**Reviewers:**

Prof. John H. McDonald

Biology Department, Western Washington University, Bellingham, Washington 98225

mcdona44@wwu.edu

Dr. Kati Michalek

SAMS, Scottish Marine Institute, Oban, Argyll, PA37 1QA

[Kati.Michalek@sams.ac.uk](mailto:Kati.Michalek@sams.ac.uk)

Dr. María Angélica Larraín Barth

Departamento de Ciencia de los Alimentos y Tecnología Química, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santos Dumont Nº 964 , Independencia

[mlarrain@uchile.cl](mailto:mlarrain@uchile.cl)

Dr. F.P. (Frank) Wesselingh

Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, Netherlands

frank.wesselingh@naturalis.nl

**Introduction**

Blue mussels *Mytilus edulis* and *M. trossulus* are old evolutionary lineages of Pliocene origin (Riginos, Cunningham 2005). A more common *M. edulis* is thought to be native in the Atlantic, while the basically Pacific *M. trossulus* has colonized the northwest Atlantic in a series of multiple natural and anthropogenic invasions (Rawson, Harper 2009; Vainola, Strelkov 2011; Wenne et al. 2016). Now these two species co-occur and hybridize in at least six geographical areas of the North Atlantic and the adjacent Arctic: Western Greenland, American coast from the Gulf of Maine to Hudson Bay, Northeastern Scotland, Western Baltic Sea, Western Norway and the coasts of the Kola Peninsula in Russia (White Sea, Barents Sea) (Wenne et al. 2020 and references therein).

Ever since the existence of *M. trossulus* was recognized by molecular genetic markers (Varvio et al. 1988), the search for reliable morphometric characters allowing one to distinguish *M. edulis* and *M. trossulus* has been in progress. The discreteness of these two species was confirmed in studies employing numerous metric shell traits and a multidimensional approach, but no individually informative characters have been found (McDonald et al. 1991; Mallet, Carver 1995; Innes, Bates 1999; Telesca et al. 2018). Therefore *M. edulis* and *M. trossulus* are generally treated as cryptic species and are routinely identified genetically. While multilocus analysis is desirable for an unambitious identification of species and their hybrids, in practice singular presumably diagnostic markers are quite often employed, most often the protein coding region for the polyphenolic adhesive protein (ME 15/16 or Glu-5’) (Larrian et al. 2019).

*Mytilus edulis* and *M. trossulus* are ecologically, economically and stratigraphically important molluscs (Seed, Suchanek 1992; FAO 2020; Mangerud, Svendsen 2018). Apart from their biogeographic histories, these two species are known or suspected to differ in life traits, ecological requirements and properties as biomonitoring and aquaculture objects (Lobel et al. 1990; Katolikova et al. 2016; Michalek et al. 2016; Beyer et al. 2017 and references therein). The most illustrative example is the harm associated with *M. trossulus* invasion on longline aquaculture designed for *M. edulis*. A cryptic presence of *M. trossulus* in *M. edulis* plantations in Loch Etive (Scotland) in the 2000s resulted in significant production losses because *M. trossulus* had lower consumer properties and shells too fragile for harvesting and grading (Beaumont et al. 2008; Dias et al. 2011). Considerable differences between species were also found in Canadian aquaculture (Mallet & Carver, 1995; Penny et al. 2002), where the commercial value of *M. trossulus* was estimated to be 1.7 times lesser than that of *M. edulis* (Mallet & Carver 1995). The impossibility of identifying *M. edulis* and *M. trossulus* by the shells is frustrating, and any cue for distinguishing these species in sympatry without genotyping would be a welcome addition to the toolkit of mussel studies.

Our recent studies have shown that *M. edulis* and *M. trossulus* in the White Sea differ by a simple conchological trait (morphotype): the presence or absence of an uninterrupted prismatic strip under the ligament on the inner side of the shell. 74% of *M. trossulus*-like mussels (i.e. mussels with multilocus genotypes dominated by *M. trossulus* genes; the group comprises mostly purebreds and some hybrids) bear a strip (T-morphotype), while 96% of *M. edulis*-like mussels lack this character (E-morphotype) (Katolikova et al. 2016; Khaitov et al. 2018). This finding raises two questions.

The first question is how to apply this marker for individual and population assignment correctly and efficiently. Species identification is usually based on fixed diagnostic traits, which have a unique state for all individuals of a species. The conchological trait under consideration is not diagnostic but semi-diagnostic, i.e. polymorphic within species but with states distributed in different frequencies across species (see Padial et al. 2010). Since there are strong (70%) differences in the morphotype frequencies between the mussel species in the White Sea, one can fall into a trap of deciding that any randomly taken White Sea mussel of T-morphotype can be assigned with a high probability to *M. trossulus* while any mussel of E-morphotype can be assigned to *M. edulis*. In fact, however, the probabilities of correct identification depend on the proportion of *M. trossulus* and *M. edulis* in the population under study. A mussel of any morphotype sampled from a “pure” *M. trossulus* population (expected T-morphotype frequency *PT* = 74%) would be *M. trossulus* anyway. By the same token, any mussel sampled from a “pure” *M. edulis* population (*PT* =4%) would be *M. edulis*. At the same time, in a 1:1 mixture of species (expected *PT* = (74+4)/2 = 39%), 95% of mussels of T-morphotypes would be *M. trossulus* (*P(tros|T)* = 0.74\*0.5/(0.39) = 0.949), while 79% of mussels of E-morphotypes would be *M. edulis* (*P(edu|E)* = 0.96\*0.5/(1-0.39) = 0.787). However, these calculations can be considered as accurate only if the frequencies of the morphotypes within species-specific genotypes do not vary with the taxonomic structure of populations.

In such a situation, taxonomists may profit from the experience of clinicians, who often have to deal with semi-diagnostic characters since many clinical diagnostic tests are semi-diagnostic or considered as such. A formal procedure has been developed in evidence-based medicine to evaluate the ability of clinical tests to classify patients as having or not having the target condition relative to the reference standard (e.g. Banoo et al. 2007). We suggest that this methodology might be useful for the evaluation of semi-diagnostic taxonomic tests for cryptic species relative to the species-specific genotype. To emphasize the analogy with the clinical approach, we denote the procedure of mussel species identification based on the morphotype as a “morphotype test”.

The second question is whether the basic morphological differences between *M. trossulus* and *M. edulis* revealed in the White Sea are a local phenomenon or whether these two species can be distinguished by the morphotype in other populations and contact zones as well. Should the latter prove true, the morphotype test would considerably facilitate local mussel studies in the Atlantic. Since difference between species in the trait under consideration were overlooked in previous morphometric studies that were all based on references from other populations (see references above) it is not improbable that this difference is valid only at the White Sea. The reasons may be associated with the unusual environmental features of the White Sea such as a combination of the subarctic climate and a relatively low salinity (below 25 ppt — Derjugin 1928) and/or the history of the local *M. trossulus*. This species is thought to have invaded the Kola Peninsula through marine traffic very recently, in the middle of the 20th century, while most of its other Atlantic populations are probably much older (Vainola, Strelkov 2011).

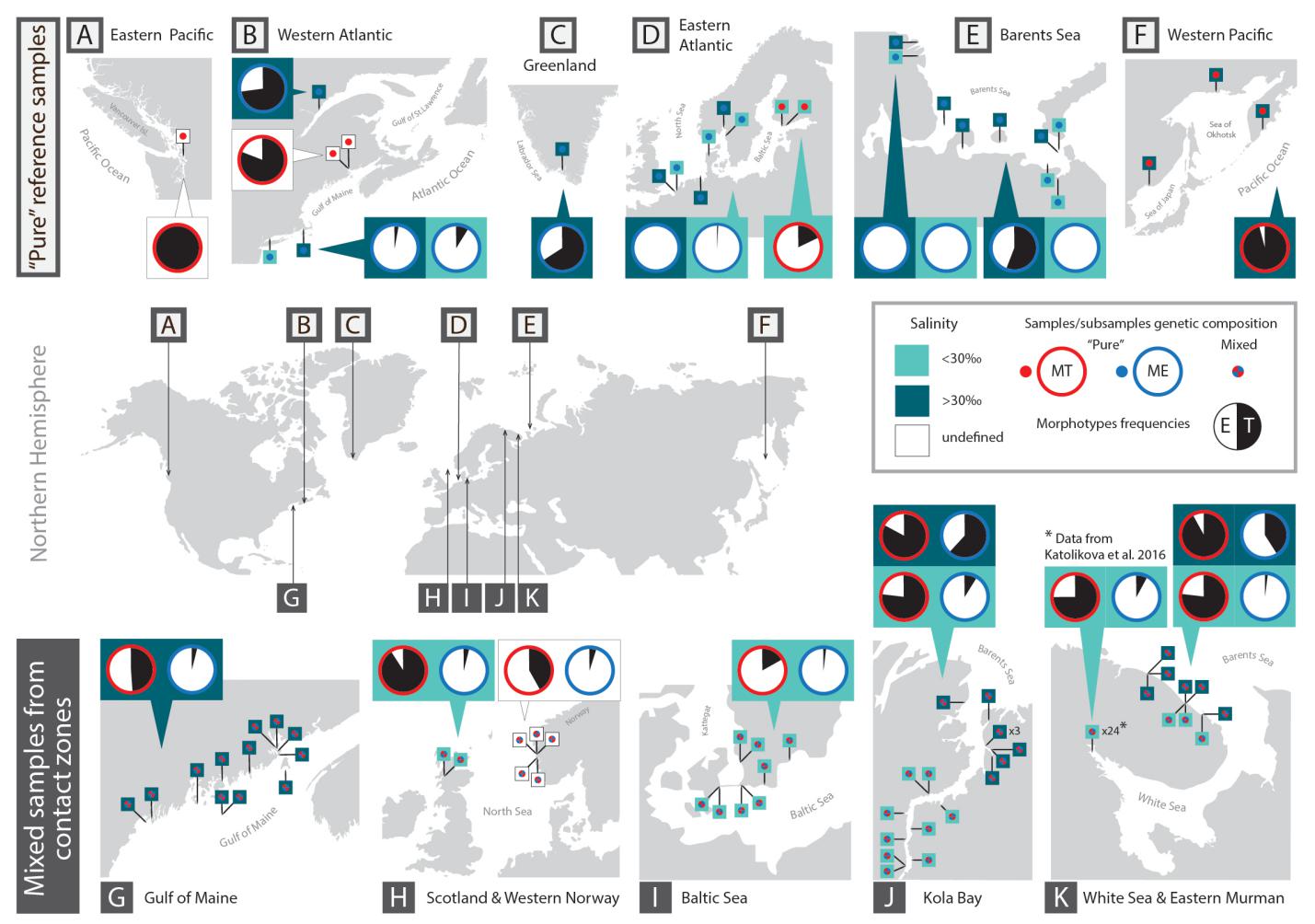
In this paper we address the above two questions. Firstly we analyze the associations between morphotypes and species-specific genotypes in an abundant material from the waters of the Kola Peninsula and in limited material from Norway, the Baltic Sea, Scotland and the Gulf of Maine. For the Kola material, we compare, firstly, populations from the marginal White Sea and from the oceanic Barents Sea coasts and, secondly, populations from the brackish and the saline localities in the Barents Sea. The purpose is to see how local geography and salinity (or associated factors) affect morphotype frequencies in populations with a similar biogeographic history existing under similar climatic conditions. Secondly, we formally evaluate the performance of the “morphotype test” for species identification using approaches from evidence-based medicine and provide practical recommendations for the use of the test for population and individual assessment.

**Materials and Methods**

**Samples.** Altogether, we considered 77 mussel samples (total sample size N = 4325, individual sample size N=18-173) representing five geographical contact zones between *M. edulis* and *M. trossulus*: the Gulf of Maine in the NW Atlantic (12 populations, N = 428), Loch Etive in Northern Scotland (2 populations, N = 160), Western Baltic Sea (8 populations, N = 638), Bergen city area in Western Norway (5 populations, N = 365) and the coasts of Kola Peninsula in Northern Russia: 24 populations from the White Sea (N =1089) and 26 populations from the Barents Sea (N = 1645). Detailed information about samples and sampling localities is provided in the **S1 Table** .

The Barents Sea samples were taken in the Kola Bay and at the open oceanic coast of Eastern Murman. Based on the salinity in the sampling localities, they were classified into brackish (salinity 5-30 ppt) and saline (>30 ppt). The first group consisted of nine samples from the freshened top of the Kola Bay and three samples from the open coast. The second group consisted of eight samples from the mouth of the Kola Bay and six samples from the open coast (**Fig. 1**). As for samples from other contact zones, all American samples and two of five Norwegian samples were from saline habitats, all the other were from brackish habitats. Salinity conditions in the sampling localities was either taken from the literature (Derjugin 1915; Ridgway, Nævdal, 2004; Bobkov et al. 2010; Dias et al. 2009; Kingston et al. 2017; Shavykin 2018) or, in case of the few American and the Barents Sea open coast localities, predicted based on the presence or absence of large rivers nearby.

In addition to the samples taken in the five target contact zones, we identified the morphotypes in 27 samples (total sample size N=912, individual sample size N=12-76) of supposedly pure blue mussel species from distant localities: *M. trossulus* from Passamaquoddy Bay and *M. edulis* from the Gulf of Saint Lawrence in Eastern Canada, *M. trossulus* from the Northern Baltic Sea, from Puget Sound in Eastern Pacific and from multiple areas of Western Pacific, *M. edulis* from SW Greenland, from the Long Island Sound and Cape Cod in the USA, and from saline and brackish localities in Europe and in the SW Barents Sea (**Fig.1**, **S2 Table**). Information about species identity of regional populations and salinity conditions in sampling localities was taken from literature. Taxonomic affinities of mussels from Canada and some of Norwegian samples, where both species could be expected, were confirmed genetically (see **S2 Table** for details).



**Fig 1.** Map of study area. Bottom maps (G-K) represent five geographical contact zones between *M. edulis* and *M. trossulus* indicated, upper maps (A - F) - other studied areas. Pins depict sampling sites. Pie diagrams depict proportions of T-morphotypes (black sector) and E-morphotypes (white sector) in *M. trossulus* (diagrams with red border) and *M. edulis* (with blue border) in combined samples from particular regions. When data on salinity in sampling localities are available, it is indicated by the color of pins (light green – brackish, dark green - saline) and proportions of T-morphotypes in combined samples from brackish and saline localities are presented separately in diagrams placed on light- and dark green background respectively. Source data are in **S1 Table** and **S2 Table**.

**Genetic characters.** Some samples from the contact zones were genotyped in previous studies (8 of 12 American samples: Kingston et al. 2017, Martino et al. 2019; all Baltic samples: Vainola, Strelkov 2011, Strelkov et al. 2017; 2 of 5 Norwegian samples: Vainola, Strelkov 2011; all the White Sea samples, Katolikova et al. 2016). The other samples were taken specially for the purpose of this study (see **S1 Table**). For mussels from published studies multilocus nuclear genotypes were available. The Gulf of Maine mussels were genotyped using 171 645 random SNPs (single nucleotide loci) (Kingston et al. 2017), while mussels from other areas were genotyped using sets of allozyme loci each time including four “nearly diagnostic” (70–95% allele frequency differences between *M. edulis* and *M. trossulus*) Est-D, Gpi, Pgm and Odh loci. New samples from the Gulf of Maine were genotyped as in Kingston et al. 2017, while other samples were genotyped by Est-D, Gpi, Pgm and Odh as in Katolikova et al. 2016. For seven samples from the Barents Sea the data on only three loci—Est-D, Gpi and Pgm—were available (see **S1 Table**). SNP set and each of the four regional 4-locus allozyme sets (from the Baltic, Norway, Scotland and Russia) were analyzed separately using STRUCTURE or fastSTRUCTURE software (Pritchard et al. 2000, Raj et al. 2014, settings as in Katolikova et al. 2016 and Kingston et al. 2017). Structure q-values defined as proportion of *M. trossulus* genes in individual genotypes were estimated (proportion of *M. edulis* genes is therefore 1-q). The material from Russia was also analyzed by three loci (Odh not considered) to show that the exclusion of Odh did not affect the inference (data not shown). 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 referred to as “*M. trossulus*” and “*M. edulis*” genotypes though each includes both purebreds and hybrids. Other genetic information but results of classification into *M. trossulus* and *M. edulis* will not be considered here. A detailed analysis of the hybrid zones under consideration, in particular, the proportions of purebreds and hybrids in mixed samples, are available in the literature (Vainola, Strelkov 2011; Katolikova et al. 2016; Kingston et al 2017; Strelkov et al. 2017; Wenne et al. 2020 and references therein).

**Morphological characters.** Data on the White Sea samples were taken from Katolikova et al. 2016 and the other samples were processed accordingly. We measured the maximum length of each shell to the nearest 0.1 mm with electronic calipers and investigated the inner surface of the valves under a dissecting stereo-microscope. The mussels were classified as having a T- or an E-morphotype based on, respectively, presence or absence of an uninterrupted strip of the prismatic layer under the ligament on the inner side of the shell. To note, this strip was additionally defined as “dark” in our previous papers (Katolikova et al 2016; Khaitov et al 2018) since mussels from the White Sea usually possess a dark prismatic layer, and T-morphotypes were illustrated with photos where the strip was both dark and quite wide. The analysis of the new material revealed some geographical variation in the coloration and width of the “strip”. We specify the definitions of the two morphotypes in the Results and provide more illustrations in the ESM.

**Predictive values.** For each sample we calculated the frequencies of *M. trossulus* (*Ptros*) and T-morphotypes (*PT*) and four indices reflecting the strength of association between genotypes and morphotypes: *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 used *P(T|edu)*=*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. *P(tros|T)* and *P(edu|E)* are the key indices because they show, respectively, how likely it is that a randomly taken mussel of the T-morphotype is *M. trossulus* and a randomly taken mussel of the E-morphotype, *M. edulis*

Here we would like to offer an analogy between the indices used in our study and those used in clinical medicine for the evaluation of the performance of diagnostic tests. If we consider *M. edulis* as a “healthy” mussel and *M. trossulus* as a “sick” mussel (which is not so far-fetching considering the threat presented by *M. trossulus* to the Scottish aquaculture, Beaumont et. al. 2008), then our terms have the following medical equivalents (Banoo et al. 2007): *Ptros* is *prevalence*, *P(T|tros)* is *sensitivity*, *P(E|edu)* is *specificity*, *P(tros|T)* is *positive predictive value* and *P(edu|E)* is *negative predictive value* of the morphotype test.

It is axiomatic that for semi-diagnostic tests positive and negative predictive values vary with prevalence (Banoo et al. 2007). With the increasing *Ptros*, *P(tros|T)* will gradually increase from 0 in pure populations of *M. edulis* to 1 in pure populations of *M. trossulus*, while *P(edu|E)* will demonstrate an opposite relationship. For the test to be meaningful, predictive values should be >0.5 since a predictive value of 0.5 indicates a random association between the genotype and the morphotype. Assuming that sensitivity and specificity do not depend on the prevalence (note that this assumption could be violated, see below), predictive values could be directly predicted basing on the *Ptros* in a sample and the known sensitivity and specificity using formulas:

*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]

In its turn, the prevalence (*Ptros*) could be predicted based on *P(E|edu)*, *P(T|tros)* and *PT* in a sample:

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

**Statistical analyses.** The following analyses were made using the data from the contact zones. Firstly, we studied variation of *PT*, *P(T|tros)*, *P(T|edu)*, *P(tros|T*), *P(edu|E)* as functions of *Ptros* within and between sample sets representing A) the White Sea (sample set *WS*) and the Barents sea coasts of the Kola Peninsula and saline (set *BH*) and brackish (set *BL*) water localities in the Barents Sea (Section “Associations between morphotypes and species-specific genotypes around the Kola Peninsula”), B) different geographical contact zones between species. Whenever possible, formulas describing empirical relationships between *Ptros* and *PT* and between positive (*P(tros|T)*)and negative (*P(edu|E)*) predictive values and *Ptros* have been derived on the basis of regression analysis (Section “Associations between morphotypes and species-specific genotypes around the Atlantic”). Secondly, we analyzed genotype-specific associations between morphotypes and the shell size in order to verify the hypothesis that morphological variation under consideration is not related to mussel size (Section “Associations between morphotypes and shell size”). Finally, we tested how well the *Ptros,* *P(edu|E)* and *P(tros|T)* could be predicted using formulas Eq 1-3 and the data on morphotype proportions among species (*P(T|tros)*, *P(T|edu)*) in only two genotyped samples. We concede that the assumption that sensitivity and specificity do not depend on the prevalence can be violated in the morphotype test, as it often is in clinical tests (Leeflang et al. 2009, 2013). Therefore our interest was focused on finding out which samples are better suited for prediction on the basis of Eq. 1-3 and consequently could be used as “calibrating” ones: the most mixed samples (*Ptros*~0.5) or the combination of the two most pure samples of each species (Section “Prediction of taxonomic structure of populations and predictive values of the morphotype test based on calibrating samples”).

All statistical analyses were performed with functions of R3.6.1 statistical programming language (REF). We used generalized linear (mixed) models, GL(M)Ms, with binomial distribution and a logit link-function. All GLM models were constructed with glm() function from the package “stats” (REF) whereas GLMM were fitted with glmer() function from the package “lme4” (REF). The validity of each model was checked by visual analysis of residual plots and the assessment of the overdispersion presence.

The goodness of fit for the models was assessed by means of pseudo-R2 (REF) using the function r.squaredGLMM() from the package “MuMIn” (REF). To assess the role of random factors in GLMM, we compared marginal and conditional pseudo R2 (REF). After the model parameters were estimated, we visualized them by means of regression lines with corresponding 95% confidence intervals.

**Associations between morphotypes and species-specific genotypes around the Kola Peninsula***.* The following three regression models were fitted for the data.

**Model 1**: Morphotype proportions (*PT*) as a function of taxonomic structure of mussel populations (*Ptros*). All mussels with a T-morphotype were coded as 1 and all mussels with an E-morphotype were coded as 0. These data were used as a dependent variable, which was regressed against *Ptros* (continuous predictor) and *Set* (discrete predictor with three levels) and interaction between them.

**Model 2**: Morphotype proportions among species (*P(T|tros)*, *P(T|edu)*) as a function of taxonomic structure of populations (*Ptros*). The dependent variable was coded as in Model 1 and modelled as a function of *Ptros*, *Set*, *Species* (a discrete predictor with two levels) and interaction between terms. The sample was included into the model as a random factor influencing the model intercept.

**Model 3**: Correctness of species identification (*P(tros|T)* and *P(edu|E)*) as a function of taxonomic structure of populations. The dependent variable was coded as 1 if *M. trossulus* had a T-morphotype or *M. edulis* had an E-morphotype and as 0 in the other cases. The set of predictors for the model was as follows: *Ptros*, *Morphotype* (discrete predictor with two levels), *Set* and interaction between terms. The sample was included into the model as a random factor influencing the model intercept.

To check whether it is possible to pool some of the geographical sets to construct a more general model without losing information, we constructed three complex data sets with different pairing combinations of *WS*, *BL* and *BH*: (*WSBL*) and *BH*; (*WSBH*) and *BL*; (*BLBH*) and *WS*. We did not consider a full combination of sets since in such a case the factor “*Set*” would be discarded from the model. We applied the structure of Model 3to these new recombined datasets. Then we compared AICs of these new models with AIC of Model 3 based on non-pooled data. If AIC of a new model was less than the AIC of the initial one, we considered this as a basis for pooling of the corresponding sample sets.

**Associations between morphotypes and species-specific genotypes around the Atlantic**. Five sample sets were considered, representing the Gulf of Maine (*GOM*), the Baltic Sea (*BALT*), Western Norway (*NORW*), saline Barents Sea (*BH*) and the White Sea combined with the brackish Barents Sea (*WSBL*, sets *WS* and *BL* were pooled since they did not differ statistically, see Results). Scotland (*SCOT*) was not included in regression analyses because it was represented by two samples only. Three models were constructed:

**Model 4.**  Taxonomic structure (*Ptros*) as a function of morphotype frequencies in populations (*PT*). The dependent variable was coded as in Model 1 and modelled as a function of PT (continuous predictor), *Set* and interaction between them. We modeled *Ptros* vs. *PT* but not vice versa, as in the previous analysis, in order to use this model as a reference for the “*Ptros* by *PT* calculator” (see below).

**Model 5.**  Morphotype proportions among species (*P(T|tros)*, *P(T|edu)*) as a function of taxonomic structure of populations (*Ptros*). The model was constructed analogously to Model 2.

**Model 6.** Correctness of species identification (*P(tros|T*) and *P(edu|E)*) as a function of taxonomic structure of populations (*Ptros*). The model was constructed analogously to Model 3.

**Associations between morphotypes and shell size.** To check the possible association of morphotypes with size we undertook the following two analyses. Firstly, we constructed a set logistic regression models for each available species-specific genotype (i.e. *M. edulis* or *M. trossulus*) from each sample. The probability of the presence of the T-morphotype was a dependent variable and mussel size was a predictor in these models. Only cases where slope-terms of the models were statistically significant (p < 0.05) after Hochberg’s correction for multiple testing (REF- Quinn and ...) were considered. Secondly, we checked the presence of any patterns in residuals from Model 6 (i.e. the main model designed to predict the probability of correct identification of an individual mussel by its morphotype) as a function of mussel size.

**Prediction of taxonomic structure of populations and predictive values of morphotype test basing on calibrating samples**

We applied Eq. 1-3 to predict *Ptros*, *P(edu|E)* and *P(tros|T)* for samples from each data set (*GOM*, *BALT*, *NORW*, *BH*, *WSBL*, *SCOT*) using estimates of morphotype proportions among species (*P(T|tros*), *P(T|edu)*) obtained from pooled samples from each set and whenever possible, from combinations of two calibrating samples selected based on the results of the following analysis.

To work out the strategy of calibrating samples selection, we considered the *WSBL* (36 samples) as a reference dataset. All 630 possible pair combinations of samples were considered. Each pair was characterized by an index of taxonomic similarity between the samples:

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

where *Ptros1* and *Ptros2* – higher and lower estimates of prevalence in samples. The index varies in a range [0; 1] and takes the value Delta=0 when both samples are pure *M. edulis* (*Ptros1* = *Ptros2* = 0) or pure *M. trossulus* (*Ptros1* = *Ptros2* = 1), Delta=0.5 when both samples are equivalent mixtures of two species (*Ptros1* = *Ptros2* = 0.5) and Delta=1 when one sample represent pure *M. trossulus* (*Ptros1* = 1) and another pure *M. edulis* (*Ptros2* = 0).

Estimates of *P(T|tros)*, *P(E|edu)* and *PT* were obtained from pooled data on each pair of samples and used for calculation of predicted values of *P(edu|E)* and *P(tros|T)* basing on Eq.1,2 for the range of *Ptros* [0;1] with the step 0.01 (“genotype by morphotype calculator”) and predicted values of *Ptros* basing on Eq.3 for the range of *PT* [0;1] with the step 0.01 (“*Ptros* by *PT* calculator”). Values of *P(edu|E)* and *P(tros|T)* obtained by the Eq. 1, 2 were contrasted with those ones predicted by the Model 6 and values of *Ptros* obtained by Eq. 3 were compared with predictions of the Model 4 using of correspondence statistics:

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

Goodness indices for each pair were plotted against the corresponding Delta values and the LOESS regression curve was fitted to find associations between them. Depending on the results of the analyses, we determined which samples could be used for predictions with best results: the most mixed samples (*Ptros1* ≈ *Ptros2* ≈ 0.5) or the combination of two most pure samples of each species (*Ptros1* ≈ 1; *Ptros2* ≈ 0).

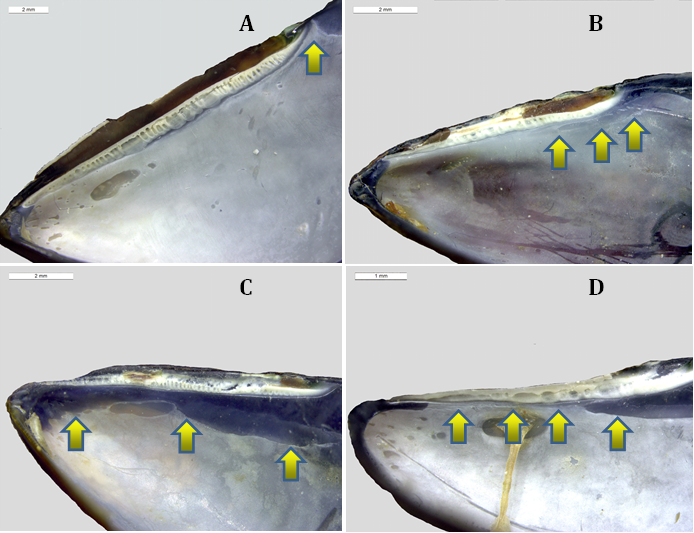
For illustrative purposes and for the convenience of users of “morphotype test” or any similar semi-diagnostic tests we provide the online “*Ptros* by *PT*” and “genotype by morphotype” calculators implementing Eq. 1-3 at +++++.

**Results**

**Geographical variation in the manifestation of mussel morphotypes.** The binary morphological character under consideration was previously defined, based on the material from the White Sea, as the “presence/absence of a distinct uninterrupted dark prismatic strip under the ligament” (Katolikova et al. 2016; Khaitov et al., 2018). In this study, which was based on the material from different geographical zones, E-morphotypes in all the populations looked the same and conformed to our previous description: the strip was absent, and the nacreous layer totally or partially covered the space under the ligament nympha (**S1 Fig A, C**). However, T-morphotypes showed some variation previously unrecorded in the White Sea. Firstly, most populations examined in this study contained, though rarely, shells in which the nacreous-free strip of the prismatic layer was quite narrow and looked like a stria rather than a strip. Secondly, in all T-morphotypes from the Gulf of Maine populations and in rare T-morphotypes from the other populations the strip was not dark but pale, as the prismatic layer itself. In such cases, T-morphotypes were difficult to notice by an unaided eye. They could be unambiguously identified only with the help of a dissecting microscope by the presence of a pronounced scar defining the boundary of the nacreous layer under the ligament nympha (**S1 Fig**).

Therefore, we propose an amended description of the character used to distinguish the E-morphotype and the T-morphotype: the presence/absence of an uninterrupted strip of the prismatic layer under the ligament nympha clearly recognizable by a scar separating the strip from the nacreous layer of the rest of the shell. This description was applicable to all the mussel populations examined in this study.

**S1 Fig.** Variation in the manifestation of mussel morphotypes. A-D. Stereoscopic micrographs of the ligament area of mussel valves. Note that scale bars differs between A-C and D. Strip of the prismatic layer under the ligament nympha is indicated by the arrows. A, B. E-morphotypes: the space under the ligament nympha is totally (A) or partially (B) covered by the nacre. C, D. T-morphotypes: a strip of uncovered prismatic layer under the ligament nympha is dark and wide (C; typical of most examined populations) or pale and narrow, recognizable by a scar separating it from the nacreous layer (D; typical of the Gulf of Maine populations). E. External and internal features of the shell valves of *M. trossulus* (left) and *M. edulis* (right) genotypes from Kola Bay (all mussels from samples ?? and ??). In most cases T- morphotypes (marked by \*) and E-morphotypes could be distinguished by an unaided eye.



**Associations between morphotypes and species-specific genotypes around the Kola Peninsula**

Variation patterns of *PT*, *P(T|tros)*, *P(E|edu)*, *P(tros|T)*, *P(edu|E)* as functions of *Ptros* in samples from the White Sea (*WS*), the brackish Barents Sea (*BL*) and the saline Barents Sea (*BH*) are visualized in **Fig. 2.** The results of the regression analysis are summarized in **Table 1**.

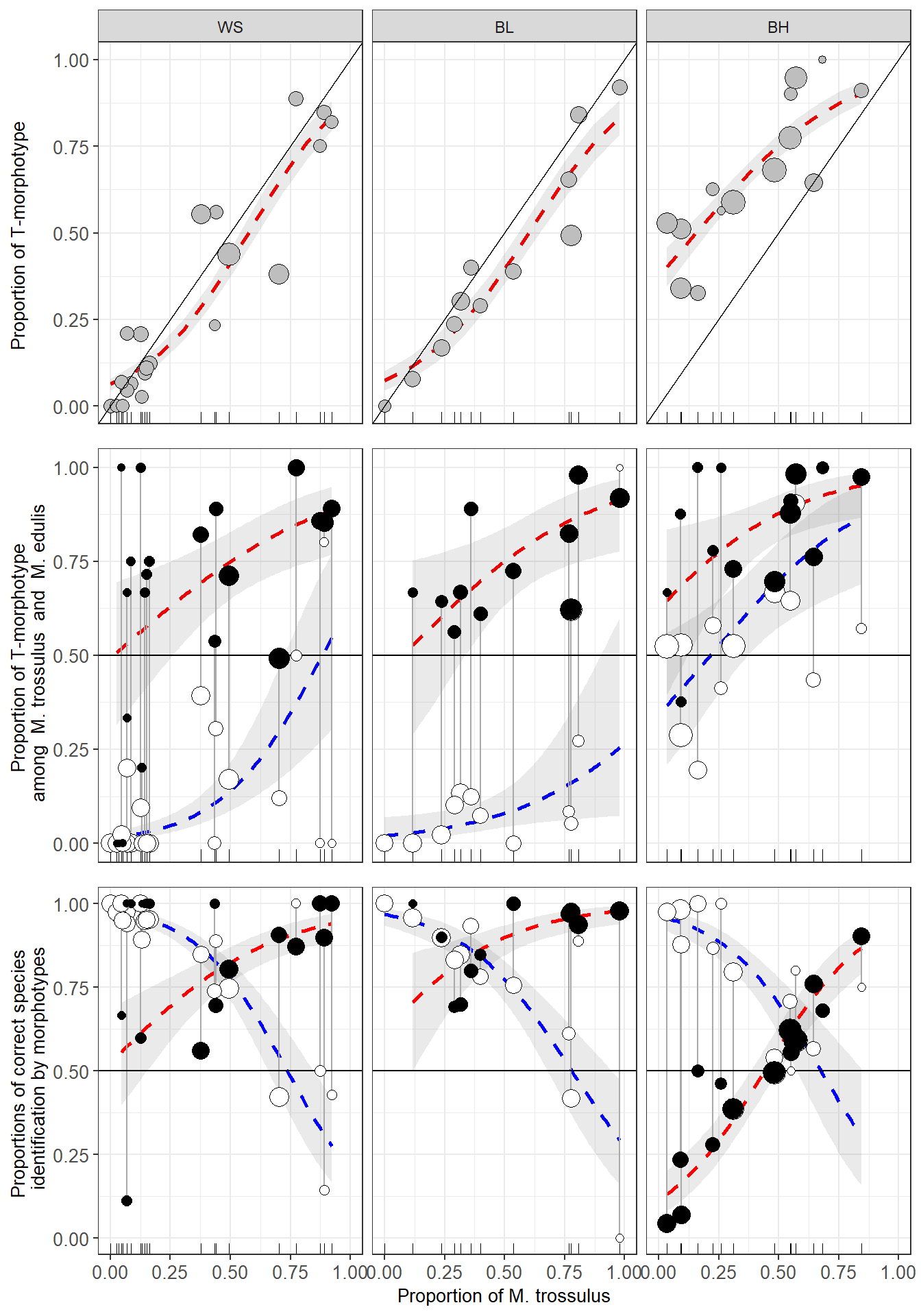
A significant positive association between the proportions of *M. trossulus* (*Ptros*) and the proportions of T-morphotypes (*PT*) in samples was revealed for all the three sample sets (Model 1, **Table 1**, **Fig. 2**). For *WS* and *BL*, the data points were generally scattered around the Y=X line, while the regression lines approached it closely, indicating a high proportionality between *Ptros* and *PT*. For BH, the data points were scattered above the Y=X line and the regression line lay higher that the regression lines constructed for *WS* and *BL*. This means that in samples with a similar taxonomic structure, the frequencies of T-morphotypes were always higher in the saline localities in the Barents Sea than in the White Sea and the brackish localities in the Barents Sea.

The analysis of the frequencies of T-morphotypes in subsamples of *M.edulis* (*P(T|tros)*) and *M.trossulus* (*P(T|edu)*) against proportions of *M. trossulus* in samples (*Ptros*) revealed the following patterns (Model 2, Table 1, Fig. 2). There was a universal tendency towards a higher frequency of T-morphotypes among *M. trossulus* than among *M. edulis*. This tendency was quite strong in *WS* and *BL* (expected differences in morphotype frequencies between species about 65% for *Ptros*=0.5). In *BH* it was rather weak (expected differences 18% for *Ptros*=0.5) due to an increased *P(T|edu)* but significant (confidential intervals for *Ptros*=0.5 did not overlap, Fig. 2). A positive correlation of *P(T|tros)* and *P(T|edu)* with *Ptros* was found in all the three subsets. This means that with the increasing contribution of *M. trossulus* to the samples the frequencies of T-morphotypes increased (??- ??% for the range *Ptros* [0, 1]) both among *M. edulis* and among *M. trossulus*.

The probability of correct identification of *M. trossulus* by the T-morphotype (the frequency of *M. trossulus* among T-morphotypes, *P(tros|T)*) expectedly increased with the increasing *Ptros*, while the probability of correct identification of *M. edulis* by the E-morphotype (*P(edu|E)*) demonstrated an opposite pattern (Model 3, Table 1, Fig. 2). In the *M. trossulus* - dominated populations, *P(tros|T)* tended to one (any mussel with a T-morphotype is 100% *M. trossulus*), while *P(edu|E)* tended to zero (any mussel with an E-morphotype is 100% *M. trossulus*), and vice versa. In the well-mixed samples (*Ptros* = 0.5) the predictive values for both species was about 0.75-0.85 in WS and BL but only 0.6 - 0.7 in BH (Fig. 2). It means that the morphotype test has a much lower predictive value in the saline Barents Sea than in the brackish Barents Sea and in the White Sea (the predictive value of 0.5 means a random association between the genotype and the morphotype). It is evident from **Fig. 2** that a low predictive value of the test in *BH* is mainly due to a generally low *P(T|tros)*: even though the great majority of *M. trossulus* have a T-morphotype, it is difficult to recognize them because many *M. edulis* have this morphotype, too. On the other hand, E-morphotypes are predominantly found in *M. edulis*. Nevertheless, the statistical analysis indicates that both *P(tros|T)* and *P(edu|E)* predicted by the model were smaller in BH than in WS and BL.

For each of the GLMM models considered (Model 2 and 3), marginal and conditional pseudoR2 were close to each other (**Table 1**). This indicates that the role of the random factor (*Sample*) as regulator of models was weak, i.e. the reproducibility of the results in different populations was satisfactory.

In intra-set **может быть все-таки inter-set** comparisons, the regression coefficients did not differ statistically for *WS* and *BL* sets, while *BH* was always different from *WS* (Table 1). To assess the possibility of pooling the data sets, we compared the AIC of full Model 3 (?????? ) with AICs of three other models based on differently pooled WS, *BL* and *BH* sets. The model based on pooled *WS* and *BL* (*WSBL*) and *BH* showed the lowest AIC. Therefore, in the following analyses we will consider two sets, *WSBL* and *BH*.



**Fig 2.** Variation of *PT*, *P(T|tros)*, *P(E|edu)*, *P(tros|T)*, *P(edu|E)* as functions of *Ptros* in the White Sea (WS), brackish Barents Sea (BL) and saline Barents Sea (BH). Points - empirical estimates, their size is proportional to sample size. Lines – regression model predictions, grey filling – 95% confidence intervals of regressions. (A) Proportions of T-morphotypes (*PT*) (Model 1). (B). Proportions of T-morphotypes among *M. trossulus* (*P(T|tros)*, filled points) and *M. edulis* (*P(T|edu)*, empty points) (Model 2). (C) Frequencies of *M. trossulus* among T-morphotypes (*P(tros|T)*, filled points) and of *M. edulis* among E-morphotypes (*P(edu|E)*, empty points) (Model 4). Vertical lines on B and C connect subsamples of *M. trossulus* and *M. edulis* from the same samples.

**Table 1.** Parameters of the fitted regression models.

**Associations between morphotypes and species-specific genotypes around the Atlantic**

The patterns of *Ptros* variation against *PT* and the patterns of *P(T|tros)*, *P(E|edu)*, *P(tros|T)* and *P(edu|E)* variation against *Ptros* in samples from different geographical zones are visualized in Fig. 3. The results of the regression analysis are summarized in Table 1. The Scottish material was not included in the regression analyses. Re-analyses of the data from the White and the Barents Sea (*WSBL* and *BH* sets) together with the data from other regions revealed the same patterns as those described above. Again, in all the cases when mixed models were used (Model 5, Model 6, Table 1) the marginal and conditional pseudoR2 were close to each other (**Table 1**) indicating a weak role of the random factor (*~~subset~~* *Set*) as regulator of models, i.e. a satisfactory reproducibility of the results from population to population in all the regions.

The proportion of *M. trossulus* in samples (*Ptros*) was positively correlated with the proportion of T-morphotypes (*PT*) in the other sets, as it did in the samples from the White and the Barents Sea. This tendency was significant for all the sets (**Fig. 2 это должна быть ссылка на Fig 3!**; Model 4, **Table 1**). Otherwise, the patterns of variation were different for different sets. For *GOM*, the regression line stretched above the Y=X line but close to it, indicating the proportionality between *PT* and *Ptros*. For *Balt*, the regression slope was very steep, and the regression line rapidly diverged from the Y=X line. This was due to the fact that the *PT* range in *Balt* was, unlike the situation in the other sets, very narrow (0-0.4) as compared with the *Ptros* range (~0-1), and the small surplus of T-morphotypes in the samples was accompanied by a strong increase in the *M. trossulus* prevalence. A similar tendency was observed in the scanty material from *Norw*. Both *SCOT* samples fell on the Y=X line. Noteworthy are single “outlier” samples from *GOM* and *NORW*, in which *PT* was close to zero but *Ptros* washigh. While *P(T|edu)* estimates were low everywhere but in *BH*, *P(T|tros)* demonstrated a strong variation among sets and a noticeable variation within some sets (**Fig. 2**; Model????; **Table 1**). Similarly to *WSBL*, most *M. trossulus* had T-morphotypes in *GOM* and *Sсot* but not in *BALT* and *NORW*. For *Ptros*=0.5, expected differences in the morphotype frequencies between the species were about 44% for *GOM*, 6% for *Balt* and 24% for *Norw*. A significant positive dependence of the frequencies of T-morphotype on *Ptros* among conspecific genotypes, which was so prominent in the White and the Barents Sea, was recorded elsewhere only in *BALT* for *P(T|tros)* (Table 1).

The structure of the dependence of *P(tros|T)* and *P(edu|E)* on *Ptros* in GOM, BALT and NORW (Model 6. Fig. ++, Table +) was the same as in the samples from the Kola Peninsula (Model 3. **Fig. 2**, **Table 1**): *P(tros|T)* increased with the increasing *Ptros*, while *P(edu|E)* showed an opposite tendency. To simplify and formalize the comparison, we provide the predictions of Model 6 for equally mixed populations (*Ptros*=0.5) together with their 95% confidence intervals in Table 2, where actual proportions of *M. trossulus* among T-morphotypes (*P(T|tros)*) and *M. edulis* among E-morphotypes (*P(T|edu)*) in pooled samples from the respected sets are also provided.

**Table 2.** Proportions of *M. trossulus* among T-morphotypes (*P(tros|T)*) and proportions of *M. edulis* among E-morphotypes (*P(edu|E)*) in pooled samples (direct count) and in equally mixed samples (predictions by the regression Model 6) in different sample sets. Low and upper boundaries of 95% confidence intervals are provided for predicted values (in brackets).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | P(edu|E) | | P(tros|T) | |
| Set | Ptros=0.5 | In the data | Ptros=0.5 | In the data |
| WBL | 0.77 (0.73-0.81) | 0.86 | 0.85 (0.82-0.89) | 0.86 |
| BH | 0.70 (0.61-0.78) | 0.84 | 0.57 (0.51-0.63) | 0.48 |
| GOM | 0.66 (0.54-0.77) | 0.86 | 0.86 (0.68-0.95) | 0.80 |
| BALT | 0.51 (0.44-0.58) | 0.46 | 0.82 (0.58-0.94) | 0.93 |
| NORW | 0.64 (0.53-0.74) | 0.51 | 0.86 (0.68-0.95) | 0.93 |
| SCOT | - | 0.90 | - | 0.96 |

For equally mixed populations the predictive values of *P(edu|E)* in *BALT* did not differ statistically significantly from 0.5, which corresponds to an equal probability of correct and incorrect identification. *P(edu|E)* in *BH*, though quite low, was statistically significantly different from 0.5*.* At the same time, the probabilities of correct identification of *M. trossulus* by the T-morphotype in *GOM*, *Balt* and *Norw* were quite high (for the range of *Ptros*>0.5). In general, high predictive values for both species were revealed in WSBL.

Variation in morphotype frequencies between *M. edulis* and *M. trossulus* within and between contact zones revealed in the study is illustrated in Fig. 1, where the estimates of *P(T|edu)* and *P(T|tros)* in pooled samples from different sets are provided. *P(T|edu)* was 53% in the saline Barents Sea and less than 10% in all the other sets. In its turn, *P(T|tros)* was 17% in Western Baltic, 42% in Western Norway, 49% in the Gulf of Maine and more than 75% in the White and Barents Seas and Northern Scotland. *P(T|tros)* estimates in Norway and the Gulf of Maine were much affected by the outlier samples (see above). If we discard these samples, *P(T|tros)* will make up 54% in Norway and 71% in the Gulf of Maine. Att! Надо везде договоритья о том, в чем измеряем доли. Я за доли от единицы

Fig. 1 also shows the morphotype frequencies in putatively pure populations of species out of the contact zones studied. Within the ancestral range of *M. trossulus* in the Pacific, the populations were nearly monomorphic for the T-morphotype. In the Passamaquoddy Bay *P(T|tros)* was 0.81, i.e. close to that in most of the conspecific populations in the Gulf of Maine. All reference *M. edulis* populations from temperate areas (Long Island Sound and Cape Cod in Western Atlantic, Northern and Norwegian Seas in Europe) were nearly monomorphic for the E-morphotype. At the northeast extreme of the species range in East Atlantic, in the Southwestern Barents Sea, *P(T|edu)* varied considerably between samples, in particular between samples from brackish (range 0-3%) and saline (0.35-0.70%) localities (see ESM Table 2), as it did along the Barents sea coast of the Kola Peninsula. Increased *P(T|edu)* was also recorded in two northernmost samples from Western Atlantic (both from saline localities), Greenland (66%) and the Gulf of St. Lawrence (73%).

**Associations between morphotypes and shell size**

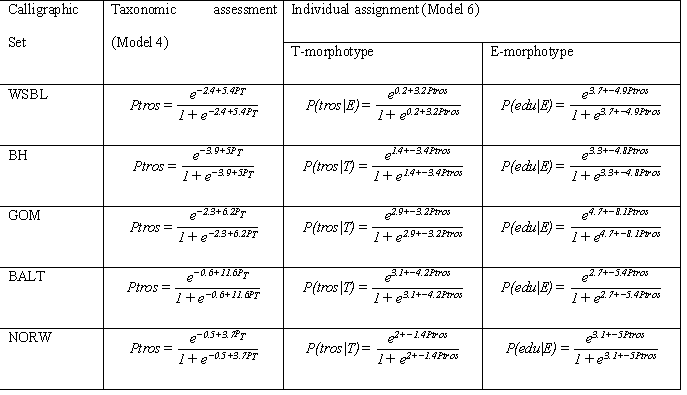
There was no clear statistical relationship between the size and the morphotype of conspecific mussels. At the level of individual samples, the probability of finding a T-morphotype increased with the mussel size (a positive slope-term of the regression) in 16 out of 34 informative comparisons (when species-specific genotypes were both present and polymorphic for morphotypes) for *M. edulis* and in 17 out of 43 comparisons for *M. trossulus*. The slope-terms of the regression models were individually significant (p<0.05) in four cases for *M. edulus* and in four cases for *M. trossulus*, but only in one case when the correction for multiple testing was applied (sample Bergen\_MV, see S3 table). We also checked the presence of any patterns in residuals from Model 6 as a function of mussel size. but no one was found.

**Prediction of taxonomic structure of populations and predictive values of the morphotype test basing on probability calculators**

Using the coefficients of the regression models Model 4 and Model 6 (Table 1), we constructed a set of formulas predicting the taxonomic structure (Ptros) and the probability of correct species identification using the morphotype test (Table 3).

Table 3. **Formulas used for taxonomic and individual assignment using morphotype tests in different sample sets (accordingly to the regression model’s coefficients represented in Table 1)**

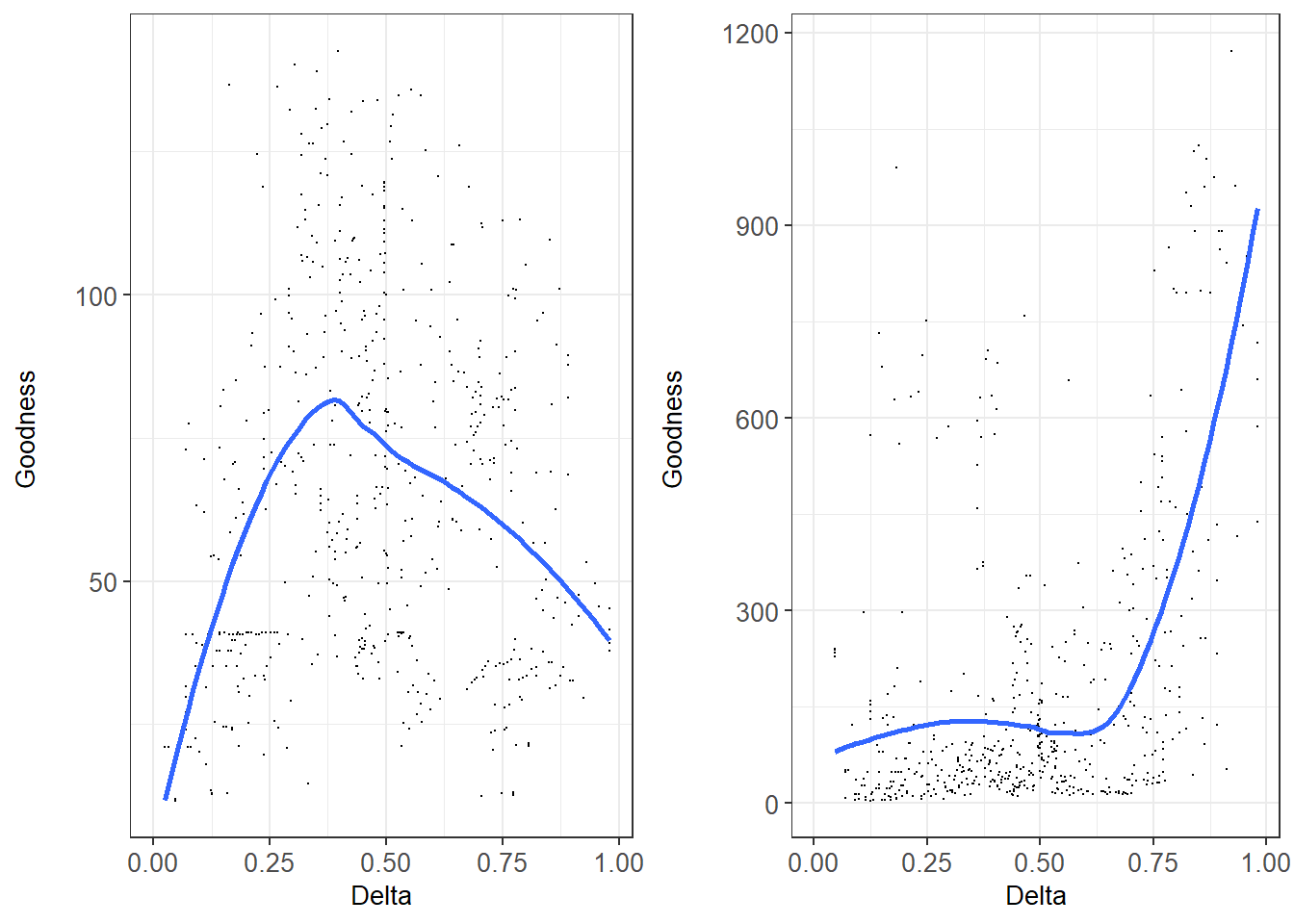
Это не последний вариант таблицы. Здесь плохо выводятся формуля с отрцательными коэффициентами. Надо вставить таблицу из Figures\_final\_version.html



Using this formulas one can predict Ptros, P(tros|T) and P(edu|E) for any new sample taken from the corresponding area if proportion of T-morphotype (PT) in the sample is known. However Eq. 1, 2 (“genotype by morphotype calculator”) and Eq 3(“Ptros by PT calculator”) also could be used for predictions of values Ptros, P(tros|T) and P(edu|E) on the base of PT if information on the P(T|tros) and P(T|edu) is available for any area (one can use the applett presented at +++++ to make the calculations accordingly to these equations). The key components of these equations are the values of P(T|tros) and P(T|edu) which could be obtained only after samples were genotyped and mussel’s mrphotype assessed. We will denote such samples as calibrating ones. To work out the strategy of searching of such calibrating samples we compare the predictions of Eq 3 with predictions of the regression Model 4 (Table 3) and Eq 1, 2 with the regression Model 6 (Table 3) using as an input the data from WSBL. For all possible pairs of samples from *WSBL* we calculated the values of P(T|tros) and P(T|edu) for the pooled data to adjust the Eq 1, 2 and 3. Then we compared the values predicted by these equations with those calculated by regression models. Fig. 4 illustrates the goodness of correspondence of the two predictions depending on the genetic constitution of the paired samples as expressed by the Delta index.

The comparison of predictions from Eq 3 and Model 4 revealed the best correspondence of both when Delta was close to one (Fig. 4 a). It means that the best results of “Ptros by PT calculator” is obtained if samples with maximally differ taxonomic structure would be used as calibrating ones to assess P(T|tros) and P(T|edu)*.*

Another result was obtained in comparison of predictions from Eq 1, 2 and regression Model 6. The highest correspondence between these predictions was observed for Delta distributed between 0.25 and 0.5. Thus the best results of “genotype by morphotype calculator” , i.e. assessment of P(edu|E) and P(tros|T) values, were obtained when P(T|tros) and P(T|edu) were assessed in the most mixed calibrating samples (*Ptros* of both samples close to 0.5).



Надо поменять местами эти два рисунка.

Fig. +. Correspondence between “genotype by morphotype calculator” (Eq. 1-2 , left graph) and “*Ptros* by *PT* calculator” (Eq. 3, right graph) and regression models (Model 6 and Model 4, respectively). Each point corresponds to a unique pair combination of samples from WSBL. . OX axis reflects dissimilarity of genetic structure in each pair (Delta) (for pure conspecific samples Delta takes a value of zero, for equally mixed samples – 0.5, for two pure heterospecific samples - 1). OY: goodness of correspondence between assessment of predictive values by equations and regression models.

We applied the “Ptros by PT calculator” (Eq 3) to all five geographical sets using, where possible, a set of the most dissimilar samples as calibrating data to assess P(T|tros) and P(T|edu) for each region (Fig. 3). Visual inspection of Fig. 3a revealed nearly ideal correspondence in the case of *WSBS* and *GOM*. In both these cases logistic regression curves were in closeness with straight lines obtained from the Eq 3. To note in the case of SCOT where only two samples were available the line derived from Eq3 approached closely to Y=X line.

Надо в таблицу, характеризующую выборки, которая в электронном приложении, вставить две колонки «Used as calibrating sample for Eq3” и “Used as calibrating sample for Eq1 and 2”.

НО! Где-то, я пока не понял где, надо написать, что в качестве калибровочных выборок для построения линий калькуляторов были использованы не по две выборки взятых формально, а по нескольку выборок, которые удовлетовряли следующим критериям

В категорию наиболее различных относили те выборки, которые имели Ptros<0.1 and Ptros>0.8

Соответственно, для каждого географичского сета в число калибровочных попали следующие выборки:

Для WSBL: belok2", "berzakol", "luv\_korg", "nm", "padan", "salnij", "umba\_06", "umba\_bridge","umba\_kamni", "umba\_pioner", "vor1", "vor2", "vor5", "kanal", "oenij", "zmis"

BH: "banka", "dz\_banka","tu\_old"

GOM: "BI", "CBE", "JPC", "KIM", "MDICOA", "PH", "VH"

BALT: "kast05", "Solvesborg", "Ystad05"

SCOT: Две наличные выборки

NORW: "Esp04", "Bergen\_MV"

В категорию наиболее смешаных относили те выборки, у которых 0.45 < Ptros < 0.65.

Соответственно это были следующие выборки:

WSBL: "abram", "niva\_sl", "sevsk", "umba", "umba\_pikut"

BH: "kuvsh", "seredina", "seredina\_sub", "ustie"

GOM: "CBSL"

SCOT: Две наличные выборки

NORW: Поскольку в этот диапазон не попала ни одна выборка, то калибровка производилась по всем имеющимся выборкам, которые были объединены.

Not ideal but rather close correspondence was observed in the case of BALT. Two notes should be made for this case. Firstly the range of PT values in BALT was narrow (0 - ++) but Ptros vary in broader limits (++ - 1). That’s why the slope for the Eq3 was rather high i.e. predicted Ptros will vary significantly after negligible changes in PT. It means that the “Ptros by PT calculator” allow making only rough assessment of Ptros by PT in BALT. Secondly predictions made by Eq3 underestimate Ptros value practically for all range of PT except its highest values (for the latter Eq3 predicts unrealistic values of Ptros).

The worst correspondence between Eq3 and Model 4 was observed in the case of NORW and BH. In the former situation was similar to one observed in BALT, i.e. Eq 3 had rather high slope and severely underestimated Ptros. In the case of BH we observed opposite pattern: the straight line derived from Eq3 was upper than logistic regression line, i.e. “Ptros by PT calculator” overestimated Ptros values in the case of BH.

Visual inspection of Fig. 3c revealed that curves derived from Eq 1 and Eq 2 (“genotype by morphotype calculator”) being calibrated by a set of samples with most mixed taxonomic structure were in rather good correspondence with regression lines. In most cases the calculator’s line situated inside 95% CI calculated for regression lines. However in some cases the high divergence between curves (calculator’s line was out of CI bound) was observed: P(tros|T) predicted by calculator downward deflected from regression curve in the case of WSBL and calcultor’s P(edu|E) deviated upwardly from regression curve in the case of GOM. However in both these cases the deviations were observed in the Ptros range corresponded to minor species: low Ptros for P(tros|T) and high Ptros for P(edu|E).