## Mussel morphotype variation

The morphological character that we used for morphotype identification was previously described as “distinct uninterrupted dark prismatic strip under the ligament” (Katolikova et al. 2016). However, more broad analysis (in previous studies Katolikova at al. 2016; Khaitov et al., 2018 only mussels from the White Sea were used) revealed some additional features which should be considered to make the description of morphotype more correct.

Mussel of E-morphotypes sampled from other geographic areas invariable carried the same traits as those from the White Sea: nacreous layer totally or partially covered the space under ligament nympha (Fig. +). However in the case of T-morphotype we revealed two new features. Firstly, in some cases (rare, approximately ++ % of shells) the nacreous free strip of prismatic layer could be very narrow (Fig. +). The thinning of the strip of the prismatic layer occurs as a result of enlargement of anterior adductor and its displacement to the dorsal surface. Secondly, the color of this strip may be rather pale in dependence on the color of prismatic layer. We have to make caution against the use of color for revealing of this strip. Sometimes the nacre completely covers the entire space under the ligament nympha (i.e. the mussel should be considered as E-morphotype), but it turns out to be quite thin and transparent. In this case the dark prismatic layer being visible through nacre can be mistakenly perceived as a sign of the T-morphotype. To reveal T-morphotype correctly it is necessary to find the morphologically pronounced scar that defines the boundary of the nacreous layer under ligament nympha (Fig. ++).

Thus we have to reformulate the description of this trait as follows. Mussels will be considered as a T-morphotype if they show a clear scar separating the nacreous layer from the strip of the prismatic layer not covered with nacre throughout the length of ligament nympha.

Fig. ++. Mussel morphotypes variations. A. Two variants of E-morphotype: with space under ligament nympha totally covered by nacreous layer (left) and some part of uncovered prismatic layer present in the posterior part of nympha (right). B. Two variants of T-morphotypes: typical case with broad strip of uncovered prismatic layer (left) and the shell with narrow stripe when T-morphotype could be recognised by the presence of scar only (right).

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

The M.trossulus prevalence (Ptros) varied broadly among populations situated on the coast of the Kola Peninsula from 0 to 0.98 (with 0.3 as median). The proportion of T-morphotype between M.trossulus (P(T|tros)) varied from ++ to ++ (++ median) whereas proportion of E-morphotype between M.edulis varied from ++ to ++ (++ median). The P(T|tros) and P(E|edu) values in W were rather similar to BL (P(T|tros) ++ and ++; P(E|edu) ++ and ++ correspondingly), whereas in BH these values were quite different (P(T|tros) ++; P(E|edu) ++).

Table ++. Parameters of regression models fitted

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Terms | Estimate | SE | z-statistic | p-value |
| **Model 1** | = 0.45 |  |  |  |
| (Intercept) | -2.9 | 0.17 | -17.08 | < 0.001 |
| Ptros | 5.2 | 0.35 | 14.73 | < 0.001 |
| Subset(BL) | 0.5 | 0.3 | 1.67 | 0.096 |
| Subset(BH) | 2.6 | 0.28 | 9.26 | < 0.001 |
| Ptros:Subset(BL) | -1.1 | 0.55 | -2.01 | 0.044 |
| Ptros:Subset(BH) | -2.3 | 0.58 | -3.88 | < 0.001 |
| **Model 2** | = 0.61 | = 0.67 |  |  |
| (Intercept) | -4.3 | 0.42 | -10.17 | < 0.001 |
| Ptros | 4.6 | 0.96 | 4.81 | < 0.001 |
| Subset(BL) | 0.8 | 0.88 | 0.96 | 0.335 |
| Subset(BH) | 4.4 | 0.89 | 4.92 | < 0.001 |
| Species(*M.trossulus*) | 4.3 | 0.5 | 8.47 | < 0.001 |
| Ptros:Subset(BL) | -2.2 | 1.78 | -1.25 | 0.211 |
| Ptros:Subset(BH) | -3.2 | 1.78 | -1.82 | 0.069 |
| Ptros:Species(*M.trossulus*) | -2.3 | 0.97 | -2.41 | 0.016 |
| Subset(BL):Species(*M.trossulus*) | -1.3 | 0.92 | -1.40 | 0.162 |
| Subset(BH):Species(*M.trossulus*) | -4.9 | 0.79 | -6.21 | < 0.001 |
| Ptros:Subset(BL):Species(*M.trossulus*) | 3 | 1.78 | 1.70 | 0.089 |
| Ptros:Subset(BH):Species(*M.trossulus*) | 5.5 | 1.57 | 3.50 | < 0.001 |
| sd\_(Intercept).pop | 0.8 |  |  |  |
| **Model 3** | = 0.2 |  |  |  |
| (Intercept) | 2.7 | 0.18 | 15.42 | < 0.001 |
| Ptros | -1.7 | 0.33 | -5.04 | < 0.001 |
| Subset(BL) | -0.9 | 0.32 | -2.67 | 0.008 |
| Subset(BH) | -3.2 | 0.27 | -11.82 | < 0.001 |
| Ptros:Subset(BL) | 1.2 | 0.55 | 2.17 | 0.03 |
| Ptros:Subset(BH) | 3.5 | 0.53 | 6.62 | < 0.001 |
| **Model 4** | = 0.4 | = 0.4 |  |  |
| (Intercept) | 3.8 | 0.25 | 15.09 | < 0.001 |
| Morph(T) | -3.7 | 0.36 | -10.33 | < 0.001 |
| Ptros | -5.1 | 0.52 | -9.89 | < 0.001 |
| Subset(BL) | -0.5 | 0.42 | -1.14 | 0.253 |
| Subset(BH) | -1.7 | 0.4 | -4.12 | < 0.001 |
| Morph(T):Ptros | 8.4 | 0.63 | 13.34 | < 0.001 |
| Morph(T):Subset(BL) | 0.4 | 0.39 | 0.91 | 0.361 |
| Morph(T):Subset(BH) | -1 | 0.33 | -2.95 | 0.003 |
| Ptros:Subset(BL) | 0.8 | 0.76 | 1.06 | 0.288 |
| Ptros:Subset(BH) | 2.3 | 0.75 | 3.00 | 0.003 |
| sd\_(Intercept).pop | 0.1 |  |  |  |
| **Model 5** | = 0.48 |  |  |  |
| (Intercept) | -2.9 | 0.17 | -16.94 | < 0.001 |
| PT | 6.1 | 0.4 | 15.25 | < 0.001 |
| Subset(BL) | 0.6 | 0.31 | 1.80 | 0.072 |
| Subset(BH) | -0.3 | 0.45 | -0.64 | 0.522 |
| PT:Subset(BL) | -0.3 | 0.74 | -0.43 | 0.669 |
| PT:Subset(BH) | -1.9 | 0.69 | -2.70 | 0.007 |
| **Model 6** | = 0.53 | = 0.53 |  |  |
| (Intercept) | 3.7 | 0.22 | 16.73 | < 0.001 |
| Morph(T) | -3.7 | 0.38 | -9.79 | < 0.001 |
| Ptros | -4.9 | 0.42 | -11.62 | < 0.001 |
| Subset(BH) | -1.6 | 0.48 | -3.37 | 0.001 |
| SubsetGOM | 1 | 0.56 | 1.81 | 0.07 |
| SubsetBALT | -1 | 0.34 | -3.02 | 0.003 |
| Morph(T):Ptros | 8.6 | 0.73 | 11.79 | < 0.001 |
| Morph(T):Subset(BH) | -0.9 | 0.67 | -1.30 | 0.193 |
| Morph(T):SubsetGOM | -1.6 | 0.85 | -1.88 | 0.06 |
| Morph(T):SubsetBALT | -1.1 | 2.92 | -0.37 | 0.712 |
| Ptros:Subset(BH) | 2.2 | 0.99 | 2.25 | 0.025 |
| Ptros:SubsetGOM | -3 | 1 | -2.99 | 0.003 |
| Ptros:SubsetBALT | -0.2 | 0.62 | -0.39 | 0.694 |
| Morph(T):Ptros:Subset(BH) | -0.6 | 1.34 | -0.41 | 0.68 |
| Morph(T):Ptros:SubsetGOM | 4.2 | 1.89 | 2.24 | 0.025 |
| Morph(T):Ptros:SubsetBALT | 6.2 | 7 | 0.89 | 0.373 |
| sd\_(Intercept).pop | 0 |  |  |  |
| **Model 7** | = 0.45 |  |  |  |
| (Intercept) | -2.7 | 0.14 | -19.47 | < 0.001 |
| PT | 6.2 | 0.33 | 18.45 | < 0.001 |
| Subset(BH) | -0.5 | 0.44 | -1.08 | 0.28 |
| SubsetGOM | 0.4 | 0.23 | 1.82 | 0.068 |
| SubsetBALT | 1.9 | 0.19 | 9.99 | < 0.001 |
| PT:Subset(BH) | -1.9 | 0.66 | -2.95 | 0.003 |
| PT:SubsetGOM | 0 | 0.77 | 0.05 | 0.957 |
| PT:SubsetBALT | 8.3 | 1.75 | 4.77 | < 0.001 |

The significant positive association between M.trossulus prevalence (Ptros) and frequency of T-morphotype was revealed in all three subsets around Kola peninsula (Model 1, Table +, Fig. +). The regression lines constructed from the Model 1 approached closely to Y=X line indicating the high proportionality between Ptros and PT in the case of W and BL. Importantly, points from testing data-sets for W and BL were concentrated closely around regression lines, i.e. model predictions was in a good agreement with relationships of genetic and phenotyping structure in populations which were not involved in model construction. Thus the Model 1 reflected realistic relations between genetic and phenotypic structures of populations in the case W and BL.

In the case of BH the regression line diverged significantly from the Y=X line taking a relatively higher position in comparison with regression lines constructed for W and BL. It means that populations from BH possessed a higher proportion of T-morphotype than in the case of W and BL for comparable Ptros range. Additionally it means that in populations dominated by M.edulis (low Ptros) the proportion of T-morphotype among mussels presented in such populations is much higher than in freshened area of Barents Sea. However, again, in BH the increase of Ptros was accompanied by a growth of T-morphotype proportion indicating some association between genetic and phenotyping population structure. Points corresponding to testing data set were concentrated around BH regression line thus showing rather good concordance to model prediction. It means that deviation of BH regression line from regression lines constructed for W and BL was a reflection of some naturally existing property of populations existing in the areas of the Barent Sea with normal oceanic salinity.

Analysis of frequency of T-morphotype among M.edulis (P(T|tros)) and among M.trossulus (P(T|edu)) revealed two patterns (Model 2, Table +, Fig. +). Firstly, a significantly higher proportion of T-morphotype among M.trossulus in comparison with M.edulis was revealed in the case of W and BL, but no such significant differences (the confidential intervals for both regression lines overlap, Fig. ++) were revealed in the case of BH due to higher T-morphotype frequency among M.edulis. Secondly in all cases some positive correlation of P(T|tros) and P(T|edu) with Ptros was found. The dependence of P(T|tros) and P(T|edu) on Ptros means that in those populations which dominated by M.trossulus (high Ptros) some part of M.edulis (minor taxa) starting to possess T-morphotype. Opposite pattern is observed in the populations with M.edulis dominance when M.trossulus being a minor taxa more frequently possess E-morphotype in compareson with populations where this specie is a major taxa (high Ptros).

Analysis of frequency of T-morphotype among M.trossulus (P(T|tros)) and among M.edulis (P(T|edu)) revealed two patterns (Model 2, Table +, Fig. +). Firstly, a significantly higher proportion of T-morphotype among M.trossulus in comparison with M.edulis was revealed in the case of W and BL, but no such significant differences (the confidential intervals for both regression lines overlap, Fig. ++) were revealed in the case of BH due to higher T-morphotype frequency among M.edulis. Secondly in all cases some positive correlation of P(T|tros) and P(T|edu) with Ptros was found. The dependence of P(T|tros) and P(T|edu) on Ptros means that in those populations which are dominated by M.trossulus (high Ptros) some part of M.edulis (minor taxa) starting to possess T-morphotype. Opposite pattern is observed in the populations with M.edulis dominance when M.trossulus being a minor taxa more frequently possessed E-morphotype in comparison with populations where this species is a major taxa (high Ptros).

## The value of morphotype-test for prediction of mussel’s taxa around Kola Peninsular

As it was expected, the predictive value of morphotype-test, i.e. probability of correct species identification for a mussel with known morphotype, was in high dependence on M.trossulus prevalence (Model 4, Table +, Fig. +). Taking into account that marginal and conditional pseudoR2 were close to each other for this model (Table ++) the role of random factor (population) as regulator of the model intercept was rather weak i.e. the association of probability of correct identification with mussel species equally reproduced in all populations.

Probability of correct identification of M.trossulus by T-morphotype (P(tros|T)) increased with increasing of Ptros whereas P(edu|E) demonstrated the opposite pattern in all subsets. Both P(tros|T) and P(edu|E) approached closely to 1 if Ptros was correspondingly maximal or minimal in the case of W and BL. It means that taking randomly chosen mussel of T-morphotype from a population with high M.trossulus prevalence we may be quite sure that the mussel belongs to M.trossulus. Oppositely if we take the mussel with E-morphotype from population dominated by M.edulis we could be sure that this individual could be classified as M.edulis. However in the case of E-morphotype from population dominated by M.trossulus (high Ptros) the probability of a correct identification of this mussel as M.edulis is negligable. The same is true for mussels with T-morphotype taken from population dominated by M.edulis (low Ptros). In the case of mixed populations (Ptros near to 0.5) we can believe with probability near to 0.7-0.8 that mussels with T-morphotype is M.trossulus and mussels with E-morphotype is M.edulis.

Qualitatively the same pattern was observed in the case of BH. However both P(tros|T) and P(edu|E) predicted by the model were lesser in BH than in the case of W and BL. It means that morphotype test has less predictive value in the case of areas the Barents Sea with normal oceanic salinity. However it should be noted that if Ptros is low the predictive value P(edu|E) is high enough. The same is true for P(tros|T) if Ptros is high.

Finally we tested the association of proportion of T-morphotype in population with proportion of M.trossulus (Model 5, Table +). Since the Model 5 is conceptually similar to Model 1 we do not present the visualisation of the model. This model reflects the significant positive association of Ptros with PT in all subsets. However, agein, the results for BH were significantly differ from W and BL.

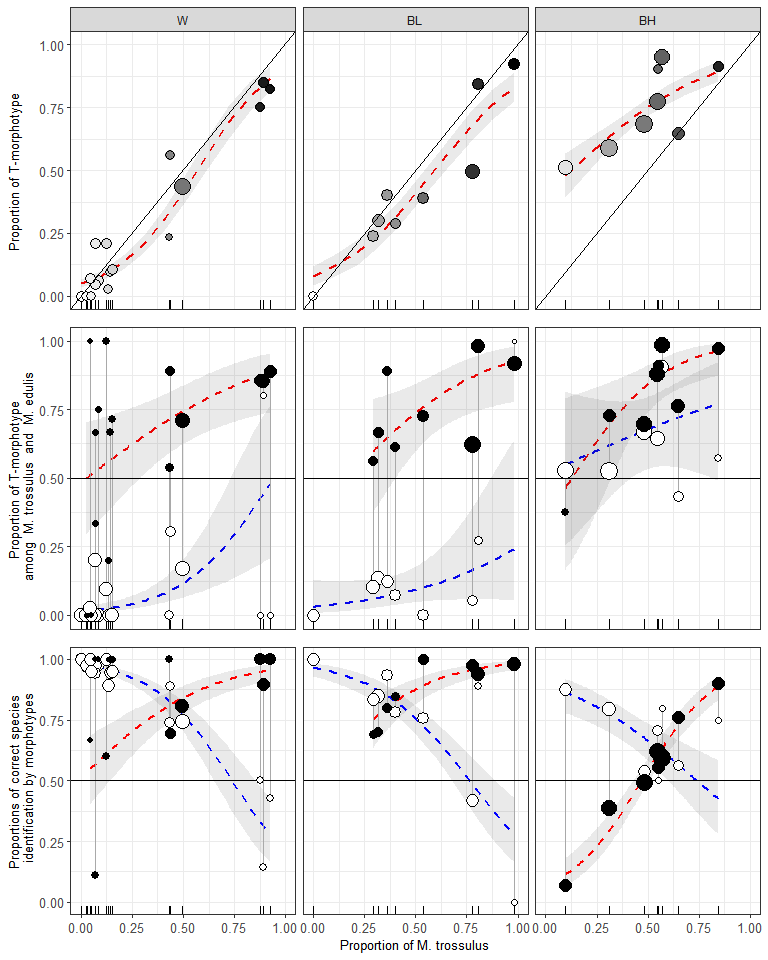


Figure ++. Visualisation of regression models. Initial data are presented as proportions of positive outcome in particular populations. Size of points is proportional to number of mussels in the particular sample. (A) Model1: the fill intensity is proportional to Ptros. (B) Model2: filled points - *M.trossulus* with T-morphotype; emty points - *M.edulis* with T-morphotype. (C) Model4: filled points - *M.trossulus* with T-morphotype; empty points - *M.edulis* with E-morphotype.

**For Material and Methods about dependence on mussel size**

Since the morphotype test was designed as a diagnostic tool for identification of mussel species and taxonomic structure of mixed populations we decided to avoid mussel size as a covariate in predictive models considered. However potentially mussel size could influence the probability of T-morphotype presence (see Katolikova et al., 2016). To check the possible correlation of T-morphotype presence with mussel size we prepared two types of assessment. Firstly we constructed a set of logistic regression models for each particular population with probability of T-morphotype presence as dependent variable and mussel size as predictor. When all models were fitted we considered those cases where slope-terms of the models were statistically significant (p < 0.05 after Hochberg’s correction for multiple testing, REF). Secondly we checked the presence of any patterns in residuals from the Model 5 (the main model which was designed to predict the probability of correct identification of individual mussel by its morphotype) as a function of mussel size.

Since the presence of T-morphotype potentially can be dependent on mussel size we assessed the statistical significance of the slope-terms for logistic regressions binding T-morphotype presence and mussel size in each of 37 populations. We found significant association only in 4 cases: one population was from W and three populations were closely situated in one common locality in BH. In all these cases slope-terms were negative indicating lesser probability to find T-morphotype for larger mussels. It should be noted that the mean mussels size calculated for these four populations was significantly larger than the mean size for mussels from all other populations (33.2 ± 16.33 and 26.7 ± 10.59 mm (Mean ± SD) respectively, t-test = 7.82, p < 0.001). Additionally we analysed the association of residuals from Model 4 with mussel size. No clear pattern was revealed in the residual plot (not shown).

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 W, BL and BH subsets. The model based on pooled W and BL (hereafter, WBL) and separated BH showed the lowest AIC = 1620.3. For the next analysis we will consider models based on this pooled data-set.

## Associations between morphotypes and species-specific genotypes around Atlantics

Material of samples from other parts of Atlantic being included into analysis did not change the patterns observed above (Model 6, Table +, Fig. ++). The probability of a correct identification of mussels with E-morphotype as M.edulis was high in both GOM and BALT populations dominated by M.edulis. This probability decreased significantly with Ptros increasing. The probability of a correct identification of M.trossulus by T-morphotype demonstrated opposite tendency.

## Associations between phenotypic and genotypic structure of populations around Atlantics

In all subsets the proportion of M.trossulus in populations was positively correlated with proportion of T-morphotype (Model 7, Table ++, Fig. ++). Similarly to WBL the regression line in the case of GOM closely approached the Y=X line. This indicates the high proportionality between frequency of T-morphotype and M.trossulus in populations situated in these two regions. In the case of SCOT both samples analyzed practically coincided with the Y=X line, indicating the same pattern as described above. However in the case of BALT the regression line slope was extremely high: regression line showing the positive association tendency rapidly diverged from Y=X line. It means that even the small proportion of mussels with T-morphotype indicate the high prevalence of M.trossulus in populations. Probably similar tendencies take place in the case of NORW.

The samples from BH showed an unusual tendency. In the case of all geographic areas, except BH, the common tendency was that most points corresponding to populations were presented above Y=X line. In the case of BH most of the points were situated below the Y=X line. It means that in this case the presence of mussels of T-morphotype in population is a weaker indicator of M.trossulus presence. However the slope was again positive and the regression line was practically in parallel with the Y=X line.

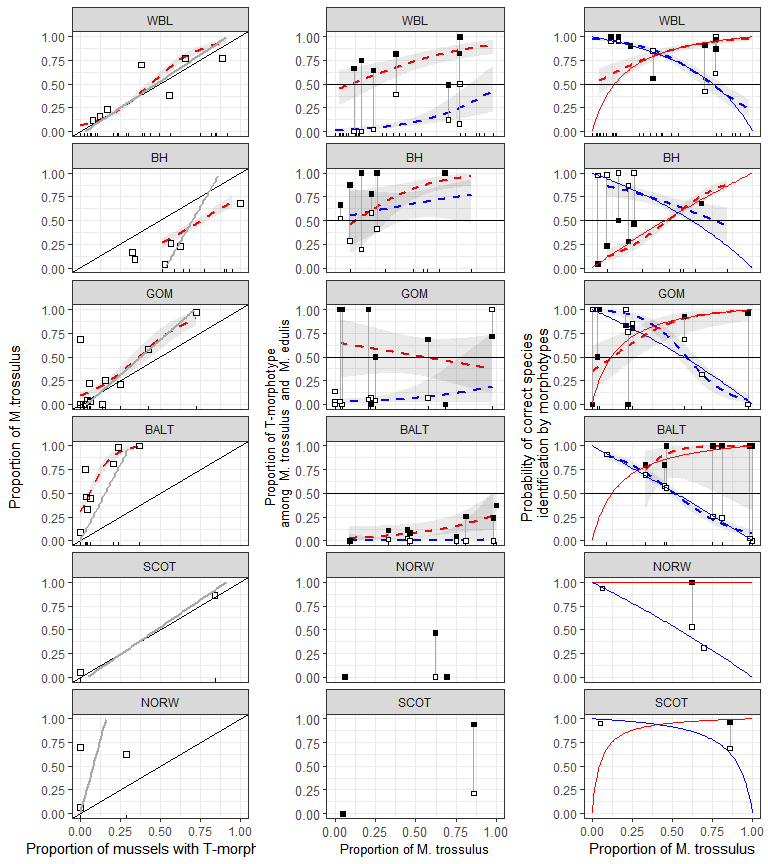


Fig ++. The applicability of morphotype test in different geographic regions. (A) Proportion of M.trossulus as a function of proportion of mussels with T-morphotype. Dashed lines represent regression Model 7. Solid gray lines indicate predictions made by Eq 3. (B) Frequency of T-morphotype among mussel of different species. Dashed lines represent regression Model 8. (С) Predictive value of morphotype test. Regression Model 6 is represented by dashed lines: P(tros|T) red dashed line, P(edu|E) blue dashed line. Solid lines indicate predictions made by Eq 1(red line) and Eq 2 (blue line).Points reflect the observed proportions in testing data set (WBL and BH) and in populations from other regions. Gray area around regression lines represent a 95% confidence interval.

## Predictive power of morphotype test and assessment of its accuracy

To compare the predictive power of morphotype test for each localities we calculated the values of P(tros|T) and P(edu|E) by Model 6 for the case when two species presented in equal proportions, i.e. Ptros = 0.5 (Table ++). If consider the level of probability of correct species identification equal to 0.5 (equal probability of correct and incorrect identification) as a threshold of acceptable probability then in some cases morphotype test produce not acceptable level of predictive value. In the case of BALT the lower boundary of 95% confidence interval for P(edu|E) is 0.46, which indicates impossibility of identification of M.edulis by E-morphotype in this region. Analogously in the case of BH, lower boundary confidence interval for P(tros|T) is 0.49, thus in this case we can not trust the identification of M.trossulus by T-morphotype. However, in all other cases the probability of correct identification was rather high indicating the possibility of using morphotype test. The highest predictive values of morphotype test was revealed in the case of WBL for both morphotypes. Thus working in this region we can be pretty sure in the results of morphotype test, i.e. randomly taking mussel of E-morphotype from the mixed population we can be sure that the mussel belong to M.edulis but randomly choosen mussel of T-morphotype could be assessed as M.trossulus. However in the case of GOM and BALT the probability of correct identification of M.trossulus by T-morphotype was also high indicating the possibility of using morphotype test for identification of M.trossulus by T-morphotype in these regions.

To assess the accuracy of morphotype test we applied the Model 7 to calculate the expected M.trossulus prevalence in populations from different subsets using the information on proportion of mussel of T-morphotype. Then the assessed Ptros was used to calculate the probability of correct species identification for each mussel (according to Model 6). Taking the value equal or more than 0.75 as acceptable level of correct identification probability we calculated the proportion of correctly identified mussels in testing data-set from WBL and BH and in all populations from GOM and BALT. Then we calculated the proportion of correctly identified mussels in the total amount of mussels sampled in each region (Table ++). The accuracy level for both morphotypes was rather high in the case of GOM and testing data-set from WBL. However in the case of testing data set from BH the high accuracy was found only if mussel of E-morphotype was identified as M.edulis. Similarly, in the case of BALT the high accuracy was found only if mussel of T-morphotype was identified as M.trossulus.

Table ++. Predicted values of probability of correct species identification by mussel morphotype in mixed populations (Ptros = 0.5) in different geographical regions. Low and upper boundaries of 95% conficencal intervals are given for predicted values.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Subset | P(edu|E) |  |  | P(tros|T) |  |  |
|  | Predicted | Low | Up | Predicted | Low | Up |
| WBL | 0.77 | 0.72 | 0.81 | 0.86 | 0.82 | 0.89 |
| BH | 0.67 | 0.59 | 0.74 | 0.54 | 0.49 | 0.59 |
| GOM | 0.67 | 0.57 | 0.76 | 0.86 | 0.7 | 0.95 |
| BALT | 0.51 | 0.46 | 0.57 | 0.94 | 0.66 | 0.99 |

Table +. Proportion of mussels correctly identified by morphotype test in different regions

|  |  |  |
| --- | --- | --- |
| Region | E-morphotype as M.edulis | T-morphotype as M.trossulus |
| WBL | 0.62 | 0.72 |
| BH | 0.89 | 0.10 |
| BALT | 0.00 | 0.96 |
| GOM | 0.82 | 0.62 |

## Using the probability theory equation (Eq 1, 2, 3) for express assessments by morphotype test

The use of Eq 1, 2 and 3 for express assessment of proportions of M.trossulus in a population (Ptros) and probability of correct species identification (P(edu|E) and P(tros|T)) using morphotype test, some calibrating samples should be collected. To work out the strategy for finding calibration samples we used the WBL subset which showed the best results of morphotype test performance (see above).

Considering all possible pairs of populations from WBL (406 possible pairs, Fig. ++) we found that predictions of Eq.3 and predictions of regression Model 7 (coefficients for WBL were used only) maximally correspond to each other if two calibrating populations possess maximal difference in M.trossulus prevalence.

The same analysis prepared to compare predictions of Eq.1 and Eq, 2 with predictions of regression Model 6 revealed the maximal congruency if populations with mixed structure (Ptros of both populations is near to 0.5) were taken as calibration samples.

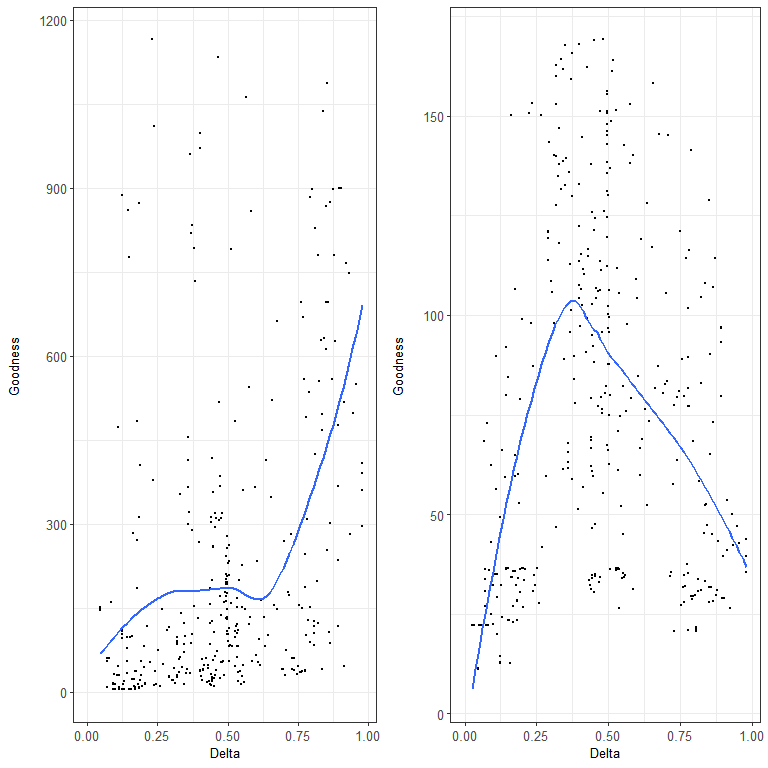


Figure +. Correspondence between regression and theoretical models. Each point corresponds to one of the possible pairs of populations from modelling data set (White Sea joined with low salinity Barens Sea). OX axis represents the differencу in genetic structre for each pair of populations. OY axis represents correspondence between prediction of regression model and theoretical model. Lines represent loess-smoother. (A ) Model 6 describing the dependence of proportion of M.trossulus (Ptros) on proportion of T-morphotype (P\_T) ; (B) Model 7 describing the dependence of probability of correct species identification (Pcorrect) on proportion of M.trossulus (Ptros) and morphotype (Morph).

Using the strategy described above we selected two maximally mixed populations in each region (except SCOT and NORW where we joined all samples and used all of them as calibrating sample) and calculated predictions by Eq1 and Eq2 for each geographical subsets (Fig. ++). In the case of WBL, BH and BALT predictions made by both equations were rather close to regression lines. It indicates high performance of these equations for prediction of P(tros|T) and P(edu|E) in these areas. In the case of GOM predictions of Eq1 was also in a good agreement with corresponding regression line. However the line predicted by Eq2 diverged from the regression line. Probability P(edu|E) predicted by Eq2 in the later case was slightly lesser than values predicted by regression analysis.

When the theoretical lines parameters of which were calculated from Eq. 3 were constructed we found rather good correspondence of these lines to regression line calculated from Model 7 (Fig. ++) in the case of WBL and GOM. In the case of BALT the line constructed by Eq3 diverged from the regression line but the slopes of both models were close to each other. Line constructed by Eq3 in the case of SCOT approached closely to the Y=X line and corresponded well to observed data. However in the case of BH and NORW poor correspondence of observed data and line predicted by Eq3 was found.