Dear Dr. Melzner,

Thank you for the evaluation of our manuscript. Our responses to the reviewers are given below.

We have provided the missing initial data. We apologise for the omission. The phrase “data not shown” has been removed. The map in Fig. 1 was produced using Inkscape 0:92, and we have added this information to the figure legend.

Reviewer #1: This is an interesting manuscript describing a new semi-diagnostic character that could be used to roughly distinguish between the two mussel species Mytilus edulis and Mytilus trossulus in the field without the need for genotyping. I want to commend the authors for collecting this broad sample set. The analyses are in my opinion well done, although I had a few questions and concerns, especially regarding the use of allozyme markers and the lack of available raw data, that should be addressed.

Thank you for a careful evaluation of our manuscript.  
  
- The manuscript is a bit long. Sometimes it is hard to follow what was done.

We have shortened Discussion. We hope the text is now easier to follow.

- L94: unambiguous

Corrected (L?? cleaned ms)

- L191: Mussels in the Baltic Sea represent a hybrid swarm and are not pure M. trossulus. This is explained later in the discussion, but I think this should be recognized already here.

We introduced the sentence: “To note, the Baltic M. trossulus is more strongly introgressed by M. edulis alleles than other populations.” (L?? cleaned ms)

All Atlantic populations of M. trossulus (as well as M. edulis) are introgressed by alien genes. The Baltic population is introgressed more strongly than any other, but its M. trossulus ancestry is evident nevertheless (see genomic data in <https://doi.org/10.1111/mec.13299> and <https://doi.org/10.1111/jeb.13709>). For the purpose of this study, the M. trossulus ancestry of the Baltic mussels is more important than its “swarm” status. This term has different meanings (a unimodal hybrid zone; a randomly mating hybrid population that is independent of further outside recruitment), and this may be confusing for the reader.

- L198: I did not find a table with the allozyme and SNP genotypes and also no accession numbers for Illumina sequence data for samples that were new to this study. Could you please provide and deposit the raw data for full reproducibility? I think the input STRUCTURE files should be provided as well.

We apologize for the omission. The initial data are now provided.

Data Availability: Initial allozyme data and data on individual shell size and morphotype are available from St. Petersburg State University database (http://hdl.handle.net/11701/22292). Gulf of Maine SNP genotypes are available from <https://datadryad.org/stash/share/N11WUV2APlT_KAg01Lm1857K_6in2M431xsfh5uOJnQ>

All the other relevant data are contained in the paper and its Supporting Information files.

- L203: Why were the allozyme markers used? These markers are also only partially diagnostic and provide less resolution than multilocus DNA markers. It does not make sense to me to try to confirm a semi-diagnostic characteristic with methods that are not fully diagnostic themselves. Then another part of the dataset is genotyped with a genome-wide SNP panel. I think the authors need to give a better explanation / justification of their analyses and a reason why they did not choose a consistent method across samples.

We inserted the following explanation about allozymes (L?? cleaned ms): “These four loci were involved in the initial differential diagnosis of *M. edulis* and *M. trossulus* and the description of all contact zones under consideration. They individually show 70–95% allele frequency differences between species, and, being less affected by introgression than most of conventionally used PCR-based “diagnostic” markers, are reliable markers for species identification everywhere.”

To use 100 000 DNA markers for such a simple classification task as distinguishing between M. edulis and M. trossulus (or genotypes dominated by their genes in case of unimodal hybrid zones) is not unlike using a sledge-hammer to crack a nut. However, as we had these data anyway, we saw nothing wrong in using them. Many genetic markers are desirable for a detailed analysis of hybrid zones but we do not need a more fractional classification of genotypes since the test is based on a binary marker and our target contact zone is strongly bimodal.

Our opinion about the reliable diagnostic markers for hybridizing species of blue mussels is expressed in the Discussion (L?? cleaned ms). Our assignment is based not on individual markers but on their complex analysis by Structure method.

- L269: What does a few genotyped samples mean? Please specify the sample size for this analysis.

The sentence is modified: “Finally, we tested how well *Ptros*, *P(edu|E)* and *P(tros|T)* could be predicted using formulas Eq 1-3 and the data on the morphotype proportions among species (*P(T|tros)*, *P(T|edu)*) in a few (minimum two, see below) genotyped samples”. (L?? cleaned ms)

Would it not be interesting to see if a general model with pools of all samples is possible, using salinity and location as variables? I don’t completely understand why this was not tested. I would think that such an analysis would be useful for researchers that work on regions not included in this study.

Thank you for noticing that. Exactly such model was tested using extensive Kola data (see “Associations between morphotypes and genotypes around the Kola Peninsula”). The data on other geographical regions do not allow for such an analysis due to lack of samples and (or) salinity variation between sampling localities.

As for a general model with pools of all samples from all sets, using salinity and location as variables, it should be considered that there is pronounced geographical variation in morphotype frequencies in M. trossulus in addition to salinity related variation in Arctic populations of M. edulis. The predictive power of any model “ignoring” region (technically, treating region as a random factor regulating both intercept and slope of the model) would be unsatisfactorily low.

Recommendations for researchers who work on regions not included in this study are provided in the Discussion: they have to evaluate the test de-novo for each new contact zone. We developed “calculators” to aid such research.

- L309-323: This seems redundant. Can you merge this section with the one on the Kola Peninsula?

We would like to present the analysis for the Kola data and that for the geographical data separately because different models are employed.

- L601-603: I don’t think this statement is justified. Mytilus species have emerged as a non-model system for hybrid zone research, so for many scientists knowing the exact genotype composition is very important. I would soften this statement.

We suggest the following wording. “In some study areas, such as community ecology, biomonitoring and aquaculture, the knowledge about the taxonomic structure of blue mussel populations and a rough classification of individuals into “species” may be more valuable than the precise information about the genetic affinity (e.g. Structure q-value) of any given mussel.” (L?? cleaned ms)

- L672: links

Corrected. The paragraph is shortened. (L?? cleaned ms)

- L728: But probably there are some published data on salinity for those locations, even though it was not measured in this study?

Some data on salinity for those locations are indeed available (see S3 Table) but they are flawed in some way or other and cannot be used for the analysis. The St. Lawrence sample was from Baie-Sainte-Catherine, and literature data on salinity in this area are inconclusive (De Vernal, A., St-Onge, G. and Gilbert, D., 2011. Oceanography and Quaternary geology of the St. Lawrence Estuary and the Saguenay Fjord. In IOP Conference Series: Earth and Environmental Science (Vol. 14, No. 1, p. 012004). IOP Publishing). The other subarctic sample was collected beside Hvalsey Church in Davis Strait, not far from the glacier. Literature data on salinity conditions in this area are contradictory: saline according to Ribergaard, M.H., 2007. Oceanographic investigations off west Greenland. Danish Metrological Institute Centre for Ocean and Ice (DMI), Copenhagen but fresh according to <https://dce2.au.dk/Pub/arcticenvironment/reports/ArcticReport54.pdf>). This information is provided in S3 Table.

Reviewer #2: Species identification based on a semi-diagnostic marker: evaluation of a simple conchological test for distinguishing blue mussels Mytilus edulis L. and M. trossulus Gould  
  
PONE-D-20-30389  
  
Khaitov et al.  
  
The manuscript is concerned with the diagnosis of two species (Mytilus edulis, M. trossulus) that are difficult to distinguish based on morphological characters for several reason. For instance, there is a general lack of phenotypic characters in these closely related bivalves, there is large phenotypic plasticity and there is extensive interspecific gene flow. The current taxonomic status of both species is derived from multi-trait investigations of mussel shells and genetic analyses (basically allozyme investigations). Additional support comes from the fact that both species are show ecological differentiation. Despite of ecological and phenotypic differentiation, they are considered part of the Mytilus edulis species complex together with M. galloprovincialis. Given the great importance of mytilid mussels for marine science (they represent up to 90% of the benthic biomass and hence shape marine ecosystems), an easy-to-apply diagnostic marker would be more than desirable for research. This is the justification of the study presented by Khaitov et al.! The authors follow previous investigations in the White Sea and test the “dark stripe” as a semi-diagnostic shell morphological character. There aim is to investigate, whether this trait is diagnostic in all marine environments. The authors apply a mathematical approach that is common on medicine where semi-diagnostic characters are commonly used for diagnosis. Both, the scientific question and the innovative approach makes the study interesting for (marine) taxonomists and ecologists and will finally guaranty the publication. However, I have serious concerns that the manuscript in its current form is already publishable! I have various questions and recommendations to the authors that need to be addressed before I can recommend the publication of this manuscript.

Thank you very much for a careful evaluation of our manuscript.

Species concept  
My first question concerns the species concept behind the study. The authors use either allozymes or DNA markers for a priori species identification. To my understanding, this is used to calculate the Ptros-parameter. Although many of the populations are situated in hybrid zones, the individuals are either classified as M. edulis or as M. trossulus. For instance, STRUCTURE-q values (which represent hybrid indices) are used for a priori species diagnosis and the threshold is 0.5 (lines 214 to 215). This means, the authors classify F1 hybrid like individuals and early-backcross-generations as one or the other species. Furthermore, species diagnostic allozymes may mask the extend of genetic admixture at other genomic loci, e.g. in the Baltic. Consequently, classification into two categories that incorporate ANY hybrid status seems not justified. I had expected classification base on other admixture thresholds into at least three categories (including at least one hybrid category) and with reference to a valid species concept (e.g., referring to genetic characters). To avoid any misunderstanding: I agree that small proportions of introgression do justify allocation to one or the other species; I disagree that F1 like genotypes (including backcross genotypes) are not treated as hybrids but as one or the other species. In this context, what is said the paragraph about the Baltic and Norway (lines 669-687) could be discussed in another light if the hybrid category is considered. This paragraph already suggests that the pronounced hybrid character of these populations make the approach used in this manuscript doubtful.

Following your advice, we introduced an analysis of the hybrid zones and the associations between morphotypes and q-values (“Genotypic structure of samples” section of Results). To note, much of these data has already been published, but the inclusion of new samples and visualization methods probably indeed justifies a reanalysis. We also included additional reminders about the method of classification into M. edulis or M. trossulus to Discussion (L?? cleaned ms).

The questions you raise are fundamental, and in our opinion there is no straightforward answer to them. We used a standard population genetic method to distinguish between hybridizing taxa. Structure identifies in the data distinct genetic entities in HW and gametic equilibria (called clusters) and estimates the probabilities of individuals belonging to these clusters (q-values, interpreted as contributions of each cluster genome to individual genotypes). Technically, species are these clusters. Their identity with M. edulis and M. trossulus was proven in the previous studies, which are cited in the manuscript. Classification of genotypes into ancestry classes (e.g. purebreds, different types of early generation hybrids) by q-values is tricky. Neither approach can be considered as universal.

M. edulis and M. trossulus demonstrate features of sympatric species rather than a “hybrid swarm” in our target contact zone in Russia as well as in America and Scotland (and Greenland, doi: 10.1007/s00300-015-1785-x.). This is a known fact. In our opinion, this makes M. edulis and M. trossulus a reliable model for the illustration of the statistical method for identification of cryptic species in sympatry by a semi-diagnostic binary marker. It is impractical to distinguish more than two categories when a singular binary marker is used.

The case of the Baltic mussel is different. If M. edulis and M. trossulus hybridize so easily in the Baltic Sea, should they be distinguished there at all? Or is it best to say that they all are hybrids, members of the unitary and indivisible “hybrid swarm”? Perhaps the truth lies in the middle. It is reasonable to classify mussels from the unimodal hybrid zone by taxonomic ancestry, e.g. by scores of q-values, but it is improbable that any class would include only purebreds or only particular categories of early-generation hybrids. We classified them into M. edulis- and M. trossulus-like (which seems reasonable) and found that the differences in morphotype frequencies were quite small. This is why we do not recommend our method for individual taxonomic assignment of mussels in the Baltic contact zone and in Western Norway. But at least we gave it a try. In our opinion, there is nothing wrong with the fact that we can identify the proportion of species (their genes) in populations and select exclusively M. trossulus-like mussels by morphotypes in the Baltic.

The “dark stripe character” in obvious hybrid individuals  
[I am aware that “dark stripe” does not fully describe the complexity of the prismatic layer closed to the ligament but I would like to use the term for simplicity.] Including a hybrid category would allow testing how the dark-stripe-character is developed in genetic hybrids! I am not sure whether the length of the “stripe” gives an information?

Please see the new “Genotypic structure of samples” section of Results. Hybrids are intermediate in morphotype frequencies.

The length of the strip was considered in other publication (DOI: 10.1086/697944). In quantitative terms, morphotypes could be expressed as the Z-index (Z = A/L, where A is the distance from the umbo to the anterior boundary of the prismatic layer under the ligament nympha and L is the distance from the umbo to the posterior boundary of the ligament). A continuous expression of a morphological trait allows a slightly better discrimination of M. edulis- and M. trossulus-like mussels than the discrete (E-morphotype vs. T-morphotype) expression in the White Sea. Measuring the length of the strip is time-consuming and cannot be recommended as an express method for species identification.

Curiously, an analysis of museum collections shows that there is a downward trend in the Z – index (among E-morphotypes) in the White Sea mussels with time (since 1970s). We do not know whether it is due to genetic (introgression) or environmental (e.g. acidification) factors. It would be interesting to analyze other historical collections of mussel shells representing time series for morphotypes and the Z-index, if any such collections are available.

Morphological characters  
Just a question: Is the presence/absence of the “dark stripe” the basis for PT (T-morphotype)

Yes.

Taxonomic structure of populations  
I like the idea that the taxonomic structure of a population might be more important for some scientific questions than the genetic affinity of a given mussel (lines 601-603). However, I do not see the reason why mussels in a mixed population should be identified as either M. trossulus or M. edulis (lines 612-615). Again, why is a status of being a hybrid not considered? Hybrids may have properties that allow surviving in extreme marine environments and Hybrids are the rule (not the exception) in mytilid mussel contact zones.

Please see the new “Genotypic structure of samples” section of Results and above

Length of the discussion  
The discussion takes 12.5 pages and is overly long to my opinion. I feel that it need to be shortened substantially. This particularly refers to the part from line 747 to 876 (= 6 pages). I recommend summarizing the applications and the limitations of the morphotype test in a concise way.

We shortened discussion by about 10%.

Terminology  
The parameters Ptros, PT, P(E/edu), P(T/tros), P(tros/T), P(edu/E) are parts of the calculations and the abbreviations are justified and logic. However, it is hard to keep the overview about their meaning. I ask the authors to give the short explanations (as in lines 229-236) as often as possible to help the readers.

We have additionally included explanations of the predictive values in the Results (L?? cleaned ms).