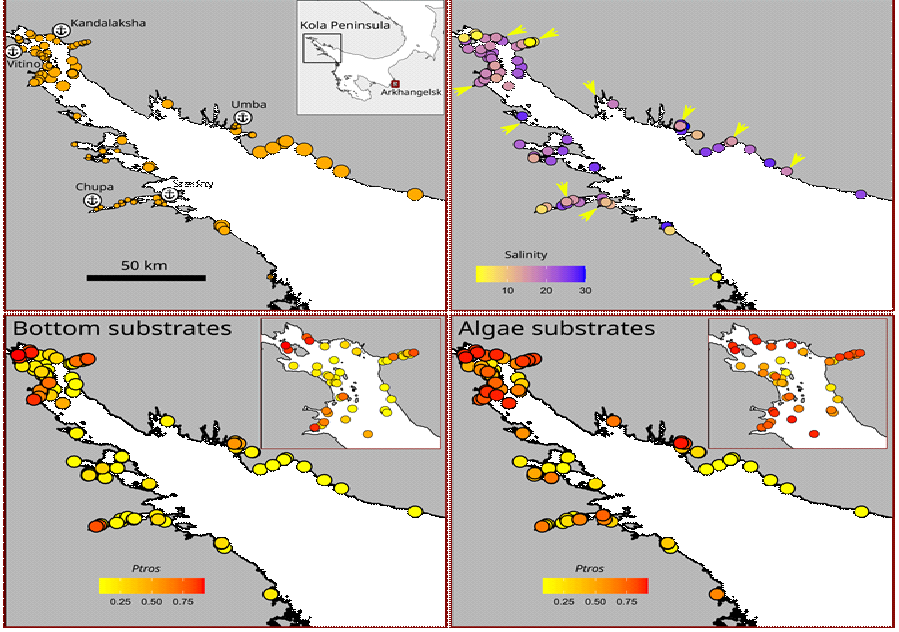
To assess the dependence of *Diff* on *PtrosSite* and mussel abundances, we could not directly operate on *ME* and *MT* densities because they cannot be calculated other than through *Ptros*; this will inevitably lead to collinearity of the predictors. Therefore, we performed principal component analysis 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 to be 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- 841 слов

The ranges and median values of the continuous predictors in the data are summarized in Table 1. Against a background of high mosaicity in distribution of *Salinity* and *Fetch* values, the most wind exposed sites were located on the southeastern coast of the Bay, and on the open shores of the islands in its top (Fig. 1 A) while the most desalinated areas - 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 located close to ports tended to have lower *Fetch* (SFig. +++ F) but no association between *DistPort* and *Salinily* was observed (SFig. +++ E). All correlations between *Salinity*, *DistRiver*, *DistPort* and *Fetch* were rather low (STable +++) with maximum correlation between *Fetch* and *DistPort* (r = ++).

~~~~

**Figure 1**. Taxonomic structure of mussel settlements and their habitat characteristics. (A) The map of Northern Europe. Box indicates the location of the Kandalaksha Bay, arrow indicates the location of Tyuva Inlet. B) Surf conditions, point size is proportional to *Fetch* value. Anchor marks with names indicate the ports. Names of abandoned ports are marked by asterisks. (C) Salinity conditions. Point filling is proportional to *Salinity*. Arrows mark the mouths of large rivers. (D-G) Proportion of *MT* in bottom (*PtrosBottom*, D-E) and a*lgal* (*PtrosAlgae*, F-G) samples. Point filling is proportional to *Ptros*. E and G show the top of the bay in higher resolution.

In separate maps of *Ptros* distribution across algal and bottom substrates, the universally elevated proportion of *MT* on the former substrate is striking (**Fig. 2 D-G**). Against a background of high spatial mosaicity in the distribution of *Ptros*, its maximum values on both substrates were observed in the top of the Bay and in some inlets deep into the shore, while its minimum values - along the open part of the southeastern coast (**Fig. 2 D-G**). Associations between *Ptros* and environmental predictors other than substrate are difficult to discern on the maps (**Fig. 2**).

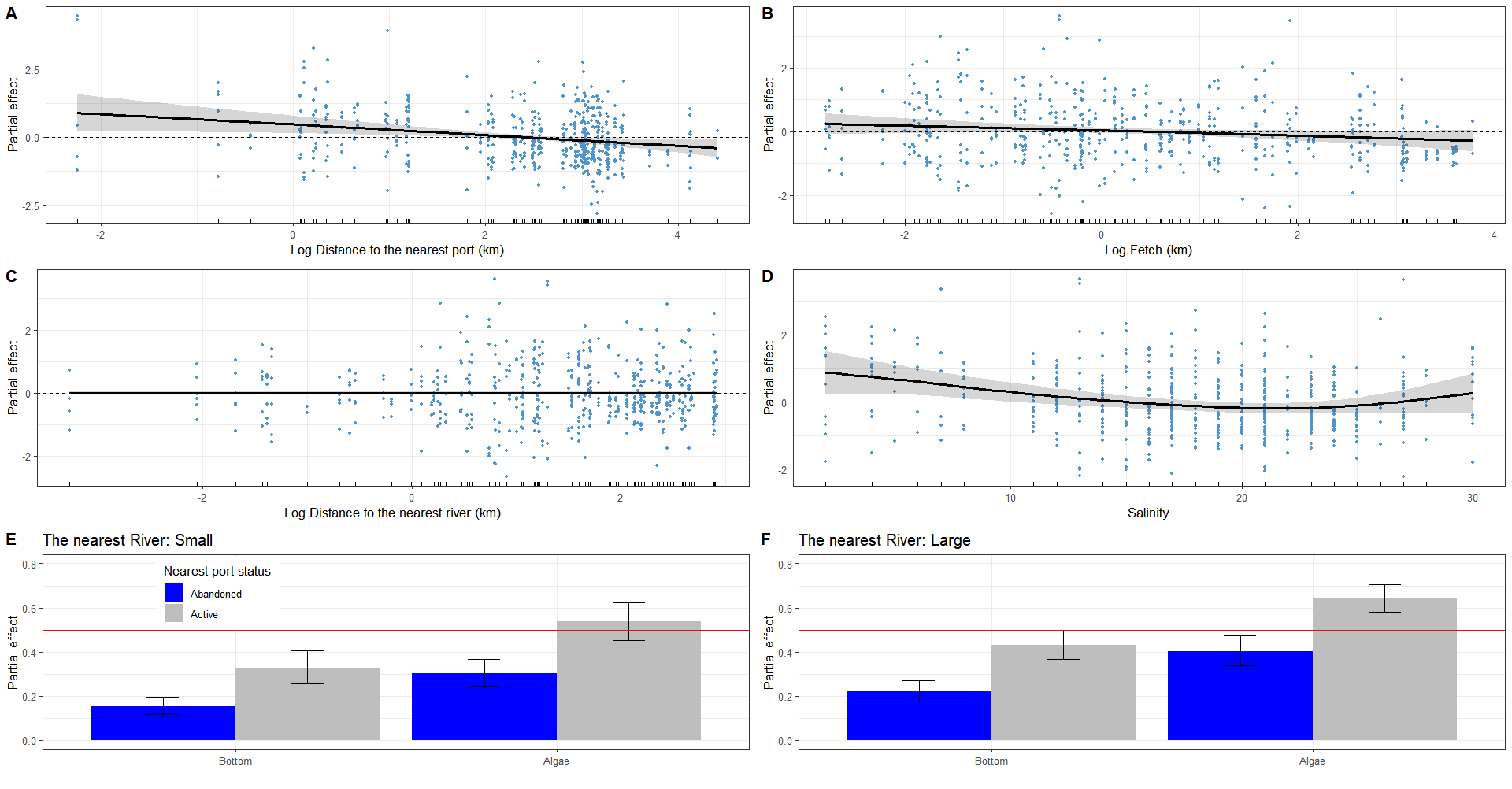
Relationship of *Ptros* to environmental parameters evaluated by Model 1.

Although some non-zero pairwise correlations between environmental factors were found (see above) the low VIF values calculated for the predictors (the maximal VIF was for *Fetch*, 1.76) can be interpreted as a negligible collinearity between them, i.e. predictors did not mask influence of each other.

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

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

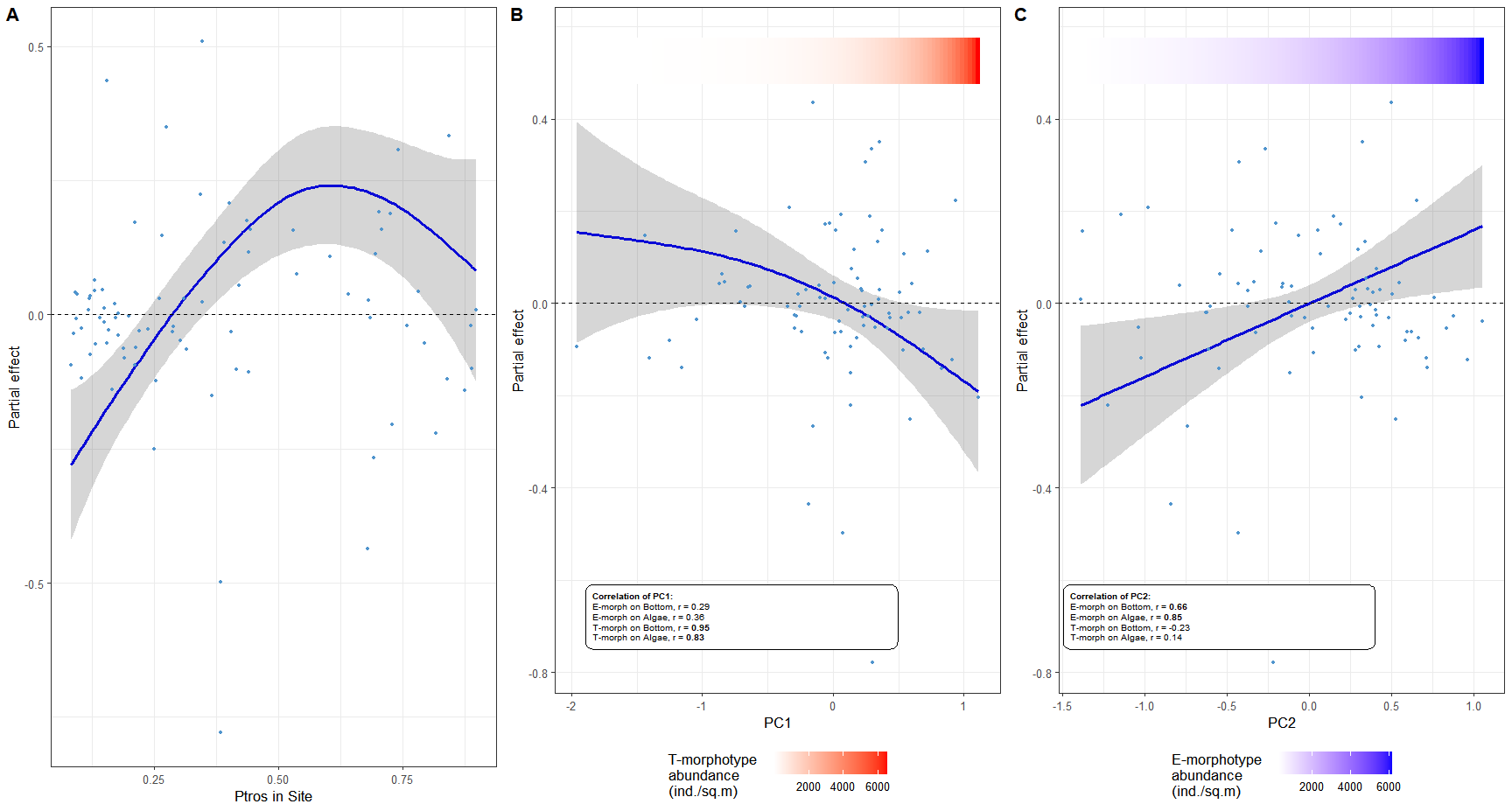
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Smoother terms** | **edf** | **ref.edf** | **Chi.sq** | **p-value** |
| s(Salinity) | 2.4 | 9 | 396.7 | 0.0033 |
| s(DistRiver) | 0.0 | 9 | 0.0 | 0.6724 |
| s(Fetch) | 0.9 | 9 | 88.2 | 0.0417 |
| s(DistPort) | 1.0 | 9 | 276.2 | 0.0016 |
| Random effect 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). The gray ribbons represent 95% confidence intervals. The dotted horizontal lines indicating partial effect equal to zero are given to show the wiggling of the fitted curves. Points on panels A-D reflect partial residuals but not the raw data. E-F. Dependency *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.

According to the model, *Ptros* decreased both with *DistPort* (Fig. 3 A) and with *Fetch* (Fig. 3, B), i.e. 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. Within the range of low to “normal” salinity (24 ppt in the White Sea) the predicted *Ptros* decreases with salinity but again increases at higher salinities (up to 30 ppt in the data) (Fig. 3, D). Further, the predicted *Ptros* was higher near large rivers than near small ones. Finally, the value of *Ptros* was higher on the algal substrate than on the bottom one (Fig. 1 C, D; Fig. 3 E, F). As already mentioned, distance to the nearest river did not affect *Ptros* (Fig 3 C).

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



**Figure 3.** The dependence of difference between proportion of *MT* on algae 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 but not the raw data. The 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).

The PC1 and PC2 of the principal component analysis of the abundance matrix of T- and E- morphotypes on different substrates explained 62% and 20% of variation, respectively. PC1 possessed high positive correlation with abundances of T-morphotypes while PC2 - with abundances of E-morphotypes on both substrates (Fig. 4 B, C). Thus, the abundance of conspecific morphotypes varied on different substrates in a consistent manner (this trend can also be seen in **Fig. 1C, D)**, and hence PC1 and PC2 can be considered as proxies of the *MT* and *ME* abundances, respectively.

Parameters of the *Model 2*, which explained 31% of deviance, are provided in ESM (STable ++). The **Figure 4** demonstrates how the difference between *MT* proportion on algal (*PtrosAlgae*) and bottom (*PtrosBottom*) substrates (*Dif*) depends on the *MT* prevalence in site (*PtrosSite*), and mussel abundances in terms of PCs accordingly to the model. The dependence of *Diff* on *PtrosSite* was significant (p < 0.001, STable ++) and expectedly bell-shaped with minimal values in sites absolutely dominated by *ME* or *MT* (*Ptros* close to 0 or 1) and maximal in sites where both species were equally present (Fig.4 A). The dependence of *Diff* on PC1 was marginally significant (p = 0.087) and demonstrated the tendency to decrease when increasing PC1 (Fig.4 B). The dependence of *Diff* on PC2 was significantly positive (p = 0.011, STable ++) (Fig.4 C). That is, strong species segregation by substrates was observed in sites with high *ME* abundance, but not in sites with high *MT* abundance.

Assessment of the Model 1 predictive power

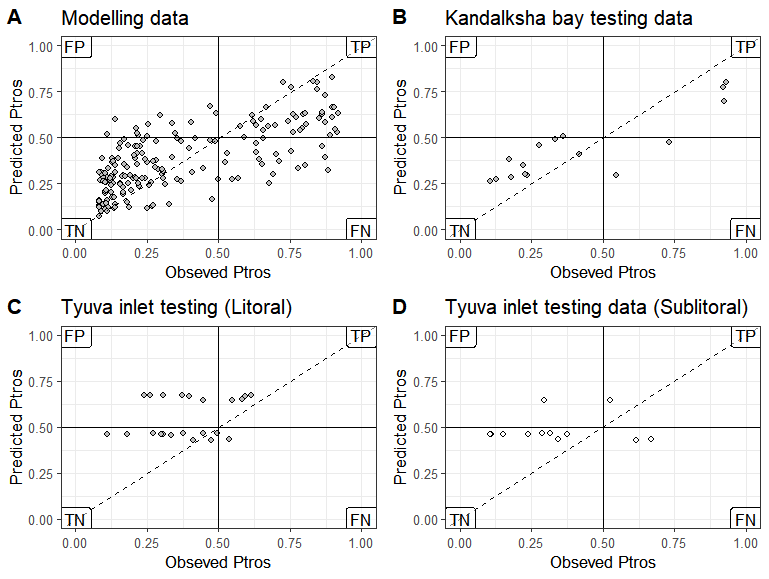


Figure 4. The performance of the SDM (Model 1) to predict proportion of *MT* (*Ptros*) in mussel samples from the modeling (A) and the testing data sets (B-D). Each plot compares empirical *Ptros* in samples from the algal and bottom substrates and that predicted by the model within a particular data set. If the empirical and 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 performance of the model to classify samples into *ME*- and *MT*-dominated ones. Set names are indicated in the chart headers.

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

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

Using an unprecedentedly large amount of material and a species distribution modeling approach (SDM), we demonstrated that such environmental predictors as surf level, distance to the port, status of the port (active vs abandoned), salinity at low tide, size of the nearest river, and substrate of fouling (fucoid algae vs bottom substrates) each influence the distribution of *Mytilus edulis* (*ME*) and *M. trossulus* (*MT*) in the White Sea at scales ranging from meters to tens of kilometers reflecting the partial divergence of their ecological niches.

Below we first discuss what species adaptations could stay behind revealed patterns of *ME* and *MT* distribution against different predictors. Then, the possible role of competition in segregation by substrates. Further, whether the same set of predictors can drive segregation outside the habitat we studied, the littoral fucoid belt, outside the White Sea and outside the Kola contact zone. Finally, we review the pros and cons of our approach to assessing ecological niche partitioning of sympatric mussels.

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

In the light of our findings, the most expected habitat for *ME* in the White Sea littoral is a bottom substrate in a location exposed to surf, with a “normal” for the White Sea surface salinity (24 ppt) and away from ports and big rivers, while for *MT* it is an algal substrate in an area protected from wind, with reduced salinity and close to active ports and big rivers. Of these ecological differences between species in the White Sea, only those related to ports and substrates were previously noticed (VS 11, Katolikova et al. 2016).

*Segregation by salinity*. It is generally recognized that the Baltic mussel (*MT*) is adapted to extremely low salinity (REF +++), which is also confirmed by ecophysiological data (Knöbel et al. 2021). Comparative ecophysiological data on *MT* and *ME* outside the Baltic are non-conclusive and partly contradictory (Gardner & Thompson 2001; Qiu et al., 2002; Sokolova et al. 2024) and, until the present study, it has not been possible to convincingly prove species segregation by salinity in non-Baltic contact zones, including the Kola one (Moreau et al. 2005; REF; Katolikova et al. 2016; Marchenko et al. 2023). For the White Sea, this state of knowledge may be due to at least three reasons, of which the first two may explain the vague relationship in other contact zones as well. First, the presence of other important factors affecting species segregation and masking the role of salinity. Second, the relatively narrow range of salinity in mussel habitats, in comparison with the Baltic Sea. Third, the curvilinear dependence of the proportion of *MT* in mixed settlements (*Ptros*) on salinity, where *Ptros* increases not only at reduced salinity, but also at extremely high, for the White Sea, salinity (up to 30 ppt, **Fig. 2D**). This nonlinearity which may have prevented the dependence from being detected, can be explained in two ways. On the one hand, local summer surface salinity above 24 ppt in Kandalaksha Bay, supposedly associated with irregular episodes of upwelling (Dale & Prego 2003), may be a nonspecific stress for littoral animals adapted to lower salinity, which *MT* can tolerate better (Katolikova et al. 2016, see also below). On the other hand, as detailed studies in the Barents Sea have shown (Khaitov et al. 2021, Marchenko et al. 2023), the method we used to predict *Ptros* (“morphotype test”) may slightly overestimate it at salinities close to 30 ppt. Therefore, at this point we cannot rule out that the increased *Ptros* in sites with high salinity is not an artifact of the method used for assessment of taxonomic structure.

*Non-random distribution relative to ports.* According to the hypothesis of Vainola & Strelkov (2011), the confinement of *MT* to harbors in the White and Barents Seas reflects the history of its invasion into the region - by sea transport from the western Atlantic in the 20th century, which agrees with all genetic data (Simon, Wenne, Laakkonen ...). Katolikova et al. 2016 suggested that *MT* may also be more resistant to anthropogenic pollution and generally be a more opportunistic species, better adapted to disturbed habitats. Our observation that *MT* frequency is lower near abandoned ports than near active ones is consistent with this hypothesis. However, it should be kept in mind that unlike active ports, abandoned harbors may have been more isolated from the propagule pressure of *MT* in recent decades, which may have affected the size of populations of this species.

*Segregation by surf level*. The confinement of *ME* and *MT* to locations with different surf levels among other things may be related to their known differences in the mechanical properties of shells and an ability to form dense aggregations. *ME* has a thicker, heavier and less flexible shell (Beaumont et al. 2008, Michalek et al, 2020), and is more inclined to form tight clumps (Liu et al. 2011) than *MT*. These features of *ME* may be adaptive on the exposed coasts. Unfortunately, there is no comparative data on species differences in byssus secretion and attachment strength, which also may influence their distribution by surf level, as well as across substrates.

*Segregation by substrate.* The same differences in shell structure and in aggregation behavior that may explain the segregation of species by surf may also explain their segregation by substrate. An ability to form dense aggregations is an adaptation to live on the bottom, not on the algae. Other things being equal, *MT* should be lighter than *ME* due to thinner shells (Michalek et al, 2020), which may facilitate their life on algae. Further, fucoid thalluses may serve as shock absorbers for fragile *MT* (Katolikova 2016), as well as shelter them from starfish, which selectively prey on *MT* in mixed settlements (Khaitov et al. 2019, 2023).

*Competition for substrate.* Whatever physiological, morphological, behavioral, etc. features of *MT* and *ME* influence their segregation relative to the environmental factors studied, interspecific competition may also be invoked. We tried to examine the role of mussel abundance in the degree of species segregation across substrates and indeed found that while *MT* abundance had no significant effect on degree of segregation between substrates, *ME* abundance, on the contrary, did: with increased *ME* abundance the degree of segregation increased (Fig. 4 B,C). We interpret this pattern as a consequence of divergence of realized species niches, with *ME* outcompeting MT on bottom substrates displacing it to algal thallomes that appear to be less suitable substrate for *ME* (see above).

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 such small-scale segregation can be found in other attached organisms - terrestrial plants (Raventos et al., 2010). Based on these examples, a “biologically generated spatial pattern” model which relates inter-specific segregation with the intra-specific clustering in competing species has been formulated (Paccala and Levin, 1997; Amarasekare, 2003). This model seems to be applicable for mussels as well.

*The predictive power of the SDM Model.* Since the performance 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), we trust that predictors included in the model explains most of variation in species distribution within the studied habitat - the littoral fucoid belt. In addition, the model also performed satisfactorily with independent data from Tyuva inlet in the Barents Sea (AUC ≈ 0.7), including sublittoral data, emphasizing the versatility of this set of predictors as regulators of the distribution of *ME* and *MT* in the Kola contact zone in general.

We attribute the worst predictive value of the model for the Barents Sea data primarily to the following reasons. First, because of the large depth range of the sampling sites. It is known that *ME* and *MT*  in Tyuva are not randomly distributed by depth (Marchenko et al. 2023). Second, because of the coarser categorization of samples into algal and bottom ones. Since fouling substrate was not taken into account during sampling, we had to predict it based on the projective cover of algae at the sampling site. Thirdly, we still do not know whether species are non randomly distributed across bottom and algal substrates in the sublittoral, where fucoids are replaced by kelps. Finally, the fact that the SDM more frequently overestimated *Ptros* in the Barents Sea data (false positive predictions) is consistent with the observation that the proportion of *MT* was declining in the studied area in the 2010s amidst seemingly stable environmental conditions in terms of predictors included in our model (Marchenko et al. 2023). The latter suggests the presence of additional, as yet unassessed factors regulating taxonomic structure.

Ecological niche partitioning between *MT* and *ME* in the Kola contact zone vs other zones.

Blue mussels represent a challenging model for studying the 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. A set of contact zones between species in the Atlantics can be seen as different ecological (and evolutionary) experiments, set in very different environments (from Baltic to Spitzbergen), at different times (from the late post-glacial to the historical period, VS 11, Wenne et al. 2020 and references therein) and possibly with different “design” in the sense that in some zones one species may have been the original settler and in others another. 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 individualistically in different zones. All this suggests that zones should differ, and this hypothesis has been a leitmotif of some genetic research on blue mussel contact zones (RC, Bierne et al. 2011, Fraisse et al. 2016). We nevertheless believe that differences between species are most fundamental, and thus conspecific ecological phenotypes (“niches”) in different zones should be similar, producing comparable patterns in species distributions. Indeed, some results of our study support this assumption.

The observation that *MT* frequency is elevated in low salinity habitats not only in the Baltic Sea but also in the White Sea seems to resolve the old intrigue about seemingly contrasting salinity adaptations of the Balticand all other Atlantic *MT* populations (e.g. RC 05, Katolikova, VS11, see also above). Further, increased *MT* frequency has been repeatedly observed in calm and freshened waters, such as in the tops of fjords around Bergen in Norway (Ridgway, Nævdal 2004) and Uummannaq in Greenland (Wenne et al. 2016), as well as in Loch Etive in Scotland (Beaumont et al. 2008), which is hardly a coincidence. According to our observations, such a combination of surf and salinity is favorable for *MT* in the White Sea too.

With the exception of the Baltic contact zone and salinity as a factor of segregation, no non-random relationship between species distribution and any of the predictors significant in the White Sea has yet been convincingly demonstrated in the other contact zones. Data on the influence of surf are inconsistent (compare Bates, Innes 1995; Comesaña et al., 1999; Tam and Scrosati, 2013 and this study) and on fouling substrates are, as far as we know, completely absent. If our assumption is correct that the different preferences of *ME* and *MT* for sites with different surf and for different substrates is related to species differences in morphology and behavior (see above), then this should be manifested universally. The intrigue about the increased resilience of *MT* to stress, particularly anthropogenic pollution, as in ports, also remains open (see Brooks et al. 2015 and Beyer et al. 2017 for discussion).

Obviously, 20 years after the classical review of RC 05 on the divergence of ecological niches of *ME* and *MT* in different contact zones, the time has come for a new such survey, and our experience with the Kola zone may be useful here.

Costs and benefits associated with the approaches we used to study sympatric mussels.

The methods we used for taxa identification, environment parameters assessment, and modeling have limitations. For taxa identification we used a “morphotype test” - rapid methodology for assessment of taxonomic structure of mussel settlements without genotyping that works well in habitats with salinity below 25 ppt in the Kola contact zone (Khaitov et al. 2021) but does not allow for direct assessment of species abundances, and the distinction of hybrids as a separate category. The former makes it difficult to account for the role of competition between species, which, judging from our exercise with different substrates, is important. The latter partly excuses the fact that hybrids are relatively scarce in the Kola zone, but in other contact zones between *ME* and *MT* is not the case (Väinölä, Strelkov 2011; Wenne et al. 2020) and hybrids may be an important ecological actors in blue mussel contact zones (e.g. Schwartz et al. 2024).

Although *ME* and *MT* differ universally in morphotype frequencies, the magnitude of the differences varies between contact zones, as well as between habitats with different salinities in the Arctic (Khaitov et al. 2021). Consequently, if the “morphotype test” to be used there, it must be additionally calibrated (see Khaitov et all 2021 for recommendations). Multilocus genotyping, yet still seemingly too expensive for processing dozens of thousands of specimens needed for SDM, is certainly a gold standard of taxonomic assessment in blue mussels.

Of course, 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), and 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 in the water). However, since most of our predictors turned out to be significant, they should not be ignored going forward.

The correlative approach used does not allow for direct assessment of either relative or absolute "importance" of different factors (for example, to say whether salinity or substrate is more important). However, at this stage it is important to recognize that there is no single "leading" factor determining species distribution. This idea, related to the experience of pioneering studies in the Baltic, has dominated *ME* and *MT* studies for many years (RC 05; ???).

All these shortcomings do not detract from the advantages of the SDM approach we pioneered to study distribution of *ME* and *MT* in sympatry. Among the obvious benefits of the approach used are the ability to analyze the distribution of species in the space of multiple predictors simultaneously, the control of predictor collinearity, and no need to treat dependencies as linear.

Prospects for further studies of sympatric mussel species niche partitioning, apart from finding better approaches for SDM construction, in our opinion are the next. Parallel study of different contact zones between species to account for common and zone-specific patterns. Use of taxonomic methods that allow direct assessment of abundances of species, as well as their hybrids, for studying competition between all of them. Incorporating additional environmental factors into the SDM, not forgetting the biotic ones. Accounting for different spatial scales in species segregation within the zones, including the smallest of them.

Some REFS (дублирую из комментариев)

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K., Meier, R., Winker, K., ... & Das, I. (2007). Cryptic species as a window on diversity and conservation. Trends in ecology & evolution, 22(3), 148-155.

Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., … Dimitrov, D. (2018). Finding Evolutionary Processes Hidden in Cryptic Species. Trends in Ecology & Evolution, 33(3), 153–163. doi:10.1016/j.tree.2017.11.007

Geller, J. B., Darling, J. A., & Carlton, J. T. (2010). Genetic perspectives on marine biological invasions. Annual review of marine science, 2, 367-393.

Dufresnes, C., Poyarkov, N., & Jablonski, D. (2023). Acknowledging more biodiversity without more species. Proceedings of the National Academy of Sciences, 120(40), e2302424120.

DeMarche, M. L., Doak, D. F., & Morris, W. F. (2019). Incorporating local adaptation into forecasts of species’ distribution and abundance under climate change. Global Change Biology, 25(3), 775-793.

Hu, Z. M., Zhang, Q. S., Zhang, J., Kass, J. M., Mammola, S., Fresia, P., ... & Zhang, Z. (2021). Intraspecific genetic variation matters when predicting seagrass distribution under climate change. Molecular Ecology, 30(15), 3840-3855.

Gardner, J. P., Oyarzun, P. A., Toro, J. E., Wenne, R., & Zbawicka, M. (2021). Phylogeography of Southern Hemisphere blue mussels of the genus Mytilus: evolution, biosecurity, aquaculture and food labelling. In Oceanography and Marine Biology (pp. 139-228). CRC Press.

Koehn RK. The genetics and taxonomy of species in the genus Mytilus. Aquaculture. 1991; 94 (2–3): 125–145.

Gosling, E. (2021). Marine mussels: ecology, physiology, genetics and culture. John Wiley & Sons.

Zbawicka, M., Sańko, T., Strand, J., & Wenne, R. (2014). New SNP markers reveal largely concordant clinal variation across the hybrid zone between Mytilus spp. in the Baltic Sea. Aquatic Biology, 21(1), 25-36.

Raventós, J., Wiegand, T., & Luis, M. D. (2010). Evidence for the spatial segregation hypothesis: a test with nine‐year survivorship data in a Mediterranean shrubland. Ecology, 91(7), 2110-2120.

Pacala, S. W., & Levin, S. A. (1997). Biologically generated spatial pattern and the coexistence of competing species. Spatial ecology: the role of space in population dynamics and interspecific interactions, 204-232.

Amarasekare, P. (2003). Competitive coexistence in spatially structured environments: a synthesis. Ecology letters, 6(12), 1109-1122.

Pfennig, D. W., & Pfennig, K. S. (2020). Character displacement. Current Biology, 30(18), R1023-R1024.

Lukhtanov, V. A. (2011). Dobzhansky’s rule and reinforcement of prezygotic reproductive isolation in zones of secondary contact. Biology Bulletin Reviews, 1(1), 2-12.

Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular ecology, 22(18), 4606-4618.

Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. Molecular ecology, 20(10), 2044-2072.

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

Ridgway, G., & Nævdal, G. (2004). Genotypes of Mytilus from waters of different salinity around Bergen, Norway. Helgoland Marine Research, 58, 104-109.

Beaumont, A. R., Hawkins, M. P., Doig, F. L., Davies, I. M., & Snow, M. (2008). Three species of Mytilus and their hybrids identified in a Scottish Loch: natives, relicts and invaders?. Journal of Experimental Marine Biology and Ecology, 367(2), 100-110.

Beyer, J., Green, N. W., Brooks, S., Allan, I. J., Ruus, A., Gomes, T., ... & Schøyen, M. (2017). Blue mussels (Mytilus edulis spp.) as sentinel organisms in coastal pollution monitoring: a review. Marine environmental research, 130, 338-365.

Brooks, S. J., Farmen, E., Heier, L. S., Blanco-Rayón, E., & Izagirre, U. (2015). Differences in copper bioaccumulation and biological responses in three Mytilus species. Aquatic Toxicology, 160, 1-12.

Schwartz, L. C., González, V. L., Strong, E. E., Truebano, M., & Hilbish, T. J. (2024). Transgressive gene expression and expression plasticity under thermal stress in a stable hybrid zone. Molecular Ecology, 33(9), e17333.