Not the salinity: factors regulating distribution of sympatric blue mussels Mytilus trossulus and M. edulis in the White Sea

**Materials and methods**

**Study area**

The 185 km long Kandalaksha Bay in the north of the White Sea is oriented from southeast to northwest and is funnel-shaped (Fig. 1, 2). Climate is continental subarctic with cold winters (4-5 months of ice cover) and relatively warm summers (August SST is 13.8°C) (REF). The bay contains hundreds of banks, skerries and islands. The bay's shoreline is fringed by numerous smaller inlets. Because of this shoreline geometry, different areas are exposed to wave action to varying degrees. The summer surface salinity is 24 ppm in most of the Bay (“normal” salinity of the White Sea is 24-30 ppm), but lower in the estuary areas. A total of 24 rivers and two canals with individual catchment area of 141 - 12830 km² (average 240 km2, <http://www.mnr.gov.ru/files/part/0306_perechen.rar>; see also ESM – кому интересен список рек в статье про мидий?) flow into the bay, creating local salinity gradients. The top of the bay is the most desalinated, due to the influence of the largest river Niva (REF). Historically (through most of the 20th century), six ports were functioning in the area operating high tonnage oceanic vessels (REF - 'Лоция Белого моря'; ; Изд-во: М.: Управление гидрографической службы ВМФ, 1964 г; Красавцев Л. Б. Внешняя торговля России через порты Белого моря в начале ХХ века //Вестник Северного (Арктического) федерального университета. Серия: Гуманитарные и социальные науки. – 2011. – №. 3 и так далее). Two of these ports, Kandalaksha and Vitino, both in the top of the Bay, are currently functioning while the remaining ports have not been used to serve ocean-going ships in recent decades, but are still visited by coaster ships, according to the conventional knowledge (Fig. 2).

Mussels are ubiquitous in the littoral of the bay, and also form sublittoral beds in some areas. On the littoral, mussels are most abundant within the fucoid belt formed by Fucus vesiculosus and Ascophyllum nodosum within depth range approximately 0.2-1.2 m above the chart datum (REF) (Fig. ? – предлагаю где-нибудь показать пояс фукоидов и мидий на фукусе и на грунте).

**Mussel sampling and processing.**

When sampling, we tried to describe littoral populations of mussels in the Kandalaksha Bay in as much detail as possible, also taking into account the heterogeneity of their habitats by salinity, wave exposure and proximity to ports. A total of 95 intertidal sites were surveyed during the summer months of ???-2020 (почему 2019? - в Умбу мы с тобой ездили в 2018, а в Гридино и Черную речку за язычками с гитой и генельтом в 2016… Fig. ?, ESM ?). At each site, six samples were taken no more than 20 meters apart, three from fucoid thalli (“Algae samples”) and three from the ground (i.e. stones, sand, “Bottom samples”). Algae samples were taken from individual bundles of fucoids large enough to contain at least a few dozens of mussels (However, this condition proved to be unfeasible for several samples from sparse populations ?). Bottom samples were taken using a 0.025 m2 square frame placed on a ground in such a way that at least a few dozens of mussels could be sampled. Shells of mussels with a shell length more than 10 mm were cleaned from tissues and used in further analyses. Shell morphotypes (E-morphotype, characteristic to M. edulis and T-morphotype, characteristic to M. trossulus) were identified as in Khaitov et al. 2021. In total we processed 55500 mussels. The mean number of mussels per sample was 97 (range 2-785), per site - ? (?-??).

**Environmental parameters assessment.**

For each site, seven? environmental parameters were assessed. The distance to the nearest port (*DistPort*) was measured as a straight line distance to the nearest of the five ports located in the area. Each port was assigned to one of two categories: "functioning" or "abandoned" (*PortSize*) depending on whether it had received oceanic ships in the past 20 years according to common knowledge. By the degree of wave exposure (Аббревиатура), each site was visually assigned to one of two categories: “exposed” or “sheltered”. As another measure of wave exposure we calculated the value of wind fletch (Аббревиатура) - the unobstructed length of water surface, km, over which wind can blow over (Seers, 2018). The averaged over all the four cardinal directions value of the wind fetch was calculated using distances taken from small-scale geographical map and the package “fetchR” (Seers, 2018). Instant salinity (*Salinity*) was measured at time of mussel sampling using refractometer (++++).As another measure of salinity (and correlated factors?), the distance to the mouth of the nearest river (*DistRiver*) was measured as the distance in a straight line to the mouth of the nearest of the 26 rivers and channels flowing into the bay. Each river or channel was assigned to one of two categories (*RiverSize*): “large” or “small” depending on whether its catchment area was more of less than an average catchment area of all rivers, 240 km2 . (This operational classification should not be misleading as to the size of rivers, the largest of which are classified as "medium" by hydrologists (Khublaryan M. G. (ed.). Types and Properties of Water-Volume I. – EOLSS Publications, 2009.))

**Calculations and statistics.**

For each sample and for polled samples from each site, the proportion of T-morphotypes (PT) was calculated. The relationship between the proportion of *M. trossulus* (Ptros) and PT is an almost straightforward (Khaitov et al., 2021), yet for formal reasons to predict Ptros basing on PT for each site we used formula (Khaitov et al., 2021):

У меня сложилось впечатление, что этот логический шаг не нужен в исследовании, и никак не используется; тогда очевидно и edulis с trossulus тоже не нужны. Или убеждаешь, что не верблюд, или не лезешь в калашный ряд. Не интересно читать про распределение язычков, которое как-то связано с распределением видов.

Нужно показать не только как Ptros связан с индивидуальными предикторами, но и как предикторы связаны друг с другом. Без демонстрации «иллюзорных» корреляций между Ptros и соленостью все это не иллюстративно.

All statistical processing was performed using the statistical programming language R 4.05 (R core Team, 2021). The analysis was conducted in two steps. The first step was to determine which predictors are statistically related to the Ptros. The second step was to estimate the contribution of each predictor to the overall variation of Ptros.

At the first step, we constructed a random intercept logistic generalized linear mixed model (GLMM, Bolker et al., 2009) with beta-binomial residuals distribution where site was considered as a random factor (Model 1). The fixed part of the model included predictors *Substrate* (categorical predictor with two levels “Bottom” vs “Algae”), *Salinity* (continuous predictor), *DistRiver* (continuous predictor), *RiverSize* (categorical predictor with two levels “Large” and “Small”), *DistPort* (continuous predictor), *PortStatus* (categorical predictor with two levels “Abandoned” vs “Active”), где exposure с флечем? For all categorical predictors, the first mentioned category was used as base level in the analyses. The function glmmadmb() from the package “glmmADMB” (Fournier et al., 2012; Skaug et al., 2016 ) was used to build the GLMM.

At the second step, the fraction of total variability explained by the fixed part of the model was estimated using marginal coefficient of determination R2 (Jagger et al., 2017; 2019; Nakagawa & Schielzeth, 2013). Marginal R2 can be decomposed into semi-partial coefficients of determination, *semi-part R2*, that describe the contribution of each individual predictor to the overall coefficient of determination using the “partR2” package (Stoffel et al., 2021). Unfortunately this package is adopted to use GLMM with binomial residual distribution only. That’s why, to estimate *semi-part R2*, we constructed an additional GLMM model which, in addition to all factors considered in previous model, included an additional random factor, “the observation level random effect” (OLRE) (Model 2). OLRE is a random factor which adsorbs overdispersion in GLMM (Harrison, 2015; Harrison, 2014). This model was constructed using glmer() function from the lme4 package (Bates et al., 2015).

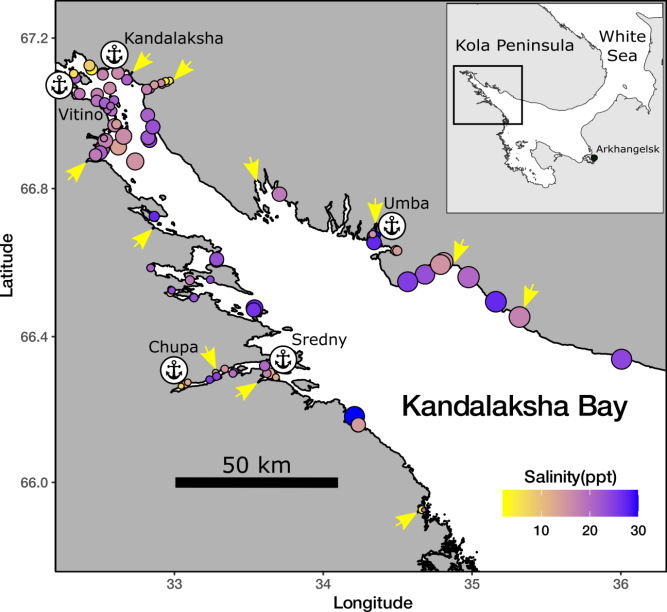
**Results**

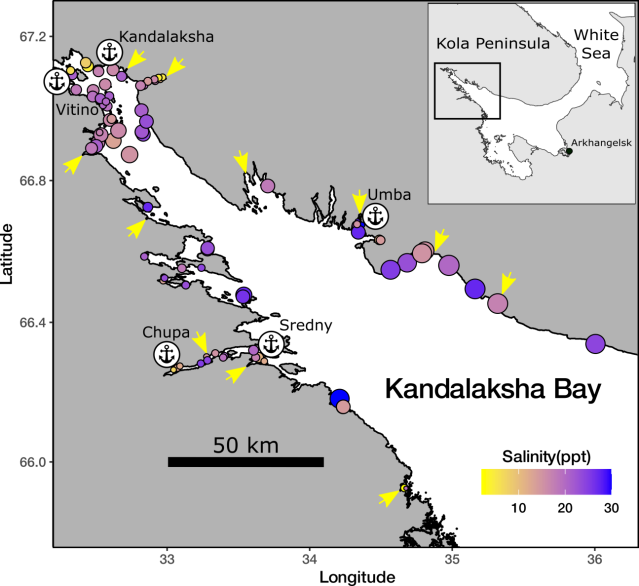
The distance from sampling sites to the nearest ports ranged from 0.11 to 82 km (Median = ???) (Fig ++, A). Sites visually categorized as "Sheltered" on average were characterized by lower values of wind fetch than those categorized as "Exposed" (ESM Fig. ++, B - ). The most exposed sites (maximum values of wind fetch) were located on the northwestern coast, in accordance with the prevailing southeastern summer wind direction (Lebedev, 2016), and on the open shores of the islands at the top of the bay (ESM Fig. ++1). Salinity measured at the time of sampling ranged 2-30 ppm (Median 19 ppm). The most desalinated areas were the top of the Bay and inlets where rivers flow into (ESM Fig. ++1). Salinity at sites close to “large” rivers was markedly lower than at sites closer to “small” ones (ESM Fig. ++, C). Salinity tended to increase as the distance to the nearest river increased (ESM Fig. ++ D). Читатель не ждет развернутого ответа на вопрос, действительно ли в КГЗ реки несут пресную воду, а ветер дует сильнее на открытых побережьях, чем на закрытых. Чем его может заинтересовать Fig. ++? В ESM его.

Читатель ждет, но так и не дождется ответа на поставленный в статье вопрос – соленость, расстояние до порта и ветер, - коррелируют, или нет? Надо ответить на вопрос и использовать ответ при обосновании выводов на основе результатов комплексного анализа. Просто картинки как на “Ptros in samples from different environmental conditions”, и тезис что они are expectedly not independent.

The values of PT at different sites ranged 0 - 0.85 (Median ???) which corresponds to the range of Ptros ?-? (Median ??). The sites with the lowest frequency of Ptros were located along the sparsely indented, wind-opened northwestern coast. The maximum values of Ptros were observed in the top of the bay (high desalination, low surf) and several inlets distant from each other, in particular, in the vicinities if the ports of Chupa and Umba (Fig. 2). On the background of these geographical mosaics, patchiness in distribution of morphotype frequencies at local scale was noticeable. For example, near Umba PT varied ?-?? at a scale of ? km.

Within sites, algal samples were generally characterized by higher *Ptros* than bottom samples (by an average of ? ? %; Fig.?; Fig. ? A). В чем смысл Fig ++. The frequency distributon of PT values at different substrates? – я бы еще понял, если бы попарно сравнивались выборки с разных субстратов в одних и тех же сайтах. Добавить, а лучше заменить. Я, кстати, не понял, эта связь с субстратом меняется или нет в зависимости от всего остального?





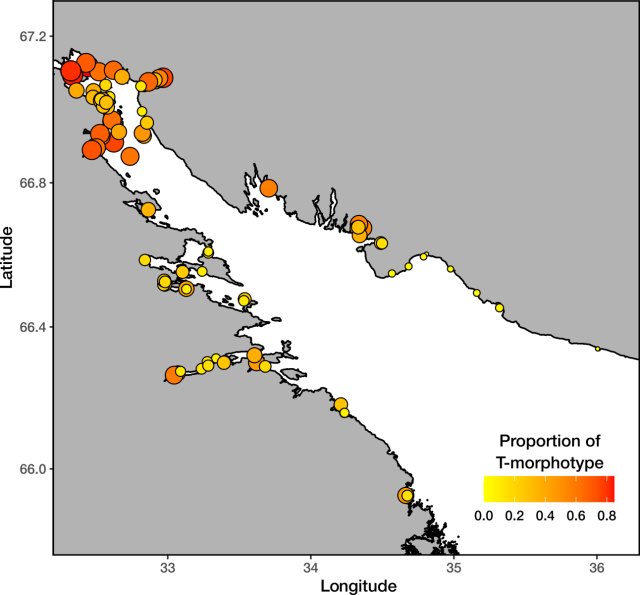
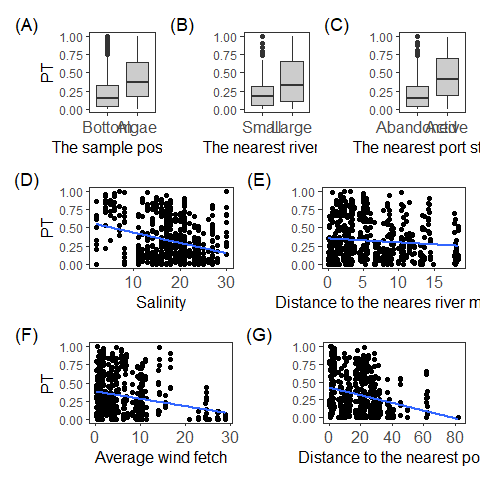


Fig. 1 Sampling sites, their environmental characteristics, and the predicted proportions of M. trossulus in mussel populations (Ptros) based on morphotype frequencies. The top of the Bay is given in separate insets on a larger scale. Sites are marked with dots; different fills denote salinity at the time of mussel sampling (A), mean wind fetch (B), and Ptros (C). Ports are labeled with anchor icons and named (for abandoned ports, the decade of closure is indicated). The mouths of the "large" rivers are indicated by arrows.

Читатели не знают, где Белое море, показанное на врезке, и им похуй на Белое море, – показать на карте Арктики или Европы, где этот залив. Предлагаю сделать отдельную иллюстрацию с такой картой и фотографиями поселений мидий.

Эти карты нужны для демонстрации паттернов распределения видов в связи с факторами среды, а не для демонстрации изменчивости местных условий в кандалакшских ебенях. Если изучаются виды, их и показывать. Соленость, прибойность и предсказанный Ptros на трех идентичных картах, где подробно изученная вершина залива показана в крупном масштабе. Стрелки – крупные реки? (в подписи не сказано, что это). Реки и порты повторять везде, для удобства ориентации, (в подписи не было сказано что такое якоря и подписи), различить на работающие и закрытые, декады закрытия предлагаю указать в названии. Не «Средний» (не было такого порта), а «Кереть 1970». Умба 1990. Потеряна Ковда 1980. Вадим, ты уверен, что Чупа когда-нибудь была портом, «обслуживающим крупнотоннажные суда»? С натяжкой в 1930, хотя она обслуживала СЛОН. Это надо выяснить! Соленость привычно давать от синего (высокая) до белого, а виды от синего (едулис) до красного (троссюлюс). Варьирование размера точек я бы убрал – это придумано для того, чтобы точки наезжали друг на друга ничего нельзя было разобрать. («Порты и якорные места. В вершине Кандалакшского залива расположен порт Кандалакша. В губах Кереть, Чупа, Ковда и Большая Пирья имеются портовые пункты с причалами. » 'Лоция Белого моря'; ; Изд-во: М.: Управление гидрографической службы ВМФ, 1964 г.; - почему бы ученым кандалакшского заповедника не познакомиться с лоциями разных лет и узнать, что где да когда.)



## Figure +. Ptros in samples from different environmental conditions. (A) In samples from bottom substrates vs. samples from algal substrates; (B) In sites closer to small rivers vs. sites closer to large rivers. (C) In sites closer to abandoned ports and closer to active ports. Ptros as a function of instant salinity (D), distance to the nearest river mouth (E), average wind strength (F), and distance to the nearest port (G). On D-G, the blue lines represent ordinal least-squares regressions. На графиках пиши хуйню свою «MinDistPort» - забыл уже, что как называется?

Among sites, *Ptros* was, on average, higher at sites located closer to large rivers than to small ones (Fig. ++ B) as well at sites located closer to active ports than to abandoned ones (Fig. ++ C). Further, *Ptros* tended to decrease as salinity, average wind fetch, and distance to the nearest port increased (Figs. ++ D, F, G, respectively). No association of *Ptros* with distance to the nearest river mouth was noticeable (Fig. ++, E). Здесь фраза о том, что эти зависимости коррелятивные и их надо проверить коплнксным анализом.

According to GLMM models (Table) *Ptros* was significantly related to the substrate type, average wind fetch, distance to the nearest port, the status of the nearest port but not to any surrogates of salinity. Model 2 was specially designed for assessment of predictors influencing power. Partial R2 for Model 1? – почему опять 1? was low (part R2 = 0.21, Table ++, Fig. ++) indicating that predictors included in the model in total explained about 20% of the variation of Ptros (так?). The semi-partial R2 indicates that the best predictor for *Ptros* was substrate type (semi-part R2 = 0.040), followed by the nearest port status (0.034), average wind fetch (0.023) and distance to the nearest port (0.013). The signal strengths from predictors characterizing salinity were in order of magnitude lower (Figure ?).

Table +. Parameters of the GLMM fitted. Здесь можно сказать как интерпретировать плюс и минус. Нельзя в два раза ужать – добавить циферки второй модели справа?

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| effect | group | term | estimate | std.error | statistic | p.value |
| Model 1 |  |  |  |  |  |  |
| fixed |  |  | -1.20 | 0.385 | -3.121 | 0.0018 |
| fixed |  |  | 1.07 | 0.070 | 15.224 | <0.0001 |
| fixed |  |  | -0.02 | 0.018 | -1.044 | 0.2965 |
| fixed |  |  | -0.03 | 0.023 | -1.473 | 0.1409 |
| fixed |  |  | 0.45 | 0.205 | 2.199 | 0.0279 |
| fixed |  |  | -0.05 | 0.016 | -3.035 | 0.0024 |
| fixed |  |  | -0.02 | 0.007 | -2.630 | 0.0085 |
| fixed |  |  | 1.18 | 0.232 | 5.100 | <0.0001 |
| random | Site |  | 0.83 |  |  |  |
| Model 2 |  |  |  |  |  |  |
| fixed |  |  | -1.40 | 0.456 | -3.072 | 0.0021 |
| fixed |  |  | 1.24 | 0.081 | 15.361 | <0.0001 |
| fixed |  |  | -0.02 | 0.021 | -0.768 | 0.4426 |
| fixed |  |  | -0.04 | 0.028 | -1.460 | 0.1442 |
| fixed |  |  | 0.45 | 0.243 | 1.865 | 0.0622 |
| fixed |  |  | -0.07 | 0.018 | -3.683 | 0.0002 |
| fixed |  |  | -0.02 | 0.008 | -2.528 | 0.0115 |
| fixed |  |  | 1.34 | 0.275 | 4.867 | <0.0001 |
| random | OLRE |  | 0.83 |  |  |  |
| random | Site |  | 0.99 |  |  |  |

Table +. Partial R2 for the full model and semi-partial R2 with bootstrap CI for predictors from the Model 1.

|  |  |  |  |
| --- | --- | --- | --- |
| Term | Estimate | Bootstap\_CI | Df |
|  | 0.211 | 0.164 : 0.271 | 8 |
|  | 0.040 | 0 : 0.109 | 7 |
|  | 0.000 | 0 : 0.072 | 7 |
|  | 0.003 | 0 : 0.075 | 7 |
|  | 0.005 | 0 : 0.077 | 7 |
|  | 0.023 | 0 : 0.093 | 7 |
|  | 0.013 | 0 : 0.084 | 7 |
|  | 0.034 | 0 : 0.104 | 7 |

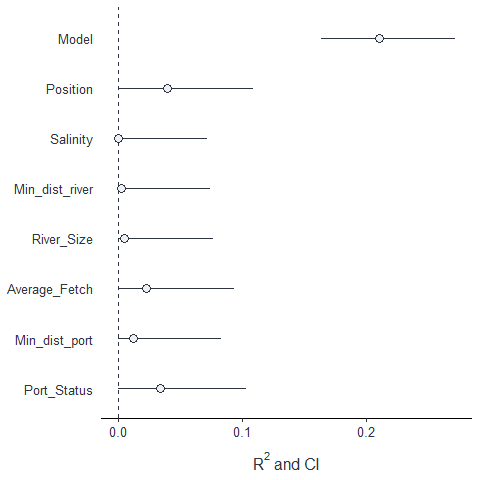


Figure ++. The importance of predictors as drivers of Ptros variation. Partial R2 reflecting the proportion of total variation explained – так что здесь точки и палочки? И нахуй таблица с тем же самым – убрать в ESM?