**Discussion**

— см. ниже Blue mussel researchers are often interested in the knowledge about taxonomic structure of populations and in row classification of individuals into “species” rather than in the exact knowledge about species identity or hybrid status of any mussel. The finding that *M. edulis* and *M. trossulus* differs sufficiently by shell morphotype frequencies in the White Sea (Katolikova et al. 2016) gave hope that these tasks could be achieved for these species by quick once-over of the inner side of shells only, without time- and cost consuming genotyping, and without soft tissues needed for genotyping (genotyping of shell material is possible (Geist et al. 2008; Der Sarkissian 2020) but is not a routine practice yet). Indeed, re-analyses of rich data from Katolikova et al. 2016 let us derive robust relationships between proportions of morphotypes in populations and their taxonomic structure, and between proportions of morphotypes in samples and the probabilities of mussels of different morphotypes being *M. trossulus* and *M. edulis*. These relationships could be safely used for prediction of taxonomic structure of any population in the White Sea. Moreover this allows identification of any mussel in equally mixed populations with the accuracy of about 80% (as it was suggested in the Introduction basing on frequencies of morphotypes in local *M. edulis* and *M. trossulus* genotypes); the greater the imbalance between species (and hence morphotypes) in population is, the more reliable is identification of dominant species but less reliable is identification of minor species.

The supreme goal of our study was to learn whether identification of *M. edulis* and *M. trossulus* by morphotypes is a unique “privilege” of the White Sea researchers or the same approach can be used for mussel identification worldwide.

The knowledge about the taxonomic structure of populations and the classification of individuals into “species” is often more valuable to the blue mussel researchers than the information about the species identity or the hybrid status of any given mussel. In the light of this, our finding that *M. edulis* and *M. trossulus* in the White Sea differed by the frequencies of the shell morphotype (Katolikova et al. 2016) seemed very promising. It gave hope that this knowledge could be obtained for these species by a quick examination of the inner side of the shells, without genotyping, which is expensive, time-consuming and requires soft tissues (genotyping of shell material is possible [Geist et al. 2008; Der Sarkissian 2020] but is not yet routine practice). In this study we reanalyzed abundant data from Katolikova et al. 2016 and derived robust relationships between the proportions of the morphotypes in the populations and their taxonomic structure as well as between the proportions of the morphotypes in samples and the probabilities of mussels of different morphotypes being *M. trossulus* and *M. edulis*. These relationships could be used for a reliable prediction of the taxonomic structure of any population in the White Sea. Moreover, any mussel in an equally mixed population could be identified with the accuracy of about 80%, based on the frequencies of the morphotypes in local *M. edulis* and *M. trossulus* genotypes (see Introduction). With the increasing imbalance between the species (and hence the morphotypes) in a population, the identification of the dominant species became more reliable though the identification of the minor species became less so.

Let our data on contact zones between species out of Northern Russia was limited, it is evident that A) the approach may be of value everywhere since interspecific differences in morphotype frequencies are ubiquitous and unidirectional, B) the utility of the approach cannot be the same for different contact zones due to considerable variation in morphotype frequencies among conspecific populations from different zones and sometimes also within zones. Below we shall first discuss patterns of morphotype frequency variation revealed and then will turn to the issue of the morphotype test application to different contact zones. In the closing section we shall touch limitations of single-marker taxonomic tests as applied to blue mussels and other taxa.

The ultimate goal of our study was to find out whether the possibility of identifying *M. edulis* and *M. trossulus* by the morphotype was a “privilege” of the researchers working at the White Sea or whether this approach could be used for identification of these two species worldwide. Though our data on the contact zones between the species outside Northern Russia were limited, our results indicate that this approach may be useful everywhere since interspecific differences in the morphotype frequencies were ubiquitous and unidirectional. However, its utility is evidently different for different contact zones due to a considerable variation in the morphotype frequencies in conspecific populations from different zones and sometimes also from the same zone.

We will start with the description of the patterns of variation of the morphotype frequencies revealed in our study. Then we will discuss the applicability of the morphotype test in different contact zones. In the closing section, the limitations of single-marker taxonomic tests for blue mussels and other taxa will be outlined.

**Geography, salinity-related factors in the subarctic, cryptic factors and taxonomic structure of mixed populations affect morphotype frequencies in conspecifics**

**Subsection как утвердительное предложение — неудачно. И вообще название длинное и непонятное.**

**Factors affecting morphotype frequencies in conspecifics — может, как-то так?**

Some variation in morphotype frequencies was observed among putatively pure conspecific populations sampled at a distance from contact zones. *M. edulis* from temperate seas (i.e. excluding northernmost samples from the eastern Barents Sea and Greenland) appeared to be nearly monomorphic for E-morphotype while northern samples were more polymorphic and diverse. In its turn, reference populations of *M. trossulus* from Western Pacific and from Washington in Eastern Pacific were nearly monomorphic for T-morphotype. Nevertheless we cannot say with confidence that *M. trossulus* lacks geographic variation in its ancestral range in the Pacific and that T-morphotype in an “ancestral” state for this species. Zolotarev (2002) identified morphotypes in small samples representing genotyped collections from the study of McDonald et al. (1991). He generally got results similar to ours but reported elevated frequencies of E-morphotypesin *M. trossulus* from Oregon (East Pacific). His data should be treated with caution because he used another, more fractional classification of morphotypes, identified morphotypes macroscopically, and because Oregon is close to a contact zone between *M. trossulus* and *M. galloprovincialis* (McDonald et al. 1991); mussels of the later species are marked by E-morphotypes (Zolotarev, Shurova 1997; Zolotarev, 2002).

Some variation in the morphotype frequencies was observed among putatively pure conspecific populations sampled at a distance from the contact zones. The samples of *M. edulis* from the temperate seas (i.e. all except those from the eastern Barents Sea and Greenland) appeared to be nearly monomorphic for the E-morphotype while the northern samples were more polymorphic and diverse. In turn, the reference populations of *M. trossulus* from Western Pacific and from Washington in Eastern Pacific were nearly monomorphic for the T-morphotype.

Nevertheless we cannot be sure that *M. trossulus* lacks geographic variation in its ancestral range in the Pacific and that the T-morphotype in an “ancestral” state for this species. Zolotarev (2002) identified the morphotypes in small samples representing genotyped collections from McDonald et al. (1991). His results, while generally similar to ours, indicated elevated frequencies of the E-morphotype in *M. trossulus* from Oregon (East Pacific). However, Zolotarev’s data should be treated with caution, because he used a more fractional classification of the morphotypes and identified them macroscopically, and also because Oregon is close to a contact zone between *M. trossulus* and *M. galloprovincialis* (McDonald et al. 1991), and the latter species is characterized by the E-morphotype (Zolotarev, Shurova 1997; Zolotarev, 2002).

In *M. trossulus* variation among contact zones was primarily due to elevated frequencies of E-morphotypes in Norway and, especially, in the Baltic Sea while variation within particular contact zones was primarily due to few “outlier” samples in the Gulf of Maine and Norway. Unlike *M. trossulus*, *M. edulis* demonstrated small variation among zones (universally low frequency of T-morphotype), with one exception. In Northern Russia, elevated (up 40%) frequency of T-morphotypes was observed, but only in samples from full saline localities (salinity above 30 ppt). Similar salinity-related variation was also present in *M. edulis* from the eastern Barents Sea, at a distance from contact zone between species along Kola Peninsula coast. Finally, analysis of rich material from the White and Barents Seas demonstrated variation in morphotype frequencies related with taxonomic structure of populations: with increasing prevalence of *M. trossulus* in samples frequencies of T-morphotypes increased both among *M. edulis* and *M. trossulus* genotypes.

In *M. trossulus* the variation in the morphotype frequencies between the contact zones was mostly associated with the elevated frequencies of the E-morphotypes in Norway and, especially, in the Baltic Sea. The variation within contact zones was mostly due to the few “outlier” samples from the Gulf of Maine and Norway. On the contrary, *M. edulis* demonstrated small variation between zones, the frequency of the T-morphotype being universally low. There was, however, one notable exception. An elevated (up to 40%) frequency of the T-morphotypes was observed in samples from full saline localities (salinity above 30 ppt) in Northern Russia. A similar salinity-related variation was also present in *M. edulis* from the eastern Barents Sea, at some distance from the contact zone between these species along the Kola Peninsula coast.

Finally, an analysis of the abundant material from the White and the Barents Sea revealed variation in the morphotype frequencies related with the taxonomic structure of the populations. The frequencies of the T-morphotype increased both among *M. edulis* and among *M. trossulus* genotypes with the increasing prevalence of *M. trossulus* in the samples.

Two hypotheses, not completely mutually exclusive, could explain extremely high frequencies of E-morphotypes in *M. trossulus* from the Baltic Sea and Norway. One hypothesis likens morphotypes (more specifically - hypothetical genes underlying morphotypes) to alleles of taxonomically diagnostic loci that can introgress between species as a result of extensive hybridization and backcrossing. From genetic studies it is known that the Baltic *M. trossulus* hybridize more freely with *M. edulis* and is stronger introgressed by *M. edulis* genes than any other Atlantic population (Vainola, Strelkov 2011; Fraisse et al. 2016); due to its mixed genetic nature the Baltic mussel is sometimes interpreted as unique *M. edulis* x *M. trossulus* hybrid swarm fundamentally different from the “oceanic” *M. trossulus* (Vainola, Strelkov 2011). Data on Norwegian population is more limited, however it is evident that hybridization there is more extensive than in other contact zones but the Baltic one (Vainola, Strelkov 2011; Wenne et al 2020) and that local *M. trossulus* populations could be strongly introgressed by *M. edulis* genes (Śmietanka, Burzyński 2017).

Two hypotheses, which are not mutually exclusive, can be offered to explain extremely high frequencies of the E-morphotype in *M. trossulus* from the Baltic Sea and Norway. One hypothesis likens the morphotypes or, more specifically, the underlying hypothetical genes, to alleles of taxonomically diagnostic loci that can introgress between species as a result of extensive hybridization and backcrossing. Genetic studies show that the Baltic *M. trossulus* hybridizes more freely with *M. edulis* and is stronger introgressed by *M. edulis* genes than any other Atlantic population (Vainola, Strelkov 2011; Fraisse et al. 2016). Due to its mixed genetic nature, the Baltic mussel is sometimes considered as a unique *M. edulis* x *M. trossulus* hybrid swarm, which is fundamentally different from the “oceanic” *M. trossulus* (Vainola, Strelkov 2011). While the data from Norway are limited, hybridization is apparently more extensive there than in most other contact zones though not as extensive as in the Baltic (Vainola, Strelkov 2011; Wenne et al. 2020). Besides, it is evident that the local Norwegian *M. trossulus* populations can be strongly introgressed by *M. edulis* genes (Śmietanka, Burzyński 2017).

Another hypothesis states that there are environmental factors, uncovered in this study, both geographical and local, reducing T-morphotype frequencies in *M. trossulus*. Our “outlier” samples with nearly zeros frequencies of T-morphotypes - one in Norway, sampled nearly in the same place as two other Bergen samples, and two from ??? (название места) in the Gulf of Maine we explain by the action of some local cryptic factors rather than methodological causes like mussel mislabeling in the collections. While local factors putatively governing morphotype frequencies in *M. trossulus* remained cryptic, in the Barents Sea we identified “complex” factor governing morphotype frequencies in *M. edilus* - the salinity or a factor(s) linked to salinity. This variation was also evident in the eastern - coldest part of the Barents Sea. The border between more temperate populations of *M.edulis* with “normally” high frequencies of E-morphotypes and more Arctic populations with decreased frequencies of E-morphotypes in oceanic habitats runs somewhere between North Cape and Kola Bay (Fig. 1) - in the area with mean annual, summer and winter sea surface temperatures about 6°C, 9°C and 4°C, respectively (*http://esimo.oceanography.ru/).* Because decreased frequencies of E-morphotypes were revealed also in populations of *M. edulis* from Greenland and the Gulf of St. Lawrence (full saline habitats), we suspect that salinity-related variation could be present in high latitudes of the West Atlantic as well.

Переписала этот абзац в меру своего понимания, не судите строго.

According to the second hypothesis, the frequency of the T-morphotype in *M. trossulus* is reduced under the influence of some environmental factors, geographical or local. We managed to identify one such factor in the Barents Sea: salinity or a factor/factors linked to salinity. The eastern part of the Barents Sea, where this variation was evident, is also the coldest. The border between the more temperate populations of *M.edulis* with “normal” (high) frequencies of the E-morphotype and the more Arctic populations with lower frequencies of the E-morphotype in oceanic habitats runs somewhere between North Cape and Kola Bay (Fig. 1). This area has mean annual, summer and winter sea surface temperatures of about 6°C, 9°C and 4°C, respectively (*http://esimo.oceanography.ru/).* Since lower frequencies of the E-morphotype were also revealed in populations of *M. edulis* from Greenland and the Gulf of St. Lawrence (full saline habitats), we suspect that salinity-related variation could be present in high latitudes of Western Atlantic as well.

The local factors possibly affecting the morphotype frequencies in *M. trossulus* remain unknown. Nevertheless, we suspect that the nearly zeros frequencies of the T-morphotype in the “outlier” samples (one from Norway, almost from the same place as the other Bergen samples, and two from ??? (название места) in the Gulf of Maine) could be explained by the impact of some cryptic local factors, though a more prosaic explanation such as the mislabeling of mussels in the collections cannot be entirely ruled out.

Мне кажется, что с этого места название subsection перестало быть валидным.

The functional significance of the morphological features underlining E- and T- morphotypes – the presence/absence of a pearl layer under the ligament is unclear, but we suspect that morphotypes could differ in conspecifics by the degree of development of the pearl layer per se, therefore the thickness and strength of the shell (the nacreous shell layer is mechanically the strongest one, Currey and Taylor, 1974). As *M. trossulus* (species usually marked by T-morphotype) generally has more poor pearl layer and fragile shell than *M. edulis* (Beaumont et al. 2008, see also below), *M. edulis* of T-morphotype could have underdeveloped pearl layer and thinner shells than *M. edulis* of E-morphotype. Are differences in the shell thickness and structure expected between high- (oceanic) and low- salinity (estuarine) environments in the Arctic? Apart from low temperatures, Arctic sea is characterized by reduced concentration of calcium carbonates in the water (Steinacher et al. 2009) and, seasonally, low concentrations of food (planktonic algae) for mussels (Zenkevitch 1963). Furthermore, estuarine habitats are generally characterized by lowest carbonate saturations states but also by highest food (seston) concentrations due to riverine discharge (Duarte et al. 2020), as exemplified by highest biomasses of *Mytilus* in estuaries in the Barents Sea (Bufalova et al. 2005) and everywhere (Seed, Suchanek 1992). Both calcium carbonates and energy are needed for shell growth and maintenance. In estuaries, the nacreous inner layer of the mussel shell is prone to dissolution and corrosion (Melzner et al. 2011) but mussels can maintain their shells strong if supplied with sufficient food (Melzner et al. 2011; Duarte et al. 2020). Under food limited conditions, energy could be allocated to somatic mass maintenance instead of shell conservation (Melzner et al. 2011 and references therein). Our hypothesis to explain assumed differences in the degree of pearl layer development between *M. edulis* from estuarine and from saline localities in the Arctic is that in estuaries mussels allocate energy into shell maintenance and keep their nacreous shell layer thick while in less acidic but more famine oceanic habitats they allocate energy in somatic growth keeping their nacreous layer thin. In the result, the majority of *M. edulis* from saline localities lacks the pearl layer under the ligament. It is noteworthy that in the same populations where *M. edulis* demonstrated the salinity-related variation, morphotype frequencies in *M. trossulus* varied negligibly. This could be attributed to the generally lower shell plasticity in “oceanic” (non-Baltic) *M. trossulus* than in *M. edulis* in reaction to environmental stressors (Lowen et al., 2013, see Khaitov et al. 2018 for more discussion).

The functional significance of the morphological character underlying the E- and the T- morphotype—the presence/absence of the nacreous layer under the ligament—is unclear. However, we suspect that the morphotypes might differ in conspecifics by the degree of development of the nacreous layer itself and thus in the thickness and strength of the shell. The nacreous shell layer is mechanically the strongest (Currey and Taylor, 1974). *M. trossulus*, which is usually marked by the T-morphotype, generally has a thinner nacreous layer and a more fragile shell than *M. edulis* (Beaumont et al. 2008, see also below). *M. edulis* of the T-morphotype might have an underdeveloped nacreous layer and a thinner shell than the conspecifics of the E-morphotype.

Can we expect the shell thickness and structure to differ in mussels from saline (oceanic) and brackish (estuarine) environments in the Arctic? Apart from the low temperatures, the Arctic Sea is characterized by a reduced concentration of calcium carbonates (Steinacher et al. 2009) and, seasonally, by low concentrations of planktonic algae, which the mussels feed on (Zenkevitch 1963). Estuarine habitats are generally characterized by the lowest saturation of carbonates but the highest concentrations of food (seston), which is due to the riverine discharge (Duarte et al. 2020). This is exemplified by the highest biomasses of *Mytilus* in estuaries in the Barents Sea (Bufalova et al. 2005) and elsewhere (Seed, Suchanek 1992). Mussels need both calcium carbonates and energy for shell growth and maintenance. In estuaries, the nacreous layer of the mussel shell is prone to dissolution and corrosion (Melzner et al. 2011) but the mussels can still keep their shells strong if the food is sufficient (Melzner et al. 2011; Duarte et al. 2020). If the food is limited, the energy is likely to be allocated to the maintenance of the somatic mass rather than the conservation of the shell (Melzner et al. 2011 and references therein).

Our hypothesis explaining the assumed differences in the degree of the nacreous layer development between *M. edulis* from the brackish and the saline localities in the Arctic is that in the estuaries the mussels allocate more energy for shell maintenance thus keeping their nacreous layer thick while in less acidic but more famished oceanic habitats they allocate more energy for somatic growth keeping their nacreous layer thin. As a result, the majority of *M. edulis* from the saline localities lack the nacreous layer under the ligament. It is noteworthy that in the populations where *M. edulis* demonstrated salinity-related variation, the morphotype frequencies in *M. trossulus* varied negligibly. This could be attributed to a generally lower shell plasticity in “oceanic” (non-Baltic) *M. trossulus* than in *M. edulis* in response to the environmental stressors (Lowen et al., 2013, see Khaitov et al. 2018 for more discussion).

A positive correlation of T-morphotype frequencies both in *M. edulis* and *M. trossulus* with *M. trossulus* prevalence in the representative data sets from the White and Barents Seas is an expected result remembering how *M. edulis* and *M. trossulus* genotypes have been defined: by dominance of conspecific genes in multilocus genotypes. Hence both genotypes included purebreds and hybrids. From detailed analysis of the White Sea data used here in Katolikova et al (2016) we know that frequencies of hybrids are about the same in all samples (18% on average), hybrids are intermediate in morphotype frequencies between purebred *M. edulis* and *M. trossulus* but usually closer to species dominating the population (Katolikova et al. 2016). This means that in our analyses “*M. edulis* genotypes” from *M. trossulus* dominated populations included mostly hybrids with increased frequency of T-morphotypes relative to such genotypes in *M.edulis*-dominated populations. In turn, “*M. trossulus* genotypes” from *M. edulis* dominated populations included mostly hybrids with decreased frequency of T-morphotypes relative to such genotypes in *M.trossulus*-dominated populations. From here the observed unidirectional variation in morphotype frequencies among *M. edulis* and *M. trossulus* genotypes with taxonomic structure of populations. To note, variation of sensitivity and specificity with disease prevalence is often observed in clinical diagnostic tests (Leeflang et al. 2009, 2013). For example, a patient population with a higher disease prevalence may include more severely diseased patients, therefore, the test performs better in this population (Leeflang et al. 2009).

A positive correlation of the T-morphotype frequencies both in *M. edulis* and *M. trossulus* with the prevalence of *M. trossulus* in the representative data sets from the White and the Barents Sea was to be expected, bearing in mind that *M. edulis* and *M. trossulus* genotypes are defined by the dominance of the conspecific genes in multilocus genotypes. Hence both genotypes included purebreds as well as hybrids. From a detailed analysis of the White Sea data (Katolikova et al. 2016) we know that the frequencies of hybrids are approximately the same in all the samples (18% on the average) and that the hybrids are intermediate in the morphotype frequencies between purebred *M. edulis* and purebred *M. trossulus* but are usually closer to the species dominating in the population (Katolikova et al. 2016). This means that in our analyses “*M. edulis* genotypes” from *M. trossulus*-dominated populations included mostly hybrids with an increased frequency of the T-morphotype as compared to the “*M. edulis* genotypes” in *M.edulis*-dominated populations. In turn, “*M. trossulus* genotypes” from *M. edulis*-dominated populations included mostly hybrids with a decreased frequency of the T-morphotype as compared to such genotypes in *M.trossulus*-dominated populations. This is the cause of the observed unidirectional variation in the morphotype frequencies among *M. edulis* and *M. trossulus* genotypes with the changing taxonomic structure of populations. To note, the variation of sensitivity and specificity of clinical diagnostic tests with the changing disease prevalence is often observed (Leeflang et al. 2009, 2013). For instance, a test is likely to perform better a patient population with a higher disease prevalence, which may include more severely diseased patients (Leeflang et al. 2009). **Applications of morphotype test**

We consider four а почему их дальше три? fields where the morphotype test could be universally useful as a cheap alternative of genotyping: (1) monitoring of taxonomic structure of commercial populations and wild monitored populations like used in the “mussel watch” contaminant monitoring programs (deviations in morphotype frequencies could be a warning of taxonomic change), (2) mapping of species distribution (detailed mapping could require numerous samples due to usually high mosaicism in distribution of species in contact zones, see Katolikova et al. 2016 and references therein), (3) interpretation of taxonomic structure of natural history collections the same as any samples of dead shells like empty shells some mussel predators leave behind.

Я понимаю желание изложить по пунктам, но тут получается, что весь абзац — одно предложение, это не годится.

In our opinion, the morphotype test can be universally applied as an alternative to genotyping in three fields. Firstly, it can be used for monitoring the taxonomic structure of commercial and wild populations, e.g. in the “mussel watch” contaminant monitoring programs, because deviations of the morphotype frequencies may be a warning sign of the taxonomic change. Secondly, it may prove useful for mapping the species distribution. Detailed mapping is likely to require a great number of samples because the distribution of the species in contact zones is usually highly mosaic (see Katolikova et al. 2016 and references therein). A cheap and fast alternative to genotyping is very desirable in this case. Thirdly, the morphotype test can be used when only dead mussel shells are available, e.g. for interpretations of the taxonomic structure of natural history collections or samples of dead shells left behind by some mussel predators.

Тут, по-моему, опять новая тема. Нужен новый подзаголовок?

The reliable application of the test requires good genotyped references, ideally - empirical relationships between morphotype frequencies and taxonomic structure of populations, different for different contact zones, as provided here (Table ?). For all contact zones but in Northern Russia our regressions require further refinement since relatively small numbers of samples were included. For mixed populations from the Baltic and the Gulf of Maine the same as for such populations from Northwestern Greenland and American coast north from the Gulf of Maine, unstudied by us, collections of genotyped mussels should remain from previous extensive population-genetic studies (e.g. MacDonald et al. 1991; Bates, Innes 1995; Rawson et al. 2001; Stuckas et al. 2017; Wenne et al. 2020). These collections could be used for further calibration of the test. If such effort will be undertaken for Greenland and subarctic American populations, salinity should be considered as a potential covariate of morphotype variation.

A reliable application of the morphotype test requires good genotyped references. Ideally, empirical relationships should be established between the morphotype frequencies and the taxonomic structure of populations in a given contact zone, as they were in our study (Table ?). Even our regressions require further refinement for all the contact zones except Northern Russia, since they are based on a relatively small number of samples. On a reassuring note, for mixed populations from the Baltic and the Gulf of Maine as well as for the populations from Northwestern Greenland and the American coast north of the Gulf of Maine unexamined in this study, collections of genotyped mussels probably remain from previous extensive population genetic studies (e.g. MacDonald et al. 1991; Bates, Innes 1995; Rawson et al. 2001; Stuckas et al. 2017; Wenne et al. 2020). These collections could be used for further calibration of the morphotype test. If such an effort is undertaken for Greenland and subarctic American populations, salinity should be considered as a potential covariate of the morphotype variation.

For understudied contact zones and for new zones, if discovered, relationships should be established de-novo; if genotyping of more than few samples covering the range of morphotype frequencies will prove impractical, relationships could be approximated using data on at least two genotyped samples with maximally contrast structure (ideally – pure *M. edulis* and *M. trossulus*) and the “genotype to morphotype calculator” (eq. 3). We claim that morphotypes could be useful both for detection of new contact zones and for their formal genetic description – preliminary selection, by morphotype frequencies, of most pure samples or subsamples needed for verification of species identity and of most mixed ones needed for assessment of the extent of hybridization and mixing. Indeed our exercise with prediction, using eq. 3, of taxonomic structure of the White Sea populations based solely on maximum and minimum morphotype frequencies in regional populations brought satisfactory results.

The relationships between the morphotype frequencies and the taxonomic structure of populations will have to be established *de novo* in understudied or, potentially, new contact zones. Should the genotyping of more than a few samples covering the range of the morphotype frequencies prove impractical, the relationships could be approximated using the data on at least two genotyped samples with the most contrasting structure (ideally, pure *M. edulis* and pure *M. trossulus*) and the “genotype to morphotype calculator” (eq. 3).

We claim that the morphotype test may be useful for the detection of new contact zones and for their formal genetic description. The procedure would involve a preliminary selection, with the help of the morphotype frequencies, of the purest samples or subsamples needed for the verification of the species identity and of most mixed ones needed for the assessment of the extent of hybridization and mixing. Indeed, we predicted the taxonomic structure of the White Sea populations based solely on maximum and minimum morphotype frequencies in regional populations using eq. 3, and the results of this prediction were quite satisfactory.

In case of historical or archaeological collections the only way to translate the proportion of T-morphotypes in samples into taxonomic structure is to use actualistic principle. If assessment of correspondence between morphotypes and genotypes was undertaken for the area of sample origin then one could use this information for retrognosis. Certainly such assessments are possible for quantitatively representative samples but not for small samples or singular shells. We are pessimistic about the utility of morphotypes for interpretation of paleontological data since both geography and local oceanographic conditions, variable at long timescale, seem to affect morphotype frequencies in conspecifics.

In case of historical or archaeological collections, the only way to translate the proportion of the T-morphotypes in the samples into the taxonomic structure is to resort to the actualistic principle. If the correspondence between the morphotypes and the genotypes was assessed in the area of the sample origin, one can use this information for retrognosis. This should be possible for quantitatively representative samples though not for small samples or single shells. Unfortunately, the morphotype test is unlikely to be useful for the interpretation of paleontological data since the morphotype frequencies in conspecifics are affected both by geography and by the local oceanographic conditions, which are variable at a large time scale.

Individual mussel identification by morphotypes indeed seems to be a confirmed “privilege” of the White Sea researchers the same as researches dealing with mussels from the Barents Sea brackish water environments. There are prospects for use of the test for individual assignment also in the Gulf of Maine, if one does not count for abovementioned outlier samples (раскрыть какая надежность для лучших не-аутлаер самплез), and possibly in Scottish populations that were presented in our analysis by only two samples, unfortunately. Due to shifted morphotype frequencies between species, only *M. trossulus* mussels could be reliably identified by morphotypes in the Baltic Sea and Norway and only *M. edulis* - in the saline water areas in the Barents Sea.

The possibility to identify individual mussels by the morphotype seems to be the “privilege” of researchers working at the White Sea and brackish environments of the Barents Sea. The morphotype test also seems to be promising for individual assignment in the Gulf of Maine, except in the outlier samples (see above) (раскрыть какая надежность для лучших не-аутлаер самплез) and, possibly, in Scotland (unfortunately, the Scottish populations were represented in our analysis only by two samples). In the Baltic Sea and Norway the morphotype test worked reliably only for *M. trossulus* mussels, while in the saline areas in the Barents Sea it did so only for *M. edulis* mussels, which was due to shifting morphotype frequencies between the species.

As an example of application of the test for individual assignment, our study can be given (Khaitov et al. 2018). Aiming to learn whether starfishes *Asterias rubens* distinguish between *M. edulis* and *M. trossulus* in the White Sea, we sampled mussels in populations with high and low T-morphotype frequencies, mixed them in equal proportions in experimental cages, and, after acclimation to ambient conditions, offered to starfishes. Starfishes selectively consumed mussels of T-morphotypes which was interpreted as their preference towards *M. trossulus* (Khaitov et al. 2018). Now we know that an alternative and probably more formal experimental design could be to use sympatric mussels from the most mixed population. Under both designs, the accuracy of individual assignment of experimental mussels would be nearly the same (on average ??% in case when most pure populations and ??% when most mixed one were used as a source of experimental mussels). По-моему, последние два предложения — это что-то очень инсайдерское, со стороны непонятное.

An example of the application of the morphotype test for individual assignment can be found in our previous study (Khaitov et al. 2018). Aiming to find out whether the starfish *Asterias rubens* distinguished between *M. edulis* and *M. trossulus* in the White Sea, we sampled mussels in populations with high and low frequencies of the T-morphotype, mixed them in equal proportions in experimental cages, and, after acclimation to ambient conditions, offered to the starfishes. These predators selectively consumed mussels of the T-morphotype, which was interpreted as a preference towards *M. trossulus* (Khaitov et al. 2018). Now we know that an alternative and probably more formal experimental design could be to use sympatric mussels from the most mixed population. Under both designs, the accuracy of individual assignment of experimental mussels would be nearly the same: on the average, ??%, when experimental mussels are taken from the purest population and ??%, when they are taken from the most mixed population).

While we do not recommend using morphotype test for individual assignment without reliable genetic references (either empirical relationships between proportions of morphotypes in samples and the probabilities of mussels of different morphotypes being M. trossulus and M. edulis, as available for Northern Russia populations, or control genotyping of mussels from populations of interest as could be recommended for other regions), the knowledge that the accuracy of individual identification of mussels from the Whites Sea populations could be predicted basing solely on morphotype frequencies in three reference population samples – with maximum, with minimum and with intermediate morphotype frequencies, and eq. 1-3 ( … раскрыть процедуру), could be helpful. Хочу отметить, что это ещё один абзац, состоящий из одного предложения.

We would like to stress that, if one plans to use the morphotype test for individual assignment, reliable genetic references are absolutely indispensable. These could be either empirical relationships between the proportions of the morphotypes in the samples and the probabilities of mussels of different morphotypes being *M. trossulus* or *M.* edulis, which is available for the Northern Russia populations, or control genotyping of mussels from the populations of interest. Still, it is noteworthy that the accuracy of individual identification of mussels from the White Sea populations could be predicted basing solely on the morphotype frequencies in three reference population samples (those with the maximum, the minimum and the intermediate morphotype frequencies) and eq. 1-3 ( … раскрыть процедуру). This knowledge might prove helpful for other researchers.

Morphotype test is not without pitfalls. Evident risks are underestimation of *M. trossulus* in some non-Russian populations like sources of “outlier samples” in Norway and Gulf of Maine, the bias generated by non-random association of morphotypes with size (or age) of conspecific mussels as was observed in very rare (2?%) samples, uncertainties in application of the test to populations from intermediate salinities (about 30 ppt) in the Barents Sea. To note four from five samples with significant non-random association of morphotypes with size were from Tyuva inlet at the very border between low- and full-saline areas of Kola Bay; hence temporal or ontogenetic trends in morphotype frequencies are possibly local Barents Sea phenomenon related by peculiar salinity conditions as in Tyuva.

The morphotype test is not without pitfalls. Evident risks are an underestimation of *M. trossulus* in some populations (such as those in Norway and the Gulf of Maine, which were the sources of the “outlier” samples), the bias generated by a non-random association of morphotypes with size (or age) of conspecific mussels such as was observed in very rare (2?%) samples, and uncertainties in its application to populations from intermediate salinities (about 30 ppt) in the Barents Sea. To note, four out of the five samples with a significant non-random association between the morphotypes and the size were from Tyuva inlet right at the border between brackish and full saline areas of the Kola Bay. It is possible that temporal or ontogenetic trends in the morphotype frequencies are a local Barents Sea phenomenon related to the unusual salinity conditions.

In spite of the fact that the hypotheses that different mussel species could differ by the extent of the nacreous layer development under the ligament nympha was suggested long time ago (Zolotarev, Shurova 1997; Vervoenen et al. 2000) the cited study of Khaitov et al. (2018) is the only one where morphotypes were purposefully used for identification of species. We hope that after the morphotype test was evaluated (i.e. associations among morphotypes and species-specific genotypes were analyzed at individual and population levels) it will receive more attention from blue mussel researchers. To note, other morphological express-method for *M. trossulus* and *M. edulis* diagnostics was suggested by Beaumont et al. 2008 who confirmed that commercially damaging “fragile mussels” recorded on *M. edulis* plantations in Loch Etive (Scotland) were genetically similar to *M. trossulus*. “Fragile” mussels appeared to differ from *M. edulis* (and reference *M. galloprovincialis*) by more flexible and elongated shells with more distinctive growth ridges and not whitish but a duller grey-blue colour of the inside. The perspective approach for identification of species basing on the shell “fragility” remains underdeveloped; By comparing photos of shells in Beaumont et al. 2008 with photos of shells from our the Barents Sea samples (ESM Fig. ?B) one can see that differences between species are not so striking in the Barents Sea as in Scotland.

Though the hypotheses that different mussel species may differ by the extent of the nacreous layer development under the ligament nympha has been suggested a long time ago (Zolotarev, Shurova 1997; Vervoenen et al. 2000), the morphotypes were actually put to use for the identification of species only in the study by Khaitov et al. (2018) (see above). Here we show that the morphotype test is a promising tool. Once it has been evaluated, i.e. associations between morphotypes and species-specific genotypes have been analyzed at the individual and the population level, it will hopefully deserve more attention from the blue mussel researchers.

To note, another morphological express method for the diagnostics of *M. trossulus* and *M. edulis* was suggested by Beaumont et al. (2008) who showed that commercially damaging “fragile mussels” recorded on *M. edulis* plantations in Loch Etive (Scotland) were genetically similar to *M. trossulus*. The “fragile” mussels appeared to differ from *M. edulis* (and the reference *M. galloprovincialis*) in having more flexible and elongated shells with more distinctive growth ridges as well as in the colour of the inside, which was not whitish but a duller grey-blue. However, this promising approach to the identification of species based on the shell “fragility” has remained underdeveloped. By comparing the photos of mussel shells in Beaumont et al. (2008) with the photos of shells from our Barents Sea samples (ESM Fig. ?B), one can see that the differences between these two species in Scotland are more striking that in the Barents Sea.

**Uses and abuses of single marker taxonomic tests**

The rule of thumb of the traditional approach to species identification is to use those organism’s traits, usually included into morphologic species diagnosis, which are diagnostic (fixed) – present in all individuals of one species but never occur in the other species. In the terms of probability theory it means that the probability of an individual possessing the species-specific diagnostic marker being the representative of a species is equal to one: P(species|trait) = 1. There are two evident sources of this probability decreasing – qualification of a researcher and condition of an individual (e.g. defective specimens), and one less obvious: ambiguity in diagnosticity of a trait. Determining whether diagnostic characters are truly fixed is generally impossible with finite sample sizes (Wiens, Servedio 2000). Hence, in practice, for diagnostic markers P(species|trait) ≤ 1.

Traditional species identification relies on diagnostic (fixed) traits of the organism. These traits, usually included in the morphological diagnosis, are present in all individuals of the species and are never found in other species. In the terms of the probability theory, it means that the probability of an individual with a species-specific diagnostic marker being a representative of the species in question is equal to one: P(species|trait) = 1. However, this probability may decrease for two reasons. An obvious reason is associated with deficient skills of the researcher or defective condition of the specimen. A less obvious source is an ambiguity in the diagnosticity of a trait. It is is generally impossible to determine whether diagnostic characters are indeed fixed if the sample size is finite (Wiens, Servedio 2000). Hence, in practice, for diagnostic markers P(species|trait) ≤ 1.

For semi-diagnostic traits, which are only known for blue mussels (McDonald et al. 1991), P(species|trait) < 1. Using semi-diagnostic markers is subjected to restrictions. With such markers we do not identify a species of a given individual, but assess the probability of assignment of the individual to one or another species. Similarly, dealing with population assessment we assess the probabilities to find representatives of one or another species in a sample but not the true proportion. The most critical point is that P(species|trait) is not constant but varies, yet in predictable manner with the prevalence of a species in a range [0;1]. As a rule, clinical diagnostic tests employ semi-diagnostic markers (REF), that’s why we followed the procedures from clinical based medicine to evaluate the performance of our morphotype test and to outline recommendations of its application.

In case of semi-diagnostic traits, the researchers do not identify the species of a given individual but assess the probability of its assignment to one or another species. For these traits, which are only known for blue mussels (McDonald et al. 1991), P(species|trait) < 1. Similarly, dealing with population assessment we assess the probabilities of finding the representatives of one or another species in a sample but not the true proportion. The most critical point is that P(species|trait) is not constant but varies, yet in predictable manner, with the prevalence of a species in a range [0;1]. Semi-diagnostic markers are employed in clinical diagnostic tests (REF), which is why we decided to profit from the experience of clinical medicine in order to evaluate the performance of the morphotype test and to outline recommendations for its application.

Correct application of tests based on semi-diagnostic markers, like clinical diagnostic tests, ultimately requires a “gold standard” – reference used for verification of the index test results (Banoo et al. 2006). In our mussel exercise, as references, groups of multilocus genotypes (from 4 to 171 645 loci depending on geographical sample set) defined by the dominance of alleles characteristic to one or the other species, were used. These groups did not represent true species but included hybrids some of which (e.g. first- and second generation hybrids) were assigned into groups randomly. It is worth mentioning that multilocus genotyping is barely employed for identification of cryptic mussel species. Most studies employ singular or few “standard” diagnostic PCR-based markers, most often nuclear Me15/16 and ITS and mitochondrial COI or 16S markers (Larrian et al. 2019). By offering the morphotype test to blue mussel researchers as a rough but cost-efficient alternative to genotyping we have to assess its reliability relative to single- and few locus tests also. A long known fact is that the efficiency of “diagnostic” markers to discriminate between *M. edulis* and *M. trossulus* differs between Western Atlantic (i.e. the Gulf of Maine) and the Baltic contact zones. While species are nearly fixed for alternative alleles at Me15/16, ITS and mitochondrial markers in the Western Atlantic, in the Baltic Sea intraspecific differences are 20%, 32% and 0% at these loci respectively, due to mass introgression of *M. edulis* genes into local *M. trossulus* genome (Riginos,Cunningham 2005); compare with 44% and 6% differences in morphotype frequencies between species in the Gulf of Maine and the Baltic. As far as we know, the efficiency of taxonomic tests based on singular or few “standard” loci have not been carefully evaluated for other *M. edulis* – *M. trossulus* contact zones (but see Vainola, Strelkov 2011 and Wilson et al. 2018 for some attempts). The recent population genomic studies of hybridizing *Mytlus* species come to the conclusion that species could lack fixed genetic differences at all due to ubiquitous introgression and that loci could introgress in unpredictable manner in different contact zones (Fraïsse et al. 2016; Simon et al. 2019); on that grounds the conventional approach of mussel species identification by singular molecular markers have been criticized (Simon et al. 2019). We do not pretend the morphotype test would be better than most single-locus taxonomic tests in any contact zone but stress that unlike morphotype test the performance of these tests still to be evaluated for most contact zones between *M. edulis* and *M. trossulus*.

A correct application of tests based on semi-diagnostic markers, such as clinical diagnostic tests, ultimately requires a “gold standard” – the reference used for verification of the index test results (Banoo et al. 2006). In our case study of the blue mussels, we used as references the groups of multilocus genotypes (from 4 to 171 645 loci depending on the geographical sample set) defined by the dominance of alleles characteristic of one or the other species. These groups did not represent true species. They included hybrids, some of which (e.g. first- and second generation hybrids) were assigned into groups randomly. To note, multilocus genotyping is seldom employed for identification of cryptic mussel species. Most studies rely on singular or few “standard” diagnostic PCR-based markers, usually nuclear Me15/16 and ITS and mitochondrial COI or 16S markers (Larrian et al. 2019). Offering the morphotype test as a rough but cost-efficient alternative to genotyping, we have to assess its reliability as compared to single- and few locus tests. It has been long known that the efficiency of “diagnostic” markers for discrimination between *M. edulis* and *M. trossulus* is different in Western Atlantic (the Gulf of Maine) and the Baltic contact zones. In Western Atlantic the species are nearly fixed for alternative alleles at Me15/16, ITS and mitochondrial markers, while in the Baltic Sea intraspecific differences at these loci are 20%, 32% and 0%, respectively, due to a mass introgression of *M. edulis* genes into the local *M. trossulus* genome (Riginos, Cunningham 2005). For comparison, the differences in morphotype frequencies between species in the Gulf of Maine and the Baltic Sea are 44% and 6%. As far as we know, the efficiency of taxonomic tests based on singular or few “standard” loci has not been carefully evaluated for other *M. edulis* – *M. trossulus* contact zones, though some attempts have been made (see Vainola, Strelkov 2011 and Wilson et al. 2018). The recent population genomic studies of hybridizing *Mytilus* species indicate that these species can altogether lack fixed genetic differences due to ubiquitous introgression and that loci can introgress in unpredictable manner in different contact zones (Fraïsse et al. 2016; Simon et al. 2019). On these grounds, the conventional approach of mussel species identification by singular molecular markers has been criticized (Simon et al. 2019). We do not claim that the morphotype test would fare better than most single-locus taxonomic tests in any contact zone between *M. edulis* and *M. trossulus*. At the same time, we would like to point out that the performance of these tests has not been evaluated in most contact zones, while that of the morphotype test has been.

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 morphological characters and identified by individual or complex (like coordinates of multifactorial analysis) semi-diagnostic traits. These are subspecies when defined according to the 75% rule (“in order to qualify as a subspecies, 75% 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 secondary lost fixed differences due to introgressive hybridization (Fitzpatrick et al. 2015 – not the best reference). 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.

A situation when one has 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, it is not unique. There are other taxa, which lack fixed diagnostic morphological characters and are identified by semi-diagnostic traits, individual or complex such as like the coordinates of multifactorial analysis. These taxa are subspecies defined according to the 75% rule (“in order to qualify as a subspecies, 75% 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 secondarily lost fixed differences due to introgressive hybridization (Fitzpatrick et al. 2015 – not the best reference). Therefore, we hope that our experience of dealing with a non-fixed taxonomic character would be interesting not only to our colleagues working with blue mussels but also to the researchers who study sympatric taxa with vague morphologies and semi-isolated gene pools.

**Acknowledgments**