**Samples.** Mussels were collected from five geographical contact zones between *M. edulis* and *M. trossulus*: Gulf of Maine in NW Atlantics, hereafter GOM (12 populations, N = ++ mussels), Loch Etive in Northern Scotland, SCOT (2 pop., N = 160), the entrance to the Baltic Sea, BALT (8 pop., N = ), Bergen area in Western Norway, NORW ( 3 pop., N = ) and coasts of Kola Peninsula in Northern Russia: 24 populations from the White Sea coast (N=1100) and 26 populations from the Barents Sea coast (N=1650) (Fig. 1). Further two latter data sets were subdivided into some subsets (see bellow). Some samples were taken from available collections of genotyped mussels used in previous studies (GOM: Kingston et al. 2017; BALT: Vainola, Strelkov 2011, Strelkov et al. 2017; 2 of 3 NORW samples: Vainola, Strelkov 2011; all samples from the White Sea, Katolikova et al. 2016) others samples were specially processed for the present investigation (see Fig. 1 and ESM table 1 for sampling details).

The samples from the White Sea and Barents Sea were considered as a main data set for analysis which was aimed to assess the diagnostic value of morphotype-test. Since the diagnostic value of any taxonomic marker could be evaluated only in comparative study involving samplings from contrast species habitats we investigated the associations between mussel genotypes and morphotypes in relation to both regional geography and local salinity conditions along coasts of Kola Peninsula. All, samples were classified by salinity conditions in the sampling localities into brackish (less 30 ppt) and saline water (more 30 ppt) ones. Thus three subsets were considered: the White Sea (all from brackish conditions since the White Sea has salinity below 25 ppt, REF), the Southern part of Kola Bay (all brackish since the top of Bay is an estuary of two large rivers, REF), the Northern part of Kola Bay (saline, REF), and the open Barents Sea coast (?? samples, ? from brackish and ? from saline waters) ones (Fig. 1). Salinity in sampling localities from the open coast was either taken from literature (REF) or, in case of few localities, predicted basing on the presence or absence of large rivers nearby (see ESM table 1 for details).

**Genetic characters.** Mussels collected in Gulf of Maine (GOM) were genotyped at 17 313 random single nucleotide (SNP) loci (Kingston et al. 2017). All mussels from other areas were genotyped by allozyme loci each time including four nearly diagnostic for M. edulis and M. trossulus loci Est-D, Gpi, Pgm and Odh (Vainola, Strelkov 2011; Katolikova et al. 2016; Strelkov et al. 2017). For ?? the Barents Sea samples genotyping by Est-D, Gpi and Pgm was available only (see ESM1 table?). SNP set and each of four regional 4-loci allozyme sets (from Baltic, Norway, Scotland and Russia) were analyzed separately using Structure software (REF, settings as in Katolikova et al. 2016 and Kingston et al. 2017; i.e. number of clusters K=2) and Structure q-values defined as proportion of M. trossulus genes in individual genotypes (proportion of M. edulis genes is 1-q) were estimated. Mussels were classified into two categories by their q-values: genotypes dominated by M. trossulus genes (q-value > 0.5) and genotypes dominated by M. edulis genes (q-value ≤ 0.5). For ease of presentation these categories will be named “M. trossulus” and “M. edulis” hereinafter in spite of the fact that each includes both purebreds and hybrids. We specially checked wether the reduction of allozyme set from 4 to 3 (Est-D, Gpi, Pgm) loci influence the q-value assessment. Since the biases were negligible we used data of geotyping by three loci in the common analysis.

The detailed analyses of hybrid zones under consideration, in particular proportions of purebreds in hybrids in mixed samples, have been provided in previous studies (Шотландская ссылка – поляки? Vainola, Strelkov 2011; Katolikova et al. 2016; Kingston et al 2017; Strelkov et al. 2017, and references therein).

**Morphological characters.** We investigated the inner surface of valves under dissecting stereo-microscope. Mussels were classified into T- and E-morphotypes by the presence/absence of an uninterrupted strip of the prismatic layer under the ligament on the inner side of the shell, respectively (Katolikova et al 2016; Khaitov et al 2018). To note, in previous papers this strip was additionally defined as “dark” since in the White Sea mussels usually possess the dark prismatic layer, and T-morphotypes were illustrated by photos where the strip was both dark and quite wide. Analyses of new material revealed some variation in the coloration and width of the “strip”, that’s why we specify the diagnosis and provide more photos of T-morphotypes (see Results).

Additionally we measured maximum length of each shell to the nearest 0.1 mm with electronic calipers. This data was used to check the influence of size on diagnostic value of morphological marker (see below).

**Statistical analyses**

**Predictive values and analysis rationale**. For each population we calculated frequencies of M. trossulus (Ptros) and T-morphotypes (PT). We further calculated five indexes that measures the validity of species identification by morphotype: P(T|tros) - the proportion of T-morphotypes among M. trossulus, P(E|edu) - the proportion of Е-morphotypes among M. edulis (for practical reasons we also used 1- P(E|edu), the proportion of T-morphotypes among M. edulis), P(tros|T) - the proportion of M. trossulus among T-morphotypes, P(edu|E) - the proportion of M. edulis among E-morphotypes and Pcorrect – the joint proportion of M. edulis with E-morphotypes and M. trossulus with T-morphotypes. The P(tros|T) and P(edu|E) answers the key question how likely is that a randomly taken mussel of T-morphotype is M. trossulus while mussel of E-morphotype is M. edulis, respectively.

It is worth mentioning that indexes similar to proposed are used in clinical medicine for evaluation of the performance of diagnostic tests (REF). If we accept, conditionally, that M. edulis is a “healthy mussel” while M. trossulus is a “seek mussel” (a reasonable assumption bearing in mind its threat of aquaculture (REF)), Ptros would be called prevalence, P(T|tros) - sensitivity, P(E|edu) - specificity, P(tros|T) - positive predictive value, P(edu|E) - negative predictive value, Pcorrect – accuracy. Using the medical jargon, species-specific morphotype could be considered as a “gold standard” i.e. diagnostic test that is the best available under reasonable conditions (REF).

It is usually assumed that for “ideal tests” sensitivity and specificity are independent on prevalence and invariable against it (REF). It means that our case P(tros|T) = 1 and P(edu|E) = 1 in any population. In such case the prevalence and predictive values for any sample could be directly predicted basing on PT in a sample.

However in more general case, when we have to deal with not ideal test (morphotype is a semidiagnostic trait) it is easy to conclude from the Bayes theorem that predictive values have to vary with prevalence: with the increasing Ptros, P(tros|T) would gradually increase from a minimal value (not necessary equal to zero) in pure populations of M. edulis to a maximal value (not necessary equal to one) in pure populations of M. trossulus while P(edu|E) would demonstrate an opposite relationship. In this case the sensitivity and specificity could be calculated using the formulas as follows:

P(tros|T) = Ptros\*P(T|tros)/(1-Ptros)\*(1 - P(E|edu)) + Ptros\*P(Т|tros) [Eq 1]

P(edu|E) = (1-Ptros)\*P(E|edu)/(1-Ptros)\*P(E|edu) + Ptros\*(1 - P(Т|tros) ) [Eq 2]

For the general case if PT, P(E|edu) and P(T|tros) are known the prevalence could be calculated as follows:

Ptros = (PT – (1 - P(E|edu)))/ (P(T|tros) - (1 -P(E|edu)) [Eq 3]

Using these formulas is straightforward if we have accurate assessment of sensitivity, P(T|tros), and specificity, P(E|edu), and these values are stable in all populations. However in reality, the relationships between the prevalence and other indexes usually are more complex (REF) and it is what we can expect for our morphotype-test. Like in the clinical practice there is often a patient spectrum from mild to hard illness that varies with prevalence affecting sensitivity (REF). In the case of mussels there is a spectrum of genotypes underlying M. edulis and M. trossulus as we define them. This spectrum includes various hybrids (e.g. early- or late generation hybrids) that could differ from purebreds and from each other by morphotype frequencies and which quality and quantity could vary with Ptros.

Taking into account the problems described we constructed our analysis accordingly to scheme as follows.

1. Using the material from the White Sea and Barents Sea we investigated the association of morphological marker and its predictive value with Ptros by means of regression analysis (section “Associations among morphotypes and species-specific genotypes around Kola Peninsular”).
2. We checked the applicability of morphotype test to materials collected from BALT, GOM, NORW, SCOT (section “Associations among morphotypes and species-specific genotypes around Atlantics”).
3. Finally, we worked out an approach to apply the equations 1, 2 and 3. This allowed to formulate some practical recommendations on how to make express assessment of mussels genotype and structure of mixed M.edulis + M.trossulus populations in any new contact zone of the species (section “+++++”)

**Associations among morphotypes and species-specific genotypes around Kola Peninsular**

For the purposes of statistical analyses all samples from the Barents Sea open coast (3 samples representing brackish and 6 samples representing saline localities) and four samples with different Ptros from the White Sea (chosen as populations with Ptrost nearest to 0.2, 0.4, 0.6 and 0.8 to maximally represent the prevalence variation) were used as a testing dataset. Other samples - all from brackish areas of the Kola Bay (subset BL), from saline areas of the Kola Bay (subset BH), and all but four from the White Sea (subset W) were used as a modeling data set.

*Analysis of modelling dataset*. All analyses were performed with functions of R3.6.1 statistic programming language (REF). We used generalized linear (mixed) models, GL(M)Ms, with binomial distribution and a logit link-function. For each analysis we first constructed the full models (included all predictors and their interactions) and after they were simplified accordingly to stepwise backward model selection protocol (REF). The model with lowest Akaike information criterion (AIC) was considered as the final one. The function drop1() from the package “stats” was used for the model simplification. The goodness of fit for the final models was assessed by the means of pseudo-R2 (REF) using the function r.squaredGLMM() from the package “MuMIn” (REF). Five next regression models were fitted for the data obtained from the modelling dataset.

Model 1: Morphotype proportions (PT) as a function of taxonomic structure of populations (Ptros). All mussels of T-morphotypes were coded as 1 and of E-morphotypes as 0. These data were used as a dependent variable which was regressed against Ptros (continuous predictor) and Subset (discrete predictor with three levels) and interaction between them. GLM was fitted with glm() function from the package “stats” (REF).

Model 2: Morphotype proportions among species (P(T|tros), P(E|edu)) as a function или functions? of taxonomic structure of populations (Ptros). The dependent variable was coded analogously to Model 1 and was modelled as function of Ptros (continuous predictor), Species (discrete predictor with two levels), Subset (discrete predictor with three levels) and interaction between them. Sample was included into model as random factor influencing the model intercept. GLMM was fitted with glmer() function from the package “lme4” (REF).

Model 3: Accuracy of morphotype-test (Pcorrect) as a function of taxonomic structure of populations (Ptros). The dependent variable was coded as 1 (mussels with T-morphotype correctly identify M. trossulus and mussels with E-morphotype correctly identify M. edulis) and as 0 (an all other cases), and was modelled as a function of Ptros (continuous predictor), Subset (discrete predictor with three levels) and interaction between them. GLM was fitted with glm() function from the package “stats” (REF).

Model 4: Correctness of species identification (P(tros|T) и P(edu|E)) as a function of taxonomic structure of populations. The dependent variable was coded analogously to Model 3. The set of predictors for the model was as follow: Ptros (continuous predictor), Morphotype (discrete predictor with two levels), Subset (discrete predictor with three levels) and interaction between them. Sample was included into model as random factor influencing the model intercept. GLMM was fitted with glmer() function from the package “lme4” (REF).

Model 5: Taxonomic structure (Ptros) as a function of morphotype frequencies in populations (PT). The dependent variable was coded analogously to Model 1 and was modelled as a functionas function of PT (continuous predictor) and Subset (discrete predictor with three levels) and interaction between them. GLM was fitted with glm() function from the package “stats” (REF).

The validity of each model was visually checked by analysis of residual plots. Additionally we analysed the associations of residuals from the Model 5 and mussel shell length to check the presence of some hidden patterns associated with mussel size which potentially can influence the predictive power of the model.

After the model parameters were estimated we visualized them by means of regression lines with corresponding 95% confidence intervals. Since the data coded as “1” or “0” is not informative to show initial data we calculated the proportions of positive outcome for each population and plotted this data as points on model visualization plots. By the same way we represented the testing data sets and considered its correspondence to fitted models.

*Я бы, честно говоря, не стал бы приводить тестовые данные на графиках с моделями, там и так каша, в которой никто не разберется. Но было бы правильно привести один единственный график с моделью №4 и моделью №5, где были бы предсказанные моделью значения VS наблюдаемые в тестинге. Да, плюс один рисунок, но зато все более или менее понятно. Это правильно и сточки зрения валидации моделей, так как важно оценить насколько модель работает/не работает на не включенных в нее данных.*

Finally, we checked would it be possible to pool some subsets to construct more general models without loosing of their predictive value. For this analysis we constructed 3 data sets with different pairing combinations of W, BL and BH (i.e. (WBL) and BH; (WBH) and BL; W and (BLBH); we did not consider full combination of all gradations since in this case the factor “Subset” would be discarded from the model). We applied the structure of models 4 and 5 to these new recombined datasets. Then we compared AIC’s of these new models with AIC’s of initial ones based on not pooled data. If AIC of a new model was less than AIC of initial one we considered this as a basis for joining of corresponding subsets.

**Associations among morphotypes and species-specific genotypes around Atlantics**

Material was divided into 6 subsets: America (GOM), Scotland (SCOT), Baltic (BALT), Norway (NORW), brackish water localities in the White and Barents Seas (WBL; W and BL subsets were pooled, see Results) and saline water localities from Barents Sea (BH). Two models were constructed:

Model 6. Correctness of species identification (P(tros|T) и P(edu|E)) as a function of taxonomic structure of populations (Ptros). Model was constructed analogously to Model 4.

Model 7. Taxonomic structure (Ptros) as a function of morphotype frequencies in populations (PT). Model was constructed analogously to Model 5.

**Assessment of M.trossulus prevalence and predictive values of morphotype-test based on minimal possible number of samples.**

To apply the simple equations (Eq. 1, 2 and 3) using information on mussel morphotypes we have to assess P(T|tros) and P(E|edu), i.e. morphotype-test could not be applied only after some calibration. We assumed that the minimal number of samples (populations) allowing to assess required values and reveal their variation is two.

Basing on the results of analyses of material sampled around Kola Peninsula (see Results) we considered as reference dataset the pooled W and BL data (WBL, ++ populations described). All possible combinations of two different samples from the set were considered. Each pair was characterized by an index of genetic dissimilarity (Delta) between samples combined:

Delta = max(Ptros1; Ptros2) \* (1 - min(Ptros1; Ptros2)) [Eq. 4],

where Ptros1 and Ptros2 - proportion of genetically identified M.trossulus in comparing populations. The index varies in [0; 1]. It takes a value equal to 1 when one sample represent pure M. edulis (Ptros = 0) and another one - pure M. trossulus (Ptros = 1) population. When both samples are equivalent admixtures of two species (Ptros1 = Ptros2 = 0.5) the index is equal to 0.25. If both populations are pure M.edulis (Ptros1 = Ptros2 = 0) or pure M.trossulus (Ptros1 = Ptros2 = 1) the index is equal 0.

For each possible pair of populations from WBL we calculated a set of values as follows. [1] We pooled the data on P(T|tros) and P(E|edu) for each pair. [2] For T- and E-morphotypes we calculated a set of P(edu|E) an P(tros|T) predicted on the basis of Eq.1 and Eq.2 using measured P(T|tros) and P(E|edu) and a set of Ptros values varying from 0 to 1 with the step equal 0.01. [3] For T-morphotype and E-morphotype we calculated predictions of regression Model 6 (coefficients corresponded to WBL were taken) for the same Ptros range. [4] Using Eq. 3 we predicted Ptros for a set of PT values varying from 0 to 1 with the step equal 0.01. [5] For the same range of PT we calculated predictions of regression Model 7.

When the values described were calculated for each possible pair of population from WBL we compared the predictions based on Eq. 1 and Eq. 2 with predictions of Model 6 using the value as follows:

Goodness = 1 / Σ(Regression prediction - Equation prediction) 2 [Eq.5]

Analogously we compared the predictions by Eq.2 and Model 7.

After the Goodness were calculated we plotted them against corresponding Delta values and fitted LOESS regression curve to find associations between them.