**Taxonomic assignment based on a semi-diagnostic marker: formal evaluation of a simple conchological test for hybridizing mussels Mytilus edulis L. and M. trossulus Gould**

**Abstract**/300 слов plos one / 350 слов peerj**.** Cryptic species as well as hybridizing species could lack diagnostic taxonomic characters leaving researchers with semi-diagnostic markers. Taxa identification by such markers is probabilistic and the probability of correct identification is dependent on the taxonomic structure of populations. Blue mussels Mytilus edulis and M. trossulus are ecologically, economically and stratigraphically important species that co-occur and hybridize along coasts of North Atlantic. Mytilus edulis and M. trossulus are generally treated as cryptic species and their studies will benefit from any cues to distinguish them in sympatry without genotyping. Recently it was shown that in the brackish White Sea species statistically differs by a simple conchological trait (morphotype) of the presence/absence of nacre under the ligament nympha. Here we analyze congruence between morphotypes and species multilocus genotypes in different contact zones between species and formally evaluate the performance of the “morphotype test” to discriminate species using methodology from clinical diagnostics. Unidirectional differences between genotypes in morphotype frequencies were ubiquitous but their scale varied among zones and also between saline and brackish localities in the Barents Sea. The latter was due to salinity-related variation within M. edulis recorded in the subarctic only. In the brackish Barents Sea and the White Sea the accuracy of individual assignment approached 75?% in the 1:1 mixtures of species. Quite strong differences were revealed in limited material from Northern Scotland and the Gulf of Maine indicating that there are prospects for use of the test for individual assignment in these regions too. In the saline Barents Sea, the Baltic Sea and Western Norway the test may be recommended for population assignment only. We provide empirical functions relating taxonomic structure of samples and the accuracy of individual assignment with morphotype frequencies in the well-described contact zones and recommendations for application of a test for the understudied zones.

**Introduction.**

Widespread cold-temperate blue mussels Mytilus edulis and M. trossulus are old evolutionary lineages of Pliocene origin (Riginos, Cunningham 2005) that co-occur and hybridize at least in six geographical areas of the North Atlantic and adjacent Arctic: Western Greenland (Wenne et al. 2016), American coast from Gulf of Maine to Hudson Bay (Seed 1992; McDonald et al. 1991), Northeastern Scotland (Beaumont et al. 2008; Dias et al. 2011), Western Baltic Sea (Väinölä, Hvilsom 1991, Stuckas et al. 2017), Western Norway and coasts of Kola Peninsula in Russia (White Sea, Barents Sea) (Vainola, Strelkov 2011). It is accepted that a more common M. edulis is a native species in the Atlantic while multiple natural and anthropogenic invasions stands behind sparse distribution of basically Pacific M. trossulus (Rawson, Harper 2009; Vainola, Strelkov 2011; Wenne et al. 2016). Two species were first recognized by molecular genetic markers (Varvio et al 1988) and a number of morphometric studies attempted to find reliable characters for their identification. Studies employing many metric shell traits and multidimensional approach indeed confirmed discreteness of species, but with no individually informative characters (McDonald et al. 1991; Mallet, Carver 1995; Innes, Bates 1999). Therefore M. edulis and M. trossulus are generally treated as cryptic species and are routinely identified genetically. While multilocus analysis is desirable for unambitious identification of species and their hybrids, quite often singular presumably diagnostic markers are employed, most often the protein coding region for the polyphenolic adhesive protein (ME 15/16 or Glu-5’) (Larrian et al. 2019).

Blue mussels are ecologically, economically and stratigraphically important molluscs (любимый обзор на экологию; FAO 2020; Mangerud, Svendsen 2018). It was proved or suggested that apart from biogeographic histories M. edulis and M. trossulus also differ in life traits, ecological requirements, properties as biomonitoring and aquaculture objects etc. (Katolikova et al. 2016; Michalek et al. 2016; Beyer et al. 2017 and references therein). The most illustrative example is harm associated with M. trossulus invasion on long line aquaculture designed for M. edulis. Cryptic invasion of M. trossulus on M. edulis plantations in Loch Etive (Scotland) in 2000s have resulted in significant losses in production because under conditions of Loch Etive M. trossulus has low consumer properties and also shells too fragile for the harvest and grading processes (Beaumont et al. 2008; Dias et al. 2011). Considerable differences between species were also found in Canadian aquaculture (Mallet & Carver, 1995; Penny et a., 2002) where the commercial value of M. trossulus was estimated to be 1.7 times lesser than that of M. edulis (Mallet & Carver, 1995). Inability to identify species by shells complicates mussel studies which will therefore benefit from any cues to distinguish M. edulis and M. trossulus in sympatry without genotyping.

Recently it was shown that in the White Sea M. edulis and M. trossulus differs by a simple conchological trait (morphotype) of the presence/absence of an uninterrupted prismatic strip under the ligament on the inner side of the shell: 74% of M. trossulus-like mussels (i.e. multilocus genotypes dominated by M. trossulus genes, group mostly includes purebreds but also 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 the marker for individual and population assignment correctly and with greater efficiency. Unlike fixed - diagnostic traits routinely used for identification of species, i.e. when for each of the compared species there is a unique state of a trait for all individuals, the trait under consideration is not diagnostic but semi-diagnostic i.e. polymorphic within species but with states distributed in different frequencies across species (definitions are from Padial et. al. 2010). Basing on the quite strong 70% differences in morphotype frequencies between species in the White Sea one can fall into the trap deciding that any randomly taken White Sea mussel of T-morphotype could be assigned to M. trossulus with high probability while any mussel of E-morphotype - to M. edulis. But in fact the probabilities of correct identification depend on the proportion of M. trossulus and M. edulis in the population under study. Mussel of any morphotype sampled from “pure” M. trossulus population (expected T-morphotype frequency 74%) would be M. trossulus anyway. Similarly any mussel sampled from “pure” M. edulis population (4%) would be M. edulis. At the same time in the 1:1 mixture of species (expected frequency =(74+4)/2=39%) 95% of mussels of T-morphotypes would be M. trossulus (=0.74\*0.5/(0.39) = 0.949) while 79% of mussels of E-morphotypes - M. edulis (= 0.96\*0.5/(1-0.39) = 0.787). These calculations certainly would be precise if only morphotype frequencies within species-specific genotypes do not vary with taxonomic structure of populations. Unlike practical taxonomists, clinicians often have to deal with semi-diagnostic characters since many clinical diagnostic tests are semi-diagnostic or considered as such. The formal procedure have 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), and one can hope that this methodology could be used for evaluation of semi-diagnostic taxonomic test for cryptic species relative to the species-specific genotype as well. To emphasize the analogy with clinical approach we denote here the procedure of mussel species identification by morhotype as “morphotype test”.

The second question is whether basic morphological differences between M. trossulus and M. edulis revealed in the White Sea is a locally restricted phenomenon or two species could be distinguished by morphotypes in other populations and contact zones as well. In the latter case the morphotype test would assist in local mussel studies over the Atlantics. Since morphological differences were overlooked in morphometric studies that all were based on references from other populations, one can think that differences appear in the White Sea only. This could be due to unusual environmental features of the Sea and (or) a history of local M. trossulus and a contact zone between species. One unusual feature of the White Sea is a combination of subarctic climate and relatively low salinity below 25 ppt (Derjugin 1928). It was supposed that M. trossulus invaded Kola Peninsula through marine traffic very recently, in the midst of the 20s century while most other Atlantic M. trossulus populations are probably much older (Vainola, Strelkov 2011).

In the paper we try to answer the above mentioned questions. First we analyze associations among morphotypes and species-specific genotypes in rich material from waters of the Kola Peninsula and in limited collections from Norway, the Baltic Sea, Scotland and the Gulf of Maine. For Kola mussels, we compare populations from the marginal White Sea and from the oceanic Barents Sea coasts on the one hand, and from brackish and saline localities in the Barents Sea on the other, in order to see how local geography and salinity (or tightly related factors) affects morphotype frequencies on the background of similar climate and biogeographic history of populations. Secondly we formally evaluate the performance of “morphotype test” for species identification using approaches from evidence based medicine and provide practical recommendations for its use for individual and population assignment.

**Materials and Methods.**

**Samples.** Altogether, 75 mussel samples (total sample size N=4087, individual sample size N=18-119) representing five geographical contact zones between M. edulis and M. trossulus were considered: the Gulf of Maine in the NW Atlantics (12 populations, N = 428 mussels), Loch Etive in Northern Scotland (2, N = 160), Western Baltic Sea (8, N =586), Bergen city area in Western Norway (5, N =165) and coasts of Kola Peninsula in Northern Russia: 24 populations from the White Sea (N=1100) and 26 populations from the Barents Sea (N=1650). Available information about samples and sampling localities is provided in the ESM table 1. The Barents Sea samples were from Kola Bay and from the open oceanic coast of Eastern Murman and were classified by salinity in sampling localities into the brackish- (salinity 5-30 ppt) and the saline water (>30 ppt) ones. The first group included nine samples from the fresh top of Kola Bay and three from the open coast. The second – eight samples from the mouth of the Bay and six from the open coast (fig. 1). Summer salinity in sampling localities was either taken from literature (Derjugin 1915; как правильно цитировать КОЛЬСКИЙ ЗАЛИВ, Bobkov et al. 2010; Shavykin 2018– ЧТО ЕЩЕ) or, in case of few open coast localities, predicted based on the presence or absence of large rivers nearby. As far as the material from other contact zones was limited, no targeted attempts to study variation in relation to local geography and salinity in America, Scotland, Baltic and Norway have been undertaken. All American samples and two of five Norwegian samples were from saline habitats, all the other from brackish habitats.

We also identified morphotypes in ?? samples (total sample size N=..., individual sample size N=...-...) of putatively pure M. edulis and M. trossulus sampled at a distance from contact zones considered (M. edulis and M. trossulus 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, from saline and brackish water localities in Europe and in the SW Barents Sea, ESM table 1).

Fig 1. Map of study area. A. Sampled regions in America: Gulf of Maine (Gom). B. Sampled regions in Europe: Scotland (Scot), the Baltic Sea (Balt), Norway (Norw), Kola Peninsula (Rus). C. Kola coast of the Barents Sea. Location of Kola Bay is indicated. D. Kola Bay. Dark and light pins depict high- and low saline sampling sites in the Barents Sea respectively. Pairs of pie diagrams depict proportions of T-morphotypes (black sector) and E-morphotypes (white sector) among M. trossulus (left diagrams) and M. edulis (right diagrams) in combined samples from particular regions. For Kola Bay and for the open Kola coast data on low-saline and high-saline areas are presented separately, on the upper and bottom diagrams respectively. Detailed data are in the ESM table 1. Придется перекраивать карту, чтобы добавить Америку; можно отказаться от Белого моря внутри Kola Peninsular, тем более что это разные вещи, и от отдельной Тювы. Референсы можно добавить мелкими кружочками, лучше пулированные данные (пресная/соленая Печора, пресное/соленое Северное море, вся восточная и вся западная Пацифика, т.пр.) Кто-то еще скажет что дб координаты на картах и мелкомасштабная карта Северной Атлантики и сопредельной Арктики. МАРИНА, СДЕЛАЙ КАРТУ.

**Genetic characters.** Some samples from contact zones represented available collections of genotyped mussels from 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), others were new (see fig. 1 and ESM table 1). 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 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 (Vainola, Strelkov 2011; Katolikova et al. 2016; Strelkov et al. 2017). New samples from the Gulf of Maine were genotyped as in Kingston et al. 2017, other samples as in Katolikova et al. 2016. For seven samples from the Barents Sea data on only three loci - Est-D, Gpi and Pgm were available however (see ESM1 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). Russian material was also analyzed by three loci (Odh not considered) to show that exclusion of Odh does 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 named as “M. trossulus” and “M. edulis” genotypes hereinafter in spite of the fact that each includes both purebreds and hybrids. Genetic information on samples but results of classification into M. trossulus and M. edulis will not be considered here. The detailed analyses of hybrid zones under consideration, in particular proportions of purebreds and hybrids in mixed samples have been provided in literature (Zbavicka et al. 2010; Vainola, Strelkov 2011; Katolikova et al. 2016; Kingston et al 2017; Strelkov et al. 2017 and references therein).

**Morphologic characters.** Data on the White Sea samples were taken from Katolikova et al. 2016 and 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 valves under a 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. To note in previous papers (Katolikova et al 2016; Khaitov et al 2018) 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 geographical variation in the coloration and width of the “strip”, that’s why we specify the diagnosis in the Results and provide more photos of morphotypes in the ESM.

**Predictive values.** For each sample we calculated frequencies of *M. trossulus* (*Ptros*) and T-morphotypes (*PT*). We further calculated four indexes that measure 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 also 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. 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 a mussel of E-morphotype is *M. edulis*, respectively. It is worth mentioning that these indexes are used in clinical medicine for evaluation of the performance of diagnostic tests (КТО ОПРЕДЕЛИТ ЛУЧШИЕ ССЫЛКИ?). If we accept, conditionally, that *M. edulis* is a “healthy mussel” while *M. trossulus* is a “seek mussel” (a reasonable assumption bearing in mind a threat of *M. trossulus* to Scottish aquaculture, Beaumont et. al. 2008), *Ptros* would be called *prevalence* while *P(T|tros)* – *sensitivity*, *P(E|edu)* - *specificity*, *P(tros|T)* - *positive predictive value* and *P(edu|E)* - *negative predictive value* of the morphotype test, in medical jargon (REF).

It is axiomatic that predictive values naturally vary with prevalence: with the increasing *Ptros*, *P(tros|T)* will gradually increase from 0% in pure populations of *M. edulis* to 100% in pure populations of *M. trossulus* while *P(edu|E)* will demonstrate an opposite relationship. For the test to be meaningful both predictive values should be >0.5 in well mixed populations (Ptros~0.5) since the predictive value of 0.5 means random association between genotype and morphotype. Under the assumption that sensitivity and specificity are independent on the prevalence, predictive values could be directly predicted basing on the *Ptros* in a sample, and 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 basing 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.** We undertake the next analyses using data from contact zones. Firstly, we studied variation of *PT*, *P(T|tros)*, *P(E|edu)*, *P(tros|T*), *P(edu|E)* as functions of *Ptros* within and among sample sets representing A) the White Sea and the Barents sea coasts of Kola Peninsula and saline and brackish water localities in the Barents Sea (Section “Associations among morphotypes and species-specific genotypes around Kola Peninsula”), B) different geographical contact zones between species. Whenever possible formulas describing empirical relationships between *PT* and *Ptros* and between *Pros* and predictive values of a test have been derived (Section “Associations among morphotypes and species-specific genotypes around Atlantics”). Secondly, we analyzed genotype-specific associations between morphotypes and shell size in order to verify the hypothesis that morphological variation under consideration is not related to mussel age or size (Section “Associations between morphotypes and shell size”). Finally, we verified how well the *Ptros,* *P(edu|E)* and *P(tros|T)* in samples could be predicted using formulas Eq 1-3 and data on morphotype proportions among species (P(T|tros), P(T|edu)) in two genotyped “calibration” samples. While allowing that the assumption of the independence of sensitivity and specificity on the prevalence could be violated in morphotype test, as it is often violated in clinical tests (REF) we were especially interesting what samples can better aid in the prediction: most mixed ones (*Ptros*~0.5) or combination of two most pure samples of each species (Section “Prediction of taxonomic structure of populations and predictive values of the morphotype-test basing on calibrating samples”).

All statistical 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. 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 the means of visual analysis of residual plots and assessment of overdispersion presence.

For each analysis we first constructed the full models (included all predictors and their interactions) and then simplified them 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” (REF) 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). 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 among morphotypes and species-specific genotypes around Kola Peninsula***.* For the purposes of statistical analyses all samples from the Barents Sea open coast (three samples representing brackish and six samples representing saline localities) and four samples with from the White Sea with the *Ptrost* nearest to 0.2, 0.4, 0.6 and 0.8 chosen to maximally represent the prevalence variation were excluded from statistical processing. These samples were used as a testing data set to assess the validity of model predictions. Other samples - all from brackish areas of the Kola Bay (hereafter, sample set *BL*), from saline areas of the Kola Bay (*BH*), and all but four from the White Sea (*WS*) were used for modeling (model parameters were estimated on the base of this data). The next tree next regression models were fitted for the data.

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 *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 analogously to Model 1 and was modelled as a function of *Ptros*, *Set*, *Species* (discrete predictor with two levels) and interaction between terms. Sample was included into the model as a random factor influencing the model intercept.

Model 4: 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 (*M. trossulus* is marked by the T-morphotype or *M. edulis* is marked by the E-morphotype) and as 0 (other cases). The set of predictors for the model was as follows: *Ptros*, *Morphotype* (discrete predictor with two levels), *Set* and interaction between terms. Sample was included into model as random factor influencing the model intercept.

To check whether it is possible to pool some of subsets to construct more general models 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 full combination of sets since in this case the factor “*Set*” would be discarded from the model. We applied the structure of the Model 4 to these new recombined datasets. Then we compared AIC’s of these new models with AIC of Model 4 based on non-pooled data. If AIC of a new model was less than AIC of the initial one we considered this as a basis for pooling of corresponding sample sets.

**Associations among morphotypes and species-specific genotypes around Atlantics**. 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, subsets WS and BL were pooled since they did not differ statistically, see Results). Scotland (SCOT) was not included because it was represented by two samples only. Three models were constructed:

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 function of PT (continuous predictor), *Set* and interaction between them. We modeled here *Ptros* vs. *PT* but not vice versa as in previous analysis in order to use the model as reference for the “*Ptros* by *PT* calculator” (see below).

Model 6. Morphotype proportions among species (*P(T|tros)*, *P(T|edu)*) as a function of taxonomic structure of populations (*Ptros*). Model was constructed analogously to the Model 2.

Model 7. Correctness of species identification (*P(tros|T*) and *P(edu|E)*) as a function of taxonomic structure of populations (*Ptros*). Model was constructed analogously to the Model 4.

**Associations between morphotypes and shell size.** For technical reasons we avoid mussel size as a covariate in predictive models considered. To check the possible association of morphotypes with size we undertaken the next two analyses. Firstly we constructed logistic regression models for each available species-specific genotype from each population. The probability of T-morphotype presence 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) were considered. Secondly we checked the presence of any patterns in residuals from the Model 6 (i.e. the main model designed to predict the probability of correct identification of 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 basing on the results of the following analysis.

We considered the WSBL (pooled modeling data from the White Sea and the brackish Barents Sea, 29 36??? samples) as reference dataset. All ~~406~~ 630??? possible pair combinations of samples were considered. Each pair was characterized by an index of taxonomic similarity between samples included:

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 obtained by the Eq. 1, 2 were contrasted with such predicted by the Model 7 for *Ptros* and the Model 4 for *P(edu|E)* and *P(tros|T)* using goodness of fit statistics:

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

Goodness indices for each pair we plotted against corresponding Delta values and fitted LOESS regression curve to find associations between them. Depending on the results of the analyses we decided what samples can better aid in the predictions and be used as “calibrating” ones: most mixed samples (*Ptros1* ≈ *Ptros2* ≈ 0.5) or 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 program …

**Results**

**Geographical variation in the manifestation of mussel morphotypes.** The binary morphological character that we studied was previously defined as the “presence/absence of a distinct uninterrupted dark prismatic strip under the ligament” (Katolikova et al. 2016; Khaitov et al., 2018). Only material from the White Sea was considered in previous studies. While E-morphotypes in all new populations studied looked the same (the srtip absent: the nacreous layer totally or partially covers the space under the ligament nympha, ESM Fig. +), analysis of geographical data revealed some variation among T-morphotypes unseen previously in the White Sea. In rare shells from most geographical populations studied the nacreous-free strip of prismatic layer was quite narrow and looked as a stria rather than a strip (ESM Fig. +). Further, in all the Gulf of Maine T-morphotypes the same as in rare mussels from other populations the color of the strip was pale (as a shell prismatic layer per se) rather than dark. It made T-morphotypes hardly noticeable to the unaided eye. To reveal the T-morphotype unambiguously it is necessary to find using dissecting microscope the morphologically pronounced scar that defines the boundary of the nacreous layer under ligament nympha (ESM Fig. ++). The most honest diagnoses of T- and E- morphotypes would be the presence/absence of the uninterrupted strip of the prismatic layer under the ligament nympha recognizable by a clear scar separating the strip from the nacreous layer of the rest of shell.

ESM Fig. ++. Mussel morphotype variation. A. E-morphotypes: space under the ligament nympha is totally (left) or partly covered by the nacre (right). B. T-morphotypes: a strip of uncovered prismatic layer under the ligament nympha is dark and wide (typical case for all populations but the Gulf of Maine ones, left) or narrow and recognizable by a scar separating the nacreous layer from the strip of the uncovered prismatic layer only (typical case for American mussels, right).

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

Patterns of *PT*, *P(T|tros)*, *P(E|edu)*, *P(tros|T)*, *P(edu|E)* variation as functions of *Ptros* among samples representing the White Sea (*WS*), the brackish- (*BL*) and the saline (*BH*) Barents Sea ~~(only modeling data considered)~~ are visualized on Fig. 2 whereas results of regression analysis are summarized in table 1.

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

Analysis of variation T-morphotype frequencies among subsamples of *M.edulis* (*P(T|tros)*) and *M.trossulus* (*P(T|edu)*) against proportions of *M. trossulus* in samples (*Ptros*) revealed the next patterns (Model 2, Table +, Fig. +). The tendency to higher frequency of T-morphotypes among *M. trossulus* than among *M. edulis* was universal, but in the *WS* and *BL* it was strong (for *Ptros*=0.5, expected differences in morphotype frequencies between species about 65%) and significant (the confidential intervals for regression lines for *P(T|tros)* and *P(T|edu)* do not overlap) while in the *BH* - quite small (differences 16% for *Ptros*=0.5) due to increased *P(T|edu),* and insignificant (the confidential intervals overlap, Fig. ++).

In all three subsets a positive correlation of *P(T|tros)* and *P(T|edu)* with *Ptros* was found, that is with increasing contribution of *M. trossulus* to samples frequencies of T-morphotypes increased both among *M. edulis* and *M. trossulus*. As a result, T-morphotype frequencies among both genotypes were usually few dozens of percent higher in M*. trossulus*-dominated samples than in *M. edulis*-dominated samples.

Probability of correct identification of M. trossulus by T-morphotype (frequency of M. trossulus among T-morphotypes, (P(tros|T)) expectedly increased with increasing of Ptros whereas probability of correct identification of M. edulis by E-morphotype (P(edu|E)) demonstrated an opposite pattern (Model 4, Table +, Fig. +). In the M. trossulus - dominated populations P(tros|T) tend to one (any mussel with T-morphotype is 100% M. trossulus) while P(edu|E)) tend to zero (any mussel with E-morphotype is 100% M. trossulus), and vice versa. In the well mixed populations, where the regression lines intercross, the predictive values for both species was about 0.75 in the WS and BL but only 0.6 in the BH (Fig. ?). It means that morphotype test has much less predictive value in the saline Barents Sea than in the brackish Barents Sea and in the White Sea, remembering that the predictive value of 0.5 means random association between genotype and morphotype. It is evident from the Fig. 2 that the low predictive value of the test in the BH is mainly due to generally low *P(tros|T)*: even though the great majority of *M. trossulus* bear T-morphotype it is difficult to recognize them because many *M. edulis* bear this morphotype too. On the other hand E-morphotypes let not that common in the BH samples are predominantly founded in *M. edulis*. Nevertheless the statistical analysis indicates that both *P(tros|T)* and *P(edu|E)* predicted by the model were lesser in the BH than in the WS and BL.

For each of the GLMM models considered (Model 2 and 4), marginal and conditional pseudoR2 were close to each other (Table ++) indicating the weak role of random factor (population) as regulator of models, i.e. the satisfactory reproducibility of results. In accordance with this, points from the testing data mostly scattered along the regression lines build for modeling data sets indicating the high predictive values of respective models (see Figure 3 and below).

In intra-set comparisons the regression coefficients did not differ statistically for the *WS* and *BL* sets while the *BH* each time was different from the *WS* (Table 1). To assess the possibility of data-sets pooling we compared the AIC of full Model 4 (AIC = 1624.3) with AICs of three other models based on differently pooled of *WS*, *BL* and *BH* sets. The model based on pooled *WS* and *BL* (*WSBL*) and *BH* showed the lowest AIC of 1620.3. Therefore in the next analyses we will consider the *WSBL* and the *BH* sets.

Figure 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, 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 (filled points) and M. edulis (empty points) (Model 2). (C) Frequencies of M. trossulus among T-morphotypes (filled points) and of M. edulis among E-morphotypes (Model 4). Vertical lines on B and C connect subsamples of M. trossulus and M. edulis from the same samples.

Table 1. Parameters of regression models fitted. Я слышал о правиле что иллюстрации д.б. понятны без чтения текста, и наоборот. Это образцовый пример такой иллюстрации.

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

Patterns of *Ptros* variation against *PT* and patterns of *P(T|tros)*, *P(E|edu)*, *P(tros|T)* and *P(edu|E)* variation against *Ptros* among samples from different geographical zones are visualized on Fig. 3 whereas results of regression analysis are summarized in table 1. Note that Scottish material was not included in the regression analyses. Re-analyses of data from the White and Barents Sea (*WSBL* and *BH* sets) together with data from other regions certainly did not change the patterns revealed above. This is additionally emphasized by the distribution of testing samples from the White and Barents Seas that generally scattered along the regression lines build for corresponding modeling data sets (P(T|tros) are P(E|edu) in the *BH* expectedly were an exclusions, Fig. 3).

As in the White and Barents Seas, in other sets the proportion of *M. trossulus* in samples (Ptros) was positively correlated with the proportion of T-morphotypes (*PT*) and this tendency was highly significant for all sets but *NORW* (Fig. 2; Model 4, Table 1). But otherwise, the patterns of variation were different for different sets. For *GOM* the regression line stretched above the Y=X line but in close proximity to it indicating proportionality between *PT* and *Ptros*. For *BALT* the regression slope was extremely large and the regression line rapidly diverged from the Y=X line. It is because unlike other sets in the Baltic one the *PT* range was very narrow (0-0.4) relative to the *Ptros* range (~0-1) and the small surplus of T-morphotypes in samples was accompanied by the strong increase in the *M. trossulus* prevalence. Similar tendency was observed in poor material from *NORW*. Both *SCOT* samples coincided with the Y=X line. Worthy of note are singular “outlier” samples in GOM and NORW with the *PT* close to zero but high *Ptros*.

While *P(T|edu)* estimates were low everywhere but in the *BH*, *P(T|tros)* demonstrated strong variation among sets and also noticeable variation within some sets (Fig. 2; …; Table 1). Like in the *WSBL*, the great majority of *M. trossulus* bore T-morphotypes in *GOM* and *SCOT* but not in the *BALT* and *NORW*. The expected *P(T|tros)* in equally mixed populations (Ptros=0.5) was about ??, ??, ?? and ?? for these four sets respectively. Significant positive dependence of T-morphotype frequencies on *Ptros* among conspecific genotypes, so evident in the White and Barents seas was recorded only for *P(T|tros)* in the *BALT* (Table 1).

In accordance with differences in *P(T|tros)* and to the lesser extent in *P(edu|E)*, distributions of predictive values were different for different sets. To simplify and formalize the comparison we provide the predictions of the Model 6 for equally mixed populations (*Ptros*=0.5) together with their 95% confidence intervals in the 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 respected sets are also provided.

Table +. Proportion of mussels correctly identified by morphotype test in different regions т.е. 95% троссюлюс в балтике он определил корректли

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 ?) 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 (0.72-0.81) | 0.62 | 0.86 (0.82-0.89) | 0.72 |
| BH | 0.67 (0.59-0.74) | 0.89 | 0.54 (0.49-0.59) | 0.10 |
| GOM | 0.67 (0.57-0.76) | 0.82 | 0.86 (0.70-0.95) | 0.62 |
| BALT | 0.51 (0.46-0.56 | 0.00 | 0.82 (0.58-0.94) | 0.93 |
| NORW | 0.63 (0.54-0.72) | 0.17 | 0.87 (0.70-0.95) | 0.93 |
| SCOT | NA | ?? | NA | ?? |

For equally mixed populations the predicted values of *P(edu|E)* in the *Balt* and *Norw* and the *P(tros|T)* in the *BH* did not differ statistically from 0.5 – a value that corresponds to equal probability of correct and incorrect identification. Further, the *P(edu|E)* was quite low yet statistically different from 0.5 in the *BH* and *GOM*. In the same time the probabilities of correct identification of *M. trossulus* by T-morphotype in the *GOM*, *Balt* and *Norw* were quite high indicating the possibility of identification of *M. trossulus* by T-morphotype in these regions (for the range of Ptros>0.5). The high predictive values for both species were revealed in the WSBL.

**The best-fit regressions for sample sets under consideration that could be used for taxonomic and individual assignment using morphotype tests were:**

P(edu|E) для балтикии норвегии и P(tros|T) для BH вряд ли можно применять для individual assignment? Значит и приводить не надо?

WSBS: Ptros=…. (PT); P(edu|E) = … ; P(tros|T) =

Адаптировать и как-то инкорпотрировать: For each of the three models considered, marginal and conditional pseudoR2 were close to each other (Table ++) indicating the weak role of random factor (population) as regulator of models, i.e. the satisfactory reproducibility of results. In intra-set comparisons the regression coefficients (Table 1).

Variation in morphotype frequencies between *M. edulis* and *M. trossulus* within and among contact zones revealed in the study is well illustrated by the Fig. 1 where 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 other sets. In its turn the *P(T|tros)* was 17% in the 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 abovementioned outlier samples. Excluding these samples, *P(T|tros)* would be 54% and 71% in two regions respectively. Fig. 1 also provides insight into morphotype frequencies in putatively pure populations of species out of contact zones studied. Within the ancestral range of *M. trossulus* in the Pacific, populations were nearly monomorphic for T-morphotype. In the Gulf of St. Lawrence *P(T|tros)* was 0.81, i.e. close to that in most the Gulf of Maine conspecific populations. All reference *M. edulis* populations from temperate areas (Long Island Sound and Cape Cod in Western Atlantics, Northern and Norwegian Seas in Europe) were nearly monomorphic for 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), just as along the Barents sea coast of Kola Peninsula. Increased *P(T|edu)* was also recorded in the northernmost samples from Western Atlantics (both two 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 size and morphotype of conspecific mussels. At the level of individual samples, the probability to find T-morphotype increased with mussel size (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 ESM1). **We also checked the presence of any patterns in residuals from the Model 6 as a function of mussel size. И ЧТО?**

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

We applied Eq.1 and Eq. 2 (predictive values as a function of *Ptros*, *P(T|tros)* and *P(E|edu)*, “genotype by morphotype calculator”) and Eq. 3 (“*Ptros* by *PT* calculator”) using as an input data on all possible pairs of populations from the *WSBL* and compared values predicted by these equations with that predicted by the regression models Model 6 and Model 4 respectively. Fig. 4 illustrates how well two predictions fit each other depending on the peculiarities of genetic constitution of paired samples as expressed by the index Delta. The best predictions of the *Ptros* were obtained when most dissimilar – nearly pure *M. edulis* and *M. trossulus* samples have been used while the best predictions of P(edu|E) and P(tros|T) values were obtained when most mixed samples (*Ptros* of both samples close to 0.5) have been taken as calibration ones. We applied “calculators” to all five geographical sets using, where possible, two most dissimilar samples for calculation of *Ptros* and two most mixed samples for calculation of predictive values (Fig. 3; note that for *Scot* only two samples were available). Visual inspection of the Fig. 3 shows good correspondence between distributions predicted by the “genotype by morphotype calculator” and regression lines in all cases but in *NORW*. The later was due to the formal choice of the only outlier sample with extremely low *P(tros|T)* as calibrating one. In its turn the “*Ptros* by *PT* calculator” was quite inaccurate for *BH*, *NORW* and *BALT* but nearly ideal for *WSBS* and *GOM*.

Мы в тупике? Ptros в половине случаев не предсказать, после чего уже не важно, хорошо ли предсказываются P(edu|E) и P(tros|T) по Ptros. Как это исправить?

Figure +. Correspondence between regression and theoretical models. Each point corresponds to one of the possible pairs of populations from modelling data set (White Sea joined with low salinity Barens Sea). OX axis represents the differencу in genetic structre for each pair of populations. OY axis represents correspondence between prediction of regression model and theoretical model. Lines represent LOESS-smoother. (A ) Model 7 describing the dependence of proportion of M.trossulus (Ptros) on proportion of T-morphotype (P\_T) ; (B) Model 6 describing the dependence of probability of correct species identification on proportion of M.trossulus and morphotype. Я не могу это компрехенд. «Correspondence between regression and theoretical models.» какие моделс, какие предикшнс? White Sea joined with low salinity Barens Sea - где они встречаются, разве что в Воронке Белого моря? Сплошное «differencу in genetic structre for»

Correspondence between predictions of Eq. 1-2 (“genotype by morphotype calculator”, left graph) and Eq. 3 (“*Ptros* by *PT* calculator”, right graph) derived basing on all possible paired combinations of “calibrating” samples from the WSBS set, and corresponding regression models. OX: peculiarities of genetic structure of calibrating samples as expressed by the index Delta (for pure conspecific samples Delta takes a value of zero, for equally mixed samples – 0.5, for pure heterospecific samples - 1). OY: goodness to fit statistics.

**Discussion**

**Основные результаты, полученные в работе**

**1. Два морфотипа (Е и Т) были найдены во всех исследованных местообитаниях, по обеим берегам Атлантики. Это не вопрос и не новость (золотарев)**

**2. Т-морфотип характеризуется не цветом полоски под лигаментом, а тем, что в этой области призматический слой не откладывается. У Т-морфотипа существует резкая граница между перламутром и призматическим слоем в этой области. Ширина полоски незакрытого призматического слоя может варьировать. Можно сослаться при формулировке рекомендаций**

**3. Частота M.trossulus в исследованных зонах контакта двух видов варьирует в широких пределах в очень небольших пространственных масштабах (этого мы, кстати, не изучали, но привести надо, хотя бы, среднее географическое расстояние между ближайшими соседями-популяциями в каждом из регионов, возможно, только для W, BL и BH, а может и для всех). Это не предмет настоящего исследования**

**4. Частота Т-морфотипа положительно коррелирует с частотой M.trossulus во всех изученных акваторриях. Аналогично частота M.trossulus положительно коррелирует с частотой Т-морфотипа. Это наблюдается во всех изученных акваториях. В случае с W, BL, GOM и, возможно, SCOT соотношение соотношение M.trossulus и T-морфотипа близко к 1:1. Пойдет для введения к обсуждению**

**5. Частота Т-морфотипа среди M.trossulus выше частоты Т-морфотипа среди M.edulis во всех акваторях. Однако в случае BH статистически значимые различия не выявлены. Как следствие п.6**

**6. В BH частота Т-морфотипа среди M.edulis выше, чем чем аналогичная величина во всех остальных акваториях. Уникальная особенность субарктических едулисов изменчивость по язычковости между солеными и пресными местообитаниями**

**7. В BALT и, видимо, в NORW частота Е-морфотипа среди M.trossulus выше, чем аналогичная величина в других акваториях.Противопоставление этих и трех других ГЗ**

**8. Чем выше частота M.trossulus в смешанных популяциях, тем выше частота Т-морфотипа, как среди M.edulis, так и среди M.trossulus. но не в GOM**

**9. Вероятность правильной идентификации случайно взятой мидии Т-морфотипа, как M.trossulus, и случайно взятой мидии E-морфотипа, как M.edulis, зависит от частоты M.trossulus в популяции. Вот это открытие так открытие**

**10. В случае W, BL, GOM морфотип-тест дает возможность надежной идентификации видов в смешанных поселениях (Ptros близко к 0.5). В поселениях с подавляющим доминированием M.trossulus (Ptros близко к 1) корректная идентификация минорного вида (M.edulis по Е-морфотипу) маловероятна. Аналогично, в поселениях с резким преобладанием M.edulis (Ptros близко к 0) корректная идентификация минорного вида (M.trossulus по T-морфотипу) маловероятна. В случае более или менее чистых поселений корректная идентификация доминирующего вида по морфотипу (M.edulis по Е-морфотипу и M.trossulus по T-морфотипу) возможна с высокой вероятностью. Это важнейший результат; individual assignment**

**11. В случае смешанных поселений в BH, корректная идентификация M.trossulus по Т-морфотипу маловероятна, но корректная идентификация M.edulis по Е-морфотипу возможна с высокой вероятностью.**

**12. В случае смешанных поселений в BALT, корректная идентификация M.edulis по E-морфотипу маловероятна, но корректная идентификация M.trossulus по T-морфотипу возможна с высокой вероятностью. А что норвегия? Там тоже язычковые троссюлюсы**

**13. Для применения уравнений теории вероятностей (Eq 1, 2) для экспресс-оценки вероятности правильного определения вида мидии по морфотипу требуются калибровочные выборки (дающие оценки частоты Е-морфотипа среди M.edulis и Т-морфотипа среди M.trossulus) с максимально смешанным составом (Ptros близко к 0.5).**

**14. Для применения уравнений теории вероятностей (Eq 3) для экспресс-оценки частоты M.trossulus по частоте Т-морфотипа требуются калибровочные выборки (дающие оценки частоты Е-морфотипа среди M.edulis и Т-морфотипа среди M.trossulus) с максимально различным составом (как минимум одна выборка с Ptros близким к 0 и одна выборка с Ptros близким к 1). Тут какой-то замкнутый круг. Ptros калькулятор нещадно лажает, что делает второй калькулятор вдвойне лажевым**

**15. Во всех акваториях теоретические кривые, вычисленные по уравнениям Eq 1 и 2, параметры которых были оценены в соответствии с выработанной стратегией, находятся в хорошем соответствии с регрессионными моделями, подобранными на основе первичных данных.**

**16. В случае с W, BL и GOM линии, описывающие связь между Ptros и частотой Т-морфотипа, построенные в соответствии с Eq3, находятся в хорошем соответствии с линиями регрессии, описывающими первичные данные. В случае с BALT соответствие хуже. В случае BH линия, подобранная, в соответствии с разработанной стратегией, существенно отклонялась от линии регрессии, построенной на основе первичных данных.**

**Как бы это трансформировать в обсуждение: если где-то калькулятор не сработал, значит он – хуйня? Что делать ученым?**

These data was supplemented by corresponding data on East Pacific populations … lacking in our collections from the study of Zolotarev (2002) who identified morphotypes in small samples of genotyped mussels from the study of McDonald et al. 1991.

Развернуто в чем главные результаты: как мы отвечаем на поставленные вопросы?

Представление обсуждения

После обзора косвенных негенетических маркеров сакраментальная фраза о том, что, насколько мы знаем это лучший морфологический признак для различения видов по раковинам из всех когда-либо использованных.

Mussels as bioindicator species. Beyer J, Green NW, Brooks S, Allan IJ, Ruus A, Gomes T, Bråte IL, Schøyen M. Blue mussels (Mytilus edulis spp.) as sentinel organisms in coastal pollution monitoring: a review. Marine environmental research. 2017 Sep 1;130:338-65.

Историческая динамика по музейным коллекциям: покуда технология ДНК идентификации широко не применяется,

Der Sarkissian, C., Pichereau, V., Dupont, C., Ilsøe, P.C., Perrigault, M., Butler, P., Chauvaud, L., Eiriksson, J., Scourse, J., Paillard, C. and Orlando, L., 2017. Ancient DNA analysis identifies marine mollusc shells as new metagenomic archives of the past. Molecular Ecology Resources, 17(5), pp.835-853.

Der Sarkissian, C., Möller, P., Hofman, C., Ilsøe, P., Rick, T., Schiøtte, T., Vinther Sørensen, M., Dalén, L. and Orlando, L., 2020. Unveiling the Ecological Applications of Ancient DNA from Mollusk Shells. Frontiers in Ecology and Evolution, 8, pp.1-21.

Like in the clinical practice there is often a patient spectrum from mild to hard illness that varies with prevalence affecting *sensitivity* (Leeflang et al. 2013) in 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.

Надо кому-нибудь прочитать Leeflang et al. 2013 и пр. и оценить стоит ли обсуждать другие источники варьирования

В заключение:

Situation when we have to rely on a singular informal – semi-diagnostic character to distinguish morphologically such old evolutionary lineages as M. edulis and M. trossulus is certainly uncommon in taxonomy. At the same time this situation is not unique in the sense that there are other taxa lacking fixed diagnostic characters and identified by individual or complex (like coordinates of multifactorial analysis – как это сказать?) semi-diagnostic morphological traits. These are subspecies when defined according to the 75% rule (“in order to qualify as a subspecies, 75% percent of one population must be separable, at a taxonomic character, from all of the members of the other population”, Amadon, 1949), cryptic species with statistical differentiation (*sensu* Chenuil et al. 2019) and hybridizing species that are secondary losing fixed differences due to continuous introgression (???). To note recent population genomic studies of hybridizing Mytlus species came to conclusion that species could lack fixed genetic differences/раскрыть due to ubiquitous introgression; on that basis the conventional approach of mussel species identification by singular molecular marker such as ME15/16 have been criticized (Simon ... 20). We therefore hope that our exercise how to deal with non-fixed taxonomic character will be interesting not only to blue mussel researchers but also to colleagues who study any sympatric taxa with vague morphologies and with semi isolated gene pools.