



Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.journals.elsevier.com/oceanologia/



ORIGINAL RESEARCH ARTICLE

Random forest assessment of correlation between environmental factors and genetic differentiation of populations: Case of marine mussels *Mytilus*

Tomasz Kijewski ^a, Malgorzata Zbawicka ^a, Jakob Strand ^b, Hans Kautsky ^c,
Jonne Kotta ^d, Merli Rätsep ^d, Roman Wenne ^{a,*}

^a Institute of Oceanology, Polish Academy of Sciences, Sopot, Poland

^b Arctic Research Centre, Department of Bioscience, Aarhus University, Aarhus, Denmark

^c Department of Ecology, Environment and Plant Sciences, Stockholm University, Sweden

^d Estonian Marine Institute, University of Tartu, Tallinn, Estonia

Received 29 April 2018; accepted 15 August 2018

Available online 30 August 2018

KEYWORDS

Marine environment;
Spatial distribution;
Seascape genetics;
Nuclear DNA markers
EFbis, Glu-5', ITS, M7
and Single Nucleotide
Polymorphism;
Baltic Sea

Summary The novel machine learning technique Random Forest (RF) was used to test if the genetic differentiation of populations of marine species may be related to any of the key environmental variables known to shape species distributions. The study was performed in North and Baltic Sea characterized by strong gradients of environmental factors and almost continuous distributions of *Mytilus* mussel populations. Assessment of the species identity was performed using four nuclear DNA markers, and previously published single nucleotide polymorphism (SNP) data. A general pattern of cline variation was observed with increasing *Mytilus trossulus* share towards the eastern Baltic Sea. Average allele share rose to 61% in Höga Kusten, Gulf of Bothnia. All Baltic Sea samples revealed a strong introgression of *Mytilus edulis* and a limited introgression of *M. trossulus* through the Danish Straits.

The studied environmental variables described 67 and 68% of the variability in the allele frequencies of *M. edulis* and *M. trossulus*. Salinity defined over 50% of the variability in the gene frequencies of the studied *Mytilus* spp. populations. Changes along this environmental gradient were not gradual but instead a significant shift from gene dominance was found at a salinity of 12 PSU. Water temperature and the trophic status of the sea area had only moderate association with the gene frequencies. The

* Corresponding author at: Institute of Oceanology, Polish Academy of Sciences, Powstańców Warszawy 55, 81-712 Sopot, Poland. Tel.: +48 58 7311763; fax: +48 58 5512130.

E-mail address: rwenne@iopan.gda.pl (R. Wenne).

Peer review under the responsibility of Institute of Oceanology of the Polish Academy of Sciences.



Production and hosting by Elsevier

<https://doi.org/10.1016/j.oceano.2018.08.002>

0078-3234/© 2018 Institute of Oceanology of the Polish Academy of Sciences. Production and hosting by Elsevier Sp. z o.o. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

obtained results showed that the novel machine learning technique can be successfully used for finding correlations between genetic differentiation of populations and environmental variables and for defining the functional form of these linkages.

© 2018 Institute of Oceanology of the Polish Academy of Sciences. Production and hosting by Elsevier Sp. z o.o. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Global climate changes and human economic activities, such as aquaculture and maritime transport, influence geographic distribution of marine species (Bach et al., 2018; Bosch et al., 2018; Gardner et al., 2016; Wenne, 2018). Extending knowledge on processes shaping species distribution will enable elaboration of adequate models for future forecasting. Changes to biodiversity and species distribution can be caused by biological or environmental factors, e.g. food availability, alterations in temperature, salinity, oxygen content in water, ocean circulation or depth (Banks et al., 2010; Silliman et al., 2011; Watson et al., 2011). Population genetic methods are tools for quantification of the genetic component of biodiversity. Studies of correlations between genetic diversity and spatial differentiation of factors along a broad range of environmental gradients can extend our understanding of species distribution changes (Riddle et al., 2008). This can be achieved through analyses of the linkage between these environmental variables and frequencies of genes in populations.

In comparison with other parts of the Atlantic Ocean, the Baltic Sea is a semi-enclosed reservoir characterized by marginal environments (Johannesson and André, 2006; Pocz-wierz-Kotus et al., 2014). Its salinity decreases in the North-East direction and strongly influences distribution and diversity of pelagic and benthic assemblages (Grabowska et al., 2015; Kijewska et al., 2009, 2016; Malachowicz et al., 2015; Vuorinen et al., 2015). Fauna of the Baltic is a mixture of fresh-water and marine species including *Mytilus* (Wennerström et al., 2013). *Mytilus* mussels have widely been used as an indicator of environmental quality and recently for monitoring of global climate changes (Dowd and Somero, 2013; Lesser, 2016; Li et al., 2015; Michaelidis et al., 2014). Morphologically similar species of *Mytilus* can hybridize in areas where their populations merge (Larraín et al., 2018; Wenne et al., 2016; Zbawicka et al., 2018). This impeded an identification of *Mytilus* species, their hybrids and backcross progenies, using classic methodology and justified using molecular markers. In the Baltic, *Mytilus* mussels have been used to study the effects of water acidification (Thomsen and Melzner, 2010), genotoxicity induced by oil spills (Barsiene et al., 2012), accumulation, the effects of pharmaceuticals (Ericson et al., 2010), and the accumulation of a variety of other chemical compounds and metals (Albalat et al., 2002; Bjork and Gilek, 1997; Dabrowska et al., 2017; Hoher et al., 2012; Kopecka et al., 2006; Larsson et al., 2018; Nyberg et al., 2015; Pempkowiak et al., 2006; Piwoni-Piórewicz et al., 2017; Potrykus et al., 2003; Protasowicki et al., 2008; Szefer et al., 2002; Szefer and Szefer, 1990; Szefer and Wenne, 1987; Szymczak-Zyla et al., 2006). *Mytilus* spp.

have several ecologically important features including water filtration and providing a link between the benthos and the pelagic zones by cycling nutrients and organic matter (Kautsky and Evans, 1987). In the Baltic, mussel farms have been set up for the purpose of mitigating eutrophication processes (Ozoliņa, 2017; Petersen et al., 2014). Commercial mussel farming and harvesting for human consumption operate in Limfjörd, Denmark, Sweden and in Kiel Bight, Germany (Bergström et al., 2017; Larsen and Riisgård, 2016; Utermann et al., 2018). Therefore, knowledge of the genetic composition of local populations of *Mytilus* can be useful not only for monitoring but also for the mariculture industry.

Mytilus populations in the Baltic have been studied with allozymes, nuclear DNA markers and sequences of mitochondrial (mt) DNA (Larsson et al., 2017; Rawson and Hilbish, 1998; Väinölä and Hvilsum, 1991; Zbawicka et al., 2007, 2014b). MtDNA of *Mytilus* is characterized by doubly uniparental inheritance, recombination and length polymorphism in the Baltic populations, which makes surveys of population differentiation more complex (Filipowicz et al., 2008; Zbawicka et al., 2003a,b). In the case of some nuclear DNA markers, polymorphism identified with, for example, e.g. PCR Amplified Fragment Length Polymorphism depends on reaction conditions and is not always reproducible (Cuéllar-Pinzón et al., 2016). New markers based on single nucleotide polymorphisms have been discovered by Zbawicka et al. (2012, 2014a). The reported studies highlighted the hybrid status of the Baltic *Mytilus* spp. populations with differential introgression power of selected genetic features. The Baltic populations of *Mytilus* spp. were established after the glaci-ation period during the Holocene and have retained unique characteristics compared to populations from other geographic areas (Kijewski et al., 2006; Smietanka et al., 2013; Väinölä and Strelkov, 2011; Wennerström et al., 2013). A steep salinity gradient was expected to act as a barrier between the North and Baltic Seas and result in an elevated mortality of larvae passing this barrier by means of currents (Gilg and Hilbish, 2003). Nonetheless, the influence of water circulation has been found to shape the genetic composition of the *Mytilus* spp. populations within the Baltic Sea and between the North and Baltic Seas (Larsson et al., 2017). Besides oceanographic connectivity and larval drift, local environmental conditions were expected to account for local adaptation and population differentiation (Valladares et al., 2014). A large array of environmental variables such as water temperature, flow velocity, food availability, either separately or interactively were considered as factors shaping the distribution and potential differentiation patterns of *Mytilus* spp. populations (Kotta et al., 2015). Salinity has been reported as an environmental factor significantly influ-

encing *Mytilus* shell variation across European and Arctic coasts (Telesca et al., 2018).

Commonly used statistical modelling may not be the most successful way to understand relationships between environmental variables and population genetics, as it starts by assuming an appropriate data model, and model parameters are then estimated from the data. Due to the lack of a solid understanding of how the external environment shapes the genetic composition of the *Mytilus* spp. populations being modelled, the predictive performance of these models would be expected to be moderate. Machine learning could provide a novel theoretical framework where modelling is seen as a sophisticated tool to improve our understanding of the relationship between environment and biota. In contrast, machine learning avoids starting with a data model, and rather uses an algorithm to learn the relationship between the response and its predictors (Hastie et al., 2009).

As with machine learning, the Random Forest (RF) method copes with different non-linear relationships, which are common in ecological data but difficult to analyse using more classic methods. The RF generates a large number of regression trees, each calibrated on a bootstrap sample of the original data (Breiman, 2001). Each node is split using a subset of randomly selected predictors and the tree is grown to the largest possible extent without pre-running. For predicting the value of a new data point, the data are run through each of the trees in the forest and each tree provides a value. The model prediction is then calculated as the average value over the predictions of all the trees in the forest. Recently, the RF has been applied to assignment analysis from population genetic data (Sylvester et al., 2018). There are other effective machine learning techniques such as Boosted Regression Trees and Maximum Entropy (Elith et al., 2006, 2008). All these methods belong to the same family of statistical models and therefore yield more or less similar results. The RF modelling is very efficient with a small sample size and this property makes this technique highly suitable for the current study.

The aim of this paper was to use random forest (RF) as the modelling algorithm to assess the relative contribution of natural environmental drivers in the allele frequencies in populations of marine species. North and Baltic Sea transects were used as an example because of known strong environmental gradients and almost continuous geographic distributions of *Mytilus* populations. Assessment of the species identity was made using four nuclear DNA markers in conjunction with already published single nucleotide polymorphism (SNP) data specific to the *Mytilus* species.

2. Material and methods

2.1. Environmental variables

The environmental layers used in the modelling (see below the section of statistical analyses) were produced from simulations of a coupled physical-biogeochemical model of the Baltic Sea (Meier et al., 2012). The model produced physical, chemical and biological data layers: annual average of salinity and current velocity, summer average of temperature and chlorophyll-*a*, winter average of nitrate (NO₃) and phosphate (PO₄). Annual and seasonal means were calculated

for the periods 1978–2007. In addition to the RCO-SCOB data layers, depth and wave exposure (simplified wave model) data were also used as modelling input variables. Depth data were acquired from the Baltic Sea Bathymetry Database (Baltic Sea Hydrographic Commission 2013). The simplified wave model is a surface wave exposure model that incorporates shoreline topography, fetch and wind data together with empirically derived algorithms to mimic diffraction (Isæus, 2004).

2.2. Sampling

Mytilus spp. samples consisting of 1466 individuals in total were collected from 32 sites in a transition from the south-eastern North Sea to the northeastern Baltic Sea mostly in 2008–2009 (Fig. 1 and Table 1). Three samples were collected in 2007, 2 in 2010 and 1 in 2011. The studied individuals varied in size from 1.5 to 8 cm. Prior to DNA extraction, mussels were frozen at –70°C or preserved in 96% ethanol.

2.3. Molecular methods

For DNA extraction, pieces of gill tissue ~3 mm × 3 mm were dissected and processed with the CTAB method of Hoarau et al. (2002). After precipitation in isopropyl alcohol, total DNA was suspended in deionized water. Four nuclear DNA markers, which tentatively diagnostically differentiate the *Mytilus* taxa in their North Atlantic and North Pacific range were analysed. A segment of gene coding an adhesive protein of byssus (Glu-5') has been characterized by diagnostic PCR products length differences among three studied taxa: *M. edulis*, *M. trossulus* and *M. galloprovincialis* (Inoue et al., 1995; Kijewski et al., 2009). Diagnostic differences between *M. trossulus* and the other two taxa in the internal transcribed spacer (ITS) regions between the 18S and 28S rDNA genes were detected with the restriction enzyme *HhaI* digestion (Heath et al., 1995; Kijewski et al., 2009). The EFbis marker is an intron in the elongation factor 1 α (Bierne et al., 2003) with double digestion RFLP variation diagnostic between *M. trossulus*, *M. edulis* and *M. galloprovincialis* (Kijewski et al., 2009, 2006). A taxonomic discrepancy in a sequence of acrosomal sperm protein (Riginos et al., 2006) was exposed using a set of 5 primers in PCR reaction as marker M7 (Kijewski et al., 2009). Amplicons and restriction fragments (*HhaI* digestion of ITS, *HhaI* + *RsaI* digestion of EFbis) were separated by electrophoresis in 3% NuSieve GTG agarose gels in TBE buffer and visualized with ethidium bromide.

2.4. Statistical analyses

Nuclear markers were scored for each individual and taxon-specific allele composition was described for each sample. Deviations from the random association of alleles within loci (Hardy–Weinberg equilibrium, HWE) were computed with the use of Arlequin 3.5.2.2 (Excoffier et al., 2005). The statistical significance of disequilibria was tested with the Markov chain. Genetic structure was analysed primarily by Structure 2.3 software (Pritchard et al., 2000), where Bayesian algorithm estimates the most plausible number of genetic clusters by 4 runs for *K* range from 1 to 10 with a length of the burn-in period of 500 000 and 15M MCMC

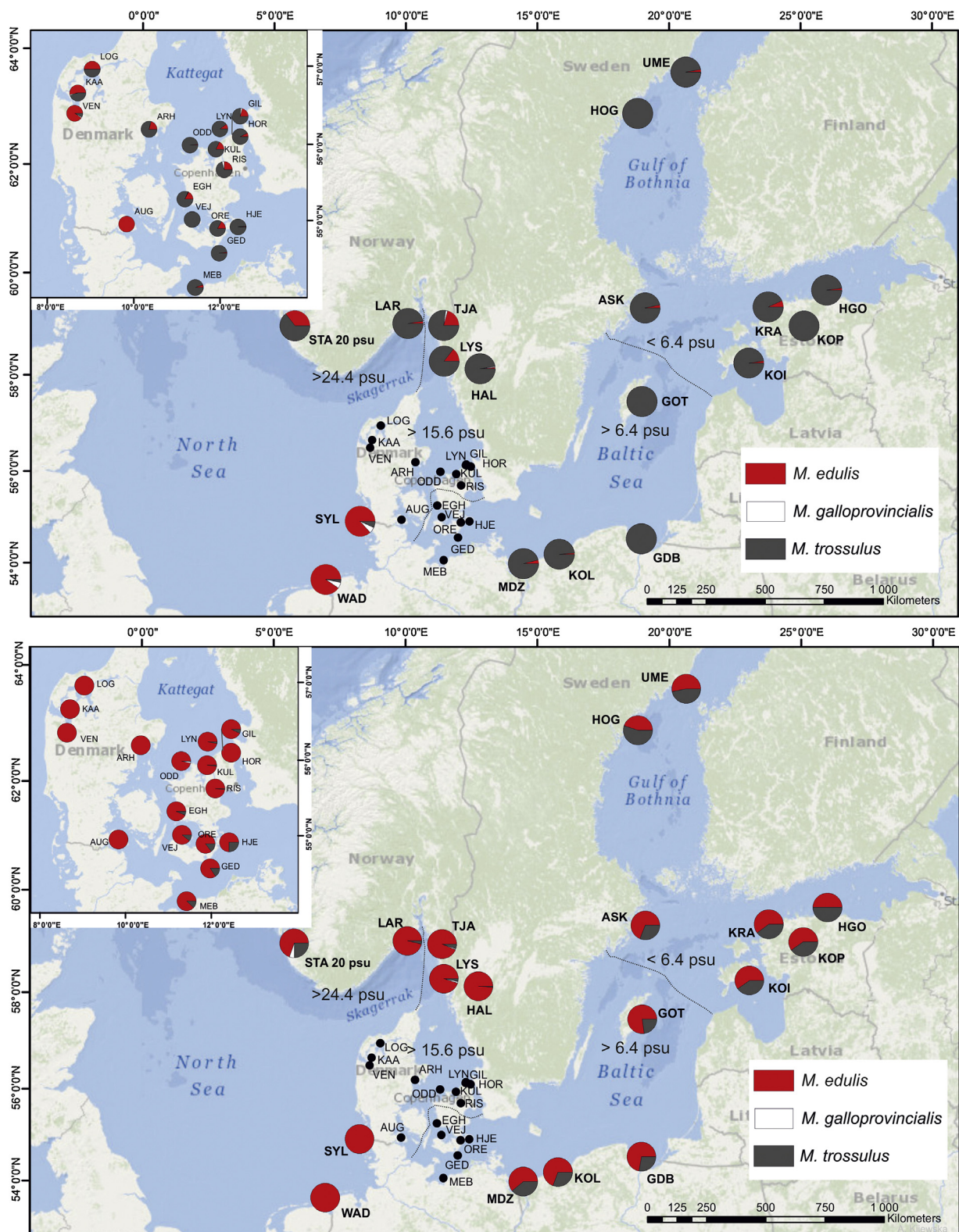


Figure 1 Map showing locations of sampling sites and genetic composition of the studied populations of *Mytilus* mussels. The pie charts visualise the frequency of taxon-specific alleles identified with 2 nuclear DNA markers: EFbis (a) and Glu-5' (b) and 26 SNPs (c).

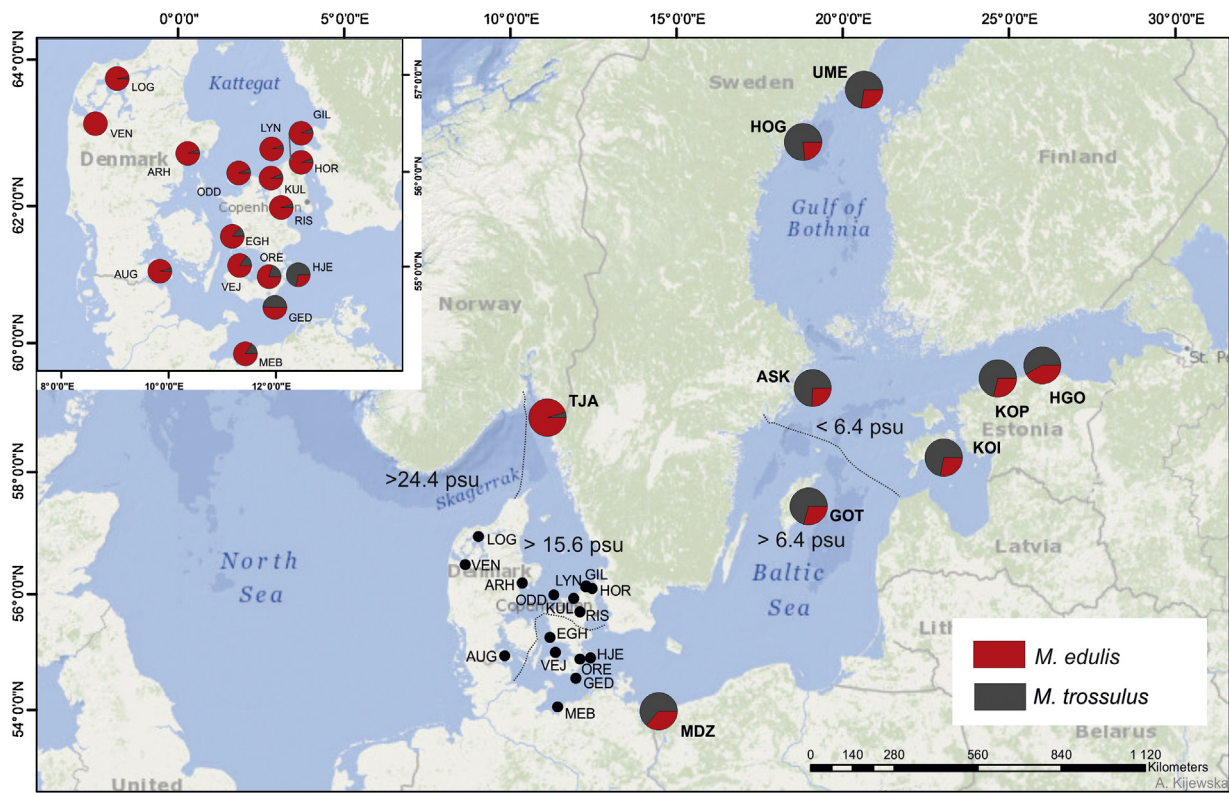


Figure 1 (Continued).

iterations after burn-in. The Evanno method highlights which number of clusters is the most probable by estimating the maximum value of ΔK (Evanno et al., 2005). The results of each run were compiled with the use of Structure Harvester v0.6.94 (Earl and vonHoldt, 2012). The different runs of the selected K were averaged in CLUMPP version 1.1.2 (Jakobsson and Rosenberg, 2007). Principal Coordinate Analysis (PCoA) was performed using the procedure implemented in GenAlEx 6.502 on the matrix of individual genotypes (Peakall and Smouse, 2012). The M7 reads were removed from this analysis due to lack of genotyping results for KOL, MDZ and MEB samples and the PCoA was conducted for three markers only. Two structure assessments obtained with Structure and PCoA were tested further by hierarchical ϕ - F -statistics using the AMOVA program Arlequin 3.5.2.2 (Excoffier et al., 2005) with fixation indices estimated overall (F_{ST}), between groups of populations (F_{CT}), within groups (F_{SC}) and F_{ST} pairwise. Significance was assessed by 9999 permutations of the original data matrix.

The package “randomForest” (Breiman and Cutler, 2015) was used to run RF models in the statistical software R 3.2.2 (The R Foundation for Statistical Computing 2015). Two main parameters can be set in RF models: the number of predictor variables to be randomly selected at each node (mtry) and the number of trees in a forest (ntree). The parameter mtry was set to the square root of the number of predictor variables as suggested by Liaw and Wiener (2002) for the classification model, and ntree was set to 1000, as 500 trees usually yield stable results. The importance of predictor variables in the model was assessed using the internal method of the package “randomForest” (mean decrease in

accuracy) using 10 permutations (Breiman and Cutler, 2015). Partial dependence plots (sensu Friedman, 2001) were produced using partialPlot function in “randomForest” to illustrate the dependence of model predictions on individual covariates. In general, a regression function will depend on many predictor variables. Partial dependence is the dependence of the probability of presence on one predictor variable after averaging out the effects of the other predictor variables in the model. The y-axis represents the modelled marginal effect on a response variable. Negative values (in the y-axis) mean that the positive class is less likely for that value of the independent variable (x-axis) according to the model. Positive values mean that the positive class is more likely for that value of the independent variable. Zero implies no average impact on class probability according to the model. In this study, allele frequencies of the *M. edulis*, *M. trossulus* and *M. galloprovincialis* specific nuclear markers were modelled along key environmental variables. Additionally, the frequency of twenty-six outlier SNP markers characteristic to *M. trossulus*, based on the already published results by Zbawicka et al. (2014a), were included in the analysis (Table S1). Eighty-four SNPs were genotyped in 630 specimens from the Baltic Sea, Belt Sea, Kattegat and North Sea using the Sequenom MassARRAY iPLEX genotyping platform. Sixty SNPs were polymorphic. Based on F_{ST} outlier analysis, 26 SNPs were significantly involved in the differentiation between the Baltic Sea and Danish Straits (Zbawicka et al., 2014a). Separate RF models were built for each *Mytilus* species.

Multicollinearity can be an issue with the RF modelling when answering if and when environmental variables are of

Table 1 Geographic localization of sampling sites. Allelic share of *Mytilus edulis*, *Mytilus galloprovincialis* and *Mytilus trossulus* for the studied genetic markers: Glu-5, Ef-bis, M7 and ITS. Samples marked with an asterisk were genotyped only with Glu-5, EF-bis and ITS markers by Kijewski et al. (2011). Frequencies of *M. trossulus* genes were used in RF model.

Sample		Coordinates		Salinity		Average allelic share over loci		
Name	Localization	N	E	ppt	n	<i>M. edulis</i>	<i>M. galloprovincialis</i>	<i>M. trossulus</i>
WAD	Wadden Sea	53°37'8.92"N	6°56'25.49"E	29.48	48	0.98	0.01	0.01
AUG	Augustenborg	54°57'2.00"N	9°50'0.00"E	17.44	39	0.97	0	0.03
VEN	Venø Bight	56°29'54.00"N	8°38'1.00"E	32.67	48	0.94	0.05	0.01
SYL	Sylt	54°54'55.14"N	8°14'43.63"E	31.23	15	0.92	0.06	0.02
KAA	Kaas Breeding	56°39'32.00"N	8°42'15.00"E	32.31	48	0.89	0	0.11
LOG	Løgstør Breeding	56°57'37.00"N	9°2'17.00"E	24.31	48	0.83	0	0.17
RIS	Riso	55°42'1.44"N	12°5'32.11"E	17.63	45	0.81	0.01	0.18
ARH	Århus	56°11'33.00"N	10°21'12.00"E	20.88	48	0.8	0	0.2
GIL	Gileleje	56°7'43.00"N	12°16'13.00"E	15.66	48	0.78	0.01	0.22
LYN	Lynetten	56°8'30.00"N	12°16'60.00"E	15.62	48	0.77	0	0.23
HOR	Hornbæk	56°5'54.90"N	12°27'49.09"E	15.79	47	0.76	0	0.24
TJA	Tjarnö	58°51'34.65"N	11°6'43.29"E	24.10	48	0.76	0.01	0.23
KUL	Kulhuse	55°56'5.00"N	11°54'10.00"E	17.27	48	0.75	0.02	0.22
ODD	Odden	55°59'17.00"N	11°18'2.00"E	15.90	47	0.75	0.01	0.24
HAL	Halse Fjord	58°7'1.48"N	11°49'46.34"E	22.82	40	0.75	0	0.25
LAR	Larvik Fjord	59°0'54.59"N	10°3'48.18"E	34.48	40	0.74	0	0.26
LYS	Lysekill	58°16'4.38"N	11°26'43.36"E	23.54	48	0.73	0.04	0.22
EGH	Egholm	55°15'49.42"N	11°11'1.77"E	13.67	48	0.71	0.04	0.25
GDB	Zatoka Gdańska	54°31'49.88"N	18°56'35.67"E	6.58	45	0.71	0.01	0.28
STA	Stavanger	58°58'9.98"N	5°45'58.15"E	20.00	48	0.71	0.01	0.28
VEJ	Vejrø	55°0'43.60"N	11°21'1.74"E	13.42	47	0.7	0.02	0.28
ORE	Ore	54°53'34.00"N	12°4'50.00"E	10.13	68	0.69	0.03	0.28
GED	Gedser	54°33'45.00"N	11°58'27.00"E	9.58	48	0.62	0.01	0.38
MEB*	Mecklemburg	54°3'32.04"N	11°25'28.92"E	11.80	63	0.6	0	0.4
HJE	Hjelm	54°54'47.99"N	12°24'45.29"E	8.16	48	0.56	0	0.44
KOI	Köiguste Bay	58°13'59.24"N	23°2'7.83"E	5.21	47	0.55	0	0.45
KOP	Kopli Bay	59°28'0.61"N	24°40'1.74"E	4.23	47	0.54	0.02	0.44
GOT	Gotland	57°27'1.25"N	18°58'15.32"E	6.40	48	0.52	0.01	0.47
HGO	Hgona	59°39'49.91"N	26°0'8.72"E	3.34	48	0.49	0	0.51
KRA	Krassi shallow	59°20'23.34"N	23°46'20.79"E	4.50	47	0.49	0	0.51
KOL*	Kotobrzeg	54°11'19.92"N	15°33'41.81"E	7.24	41	0.49	0	0.51
ASK	Askö	59°19'5.62"N	19°5'31.29"E	5.29	48	0.48	0	0.52
UME	Umeå	63°36'15.62"N	20°38'23.66"E	2.97	48	0.45	0	0.55
MDZ*	Międzyzdroje	53°58'36.84"N	14°27'36.00"E	6.70	51	0.44	0	0.56
HOG	Höga Kusten	62°53'39.97"N	18°48'20.49"E	4.10	28	0.39	0	0.61

ecological interest. Thus, prior to modelling, the Pearson correlation analysis between all environmental variables was run in order to avoid situations of including highly correlated variables into the modelling. The correlation analysis showed that most variables were only weakly intercorrelated at $r < 0.5$. However, winter average nitrate and current velocity were positively correlated ($r = 0.73$, $p < 0.001$). Nevertheless, these values were far below the critical threshold when collinearity begins to severely distort model estimation and subsequent prediction (Dormann et al., 2013).

3. Results and discussion

Mussels from 32 samples from the Baltic Sea, Sound, Great Belt, Little Belt, Kattgat, Limfjord, Skagerrak, North Sea and Wadden Sea (salinity gradient ranging from 3 to 34 PSU) were assayed for nuclear markers Glu-5', ITS, Efbis and M7 (Fig. 1

and Table 1). Already published results of the genotyping of 3 samples collected in 1995 from the southern Baltic (KOL, MDZ and MEB) for Glu-5', ITS and Efbis markers were also included in further analyses (Kijewski et al., 2006, 2011). These 4 markers differed in sensitivity of *Mytilus* spp. discrimination. None of the samples was identified as pure *M. edulis* or *M. trossulus* with all markers, however for three samples (Wadden Sea, Augustenborg and Halse Fjord) genetic composition across all the studied loci was over 95% *M. edulis*. A general pattern of cline variation was observed, with increasing *M. trossulus* share towards the eastern Baltic Sea, but average allele share rose to 61% in Höga Kusten (Bothnian Sea). All Baltic Sea samples revealed a strong introgression of *M. edulis* and a limited introgression of *M. trossulus* through the Danish Straits. A higher frequency of *M. trossulus* genes was reported in populations from the Gulf of Finland and Bothnian Sea as identified using an 8-locus allozyme character set (Väinölä and Strelkov, 2011). SNP analysis revealed

up to 76% of *M. trossulus* characteristic alleles in Höga Kusten (Zbawicka et al., 2014a), which was used in this study.

The frequency of *M. edulis* allele at the codominant Glu-5' locus varied across populations from 1 in the Wadden Sea, Jutland Peninsula to 0.45 in the Bothnian Bay (Table S2A). The *M. trossulus* specific alleles at EFBis marker were distributed across all analysed samples except Augustenborg sample where 100% of *M. edulis* alleles were observed. Average share of *M. trossulus* alleles over all samples was 77% and 5 populations from inner Baltic (Gdańsk Bay, Gotland, Höga Kusten, Kopli Bay) and Vejvø in Smålandsfarvandet, the Great Belt, were monomorphic with *M. trossulus* alleles (Table S2B). The acrosomal sperm protein M7 in this study displays moderate *M. edulis* specific alleles share with maximum (100%) in the northern entrance to the Sound (Hornbæk, Gileleje) and Limfjord (Kås Bredning). Minimum *M. edulis* allele frequency (40%) was noted with this marker in Askö in the central Baltic Sea (Table S2C).

Introgression of *M. edulis* alleles in the Baltic Sea was strong for the ITS marker. Most populations from the Danish coast and Tjärnö were monomorphic with *M. edulis*. The minimum frequency of *M. edulis* was observed in Kõiguste Bay, Estonia (66%) (Table S2D). Clinal differentiation in allele frequencies at Glu-5', EFBis, M7 and ITS loci have been also reported in populations of *Mytilus* from Sounds and South-Western Baltic by Kijewski et al. (2006), Stuckas et al. (2009, 2017) and Väinölä and Strelkov (2011). The higher percentage of *M. edulis* characteristic alleles in sample GDB in comparison with populations from the central Baltic Sea in the

present study and from the shallow site in the Gdańsk Bay presented in Kijewski et al. (2011) can be explained by the different origin of mussels from a deeper site (50 m).

The Bayesian clustering implemented in STRUCTURE for 3 and 4 markers identified two genetic clusters ($K = 2$) corresponding to *M. edulis* and *M. trossulus*, as expected. Three groups of samples could be distinguished. The first group (6 samples) with q values above 0.7 for *M. edulis* cluster (North Sea and northern Danish Straits), The second (18 samples) with values from about 0.6 to 0.2, indicating a high gene admixture (mainly in the Danish Straits), and the last group (11 samples) with q values below 0.2 covering the southern and central Baltic. The result reflects the geographical structure of sampling, with a barrier south to Zeland Island, this being similar to results based on SNP (Zbawicka et al., 2014a). The inner Baltic group consisted of samples representing 11 populations, including Gedser and Hjelm from Lolland, Falster and Møn Islands. The PCoA analysis reveals a similar split of populations into three groups along the 1st axis, confirmed with the AMOVA analysis (Fig. 2).

The large-scale patterns of *Mytilus* spp. populations are highly complex and driven by multiple environmental factors. Earlier studies have suggested that direct environmental and resource gradients define the large-scale distribution pattern of *Mytilus* spp. (e.g. Kotta et al., 2015). Direct environmental gradients represent features that have a direct physiological impact on growth but are not consumed, whereas resource gradients are substances being consumed. The most important direct environmental gradients for *Myti-*

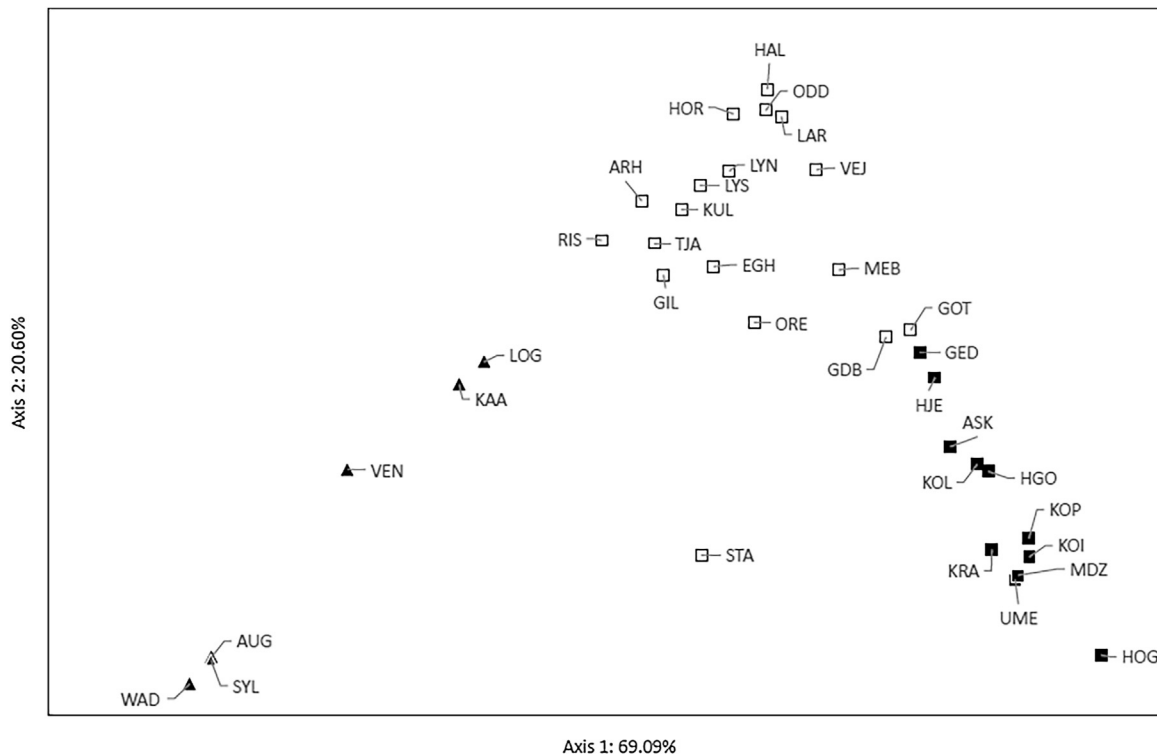


Figure 2 Scatterplot of samples constructed for 3 nuclear DNA markers (EFbis, Glu-5' and ITS). The axes represent the contribution of inertia of the data matrix as an analogue of the total variance in allelic frequencies. Filled triangles – samples from the North Sea (high share of *Mytilus edulis* genes), open squares – samples from transition zone, filled squares – samples from inner and eastern Baltic with higher *Mytilus trossulus* share.

lus spp. are salinity and water exchange and within a favourable distribution range of these direct environmental variables, resource gradients in interaction with direct environmental factors are expected to modulate the patterns of populations. The RF modelling of the allele frequencies of *M. edulis* and *M. trossulus* along environmental gradients described 67 and 68% of the variability in the respective gene frequencies. The model of *M. galloprovincialis* was insignificant and explained only 5% of its allele frequency. The analysis identified strong variability of allele frequencies of *Mytilus* spp. populations along the studied environmental gradients. The most significant source of variation was salinity and to a lesser degree, temperature. The responses of salinity and temperature were not gradual but a clear discontinuity in the allele frequencies was observed at salinity 12 PSU and temperature 16°C (around the eastern islands of Falster and Moen) (Fig. 3). It is important to note that the current statistical analyses involved only arithmetic means of environmental variables and these threshold limits only infer that variability in both salinity and temperature may be behind the current allele frequencies of *Mytilus* spp. The ecological meaning of the observed relationships between environmental variables and the biota is not always obvious and it is as plausible that local salinity and/or temperature minima define the observed patterns of allele frequencies. However, due to high covariation of natural gradients of the minima, means and maxima, their individual effects cannot be separated. Similarly, twenty-six SNP included in the RF analysis were significantly involved in

the differentiation between mussels from the North Sea and Baltic area, showing very sharp differences in frequencies forming clines at the Baltic Sea entrance (Zbawicka et al., 2014a). These SNPs based on F_{ST} outlier analysis have been identified as outliers and could be under selection of salinity stress.

Water nutrients, exposure to waves and chlorophyll-*a*, all contributed to the models but their effects were notably lower compared to the effects of salinity and temperature. As salinity and temperature were not intercorrelated with the other studied environmental variables then the observed high contributions of salinity and temperature in the RF models suggest their significance in shaping the allele frequencies of *Mytilus* spp. in the North and Baltic Seas. The existing salinity gradient is much more stable compared to the patterns of factors such as wind and food availability resulting in a strong selection pressure over a centennial-scale. Moreover, as opposed to other environmental gradients (such as exposure to waves and food availability) salinity values are always suboptimal for *Mytilus* spp. in the Baltic Sea basin. The situation may change in the near future as salinity values are expected to be dramatically changed by current climate change influences (BACC, 2015).

This study characterized the geographical distribution pattern of *Mytilus* mussels in the Baltic Sea, Danish Straits and eastern shore of North Sea in terms of PCR species-specific markers. The blue mussels in the Baltic Sea appeared as an established hybrid swarm containing relics of boreal *M. trossulus* under a strong introgression of Atlantic *M. edulis*

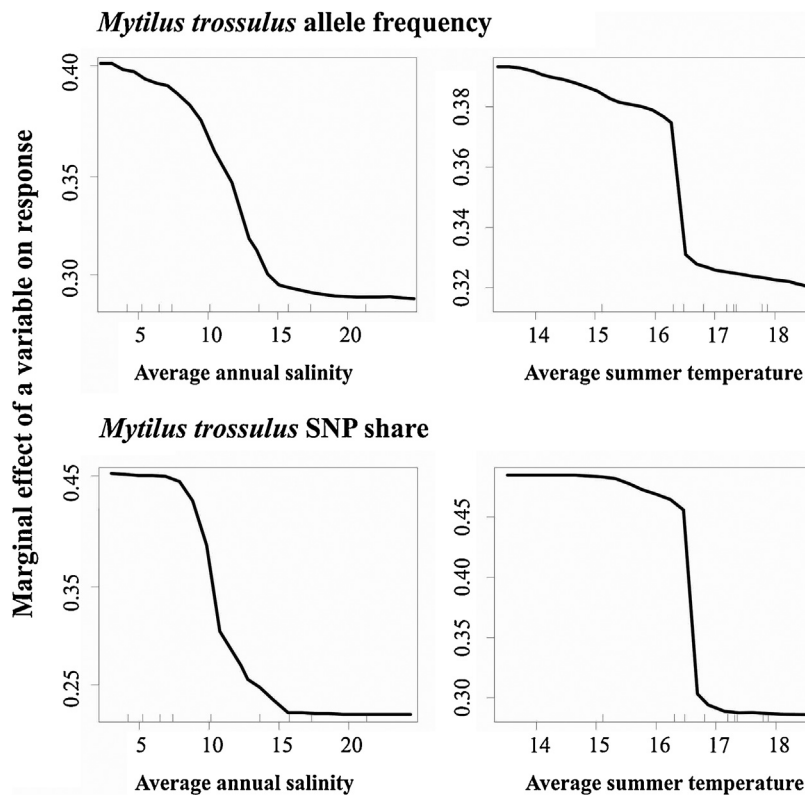


Figure 3 The plots computed in the Random Forest show the relative logit contribution of the variable on the class probability from the perspective of the model. Marginal effect of a variable on response, based on salinity gradient and average frequency of taxon-specific (*Mytilus trossulus*) alleles for EFbis, Glu-5', ITS and M7 markers and 26 SNPs.

(Väinölä and Strelkov, 2011). The genetic structure of the *Mytilus* spp. transition zone in the Danish Straits is maintained by water currents and also larvae dispersal and water salinity (Larsson et al., 2017; Stuckas et al., 2017; Väinölä and Strelkov, 2011). Our data analysis with Bayesian clustering method STRUCTURE displays a prevalent barrier to gene flow in Falster and Lolland Islands. However PCoA analysis displayed a pattern with distinct samples from the North Sea, and another group split into group east – north from Gotland, and from Gdansk Bay to Stavanger. These spatial positions reflect the main division into the Baltic area and the Danish Straits and larval dispersal pattern across the transition zone (Fraïsse et al., 2016; Stuckas et al., 2017). Salinity and temperature are critical aspects in the marginal habitat of the Baltic Sea, setting the limits for the distribution of various species (Bonsdorff and Pearson, 1999). This study demonstrated multigene clines at the Baltic Sea entrance and the main barrier to gene flow in *Mytilus* populations in the area of the south Danish Islands Falster and Lolland.

The generic result of this study is that salinity appeared to define over 50% of the variability in the gene frequencies of the studied *Mytilus* spp. populations. Changes along this environmental gradient are not gradual but instead, a significant shift from gene dominance was found at a salinity of 12 PSU. Other environmental gradients involved water temperature and trophic status of the sea area but these environmental gradients had only moderate linkage with the gene frequencies. Surprisingly, the gradients that define the availability of food resources (e.g. water chlorophyll-*a* and exposure to waves) were nearly insignificant. However, causalities remain to be determined as the observed gene frequencies along the studied environmental gradients were based on correlations, and multiple possible mechanisms not involved with the current study may have influenced these frequencies. Despite these limitations, the results presented here have helped to understand how the gene frequencies of *Mytilus* spp. were distributed along environmental gradients on a regional scale. As the study involved the most important environmental gradients known to shape distribution of *Mytilus* spp. populations at large scales, the results may have pointed to potential reasoning for causation. The study identified strong interregional variability and showed that patterns of gene frequencies of *Mytilus* spp. were primarily a function of water salinity. However, *M. trossulus* is known to be tolerant to lowered salinity (Qiu et al., 2002; Yaroslavl-seva and Sergeeva, 2005), the responsible genetic background still remains unclear (Lockwood and Somero, 2011).

4. Conclusions

This study demonstrated that the novel machine learning technique random forest can be successfully used for finding correlation between genetic differentiation of populations and environmental variables. Populations of the mussels *Mytilus* in the Baltic Sea were used as a model. The genetic differentiation of populations was assayed using 4 known *Mytilus* taxa-specific nuclear DNA markers and published single nucleotide polymorphism markers. Strong genetic differentiation in allele frequencies was correlated with environmental factors along the North Sea – inner Baltic transect.

Environmental gradients explained large variability in the studied allele frequencies with salinity contributing the most to observed genetic differences.

Conflict of interest

The authors declare no conflict of interests.

Acknowledgements

This study was partially funded by the BaltGene Bonus Plus, 03/BONUS/2009 SPB, 2011/01/B/NZ9/04352 NCN project to R.W. and Statutory topic IV.1 in the IO PAN. J.K. and M.R. were supported by the Estonian Research Council, Institutional research funding, IUT02-20. The authors thank Dr Agnieszka Kijewska for assistance in figure preparation.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at <https://doi.org/10.1016/j.oceano.2018.08.002>.

References

- Albalat, A., Potrykus, J., Pempkowiak, J., Porte, C., 2002. Assessment of organotin pollution along the Polish coast (Baltic Sea) by using mussels and fish as sentinel organisms. *Chemosphere* 47 (2), 165–171, [http://dx.doi.org/10.1016/S0045-6535\(01\)00294-6](http://dx.doi.org/10.1016/S0045-6535(01)00294-6).
- BACC Author Team, 2015. In: *II Second Assessment of Climate Change for the Baltic Sea Basin*. Springer-Open, Cham, Heidelberg, New York, Dordrecht, London, 501 pp.
- Bach, L., Zbawicka, M., Strand, J., Wenne, R., 2018. *Mytilus trossulus* in NW Greenland is genetically more similar to North Pacific than NW Atlantic populations of the species. *Mar. Biodivers.*, 7 pp., <http://dx.doi.org/10.1007/s12526-018-0870-0>.
- Banks, S.C., Ling, S.D., Johnson, C.R., Piggott, M.P., Williamson, J. E., Beheregaray, L.B., 2010. Genetic structure of a recent climate change-driven range extension. *Mol. Ecol.* 19 (10), 2011–2024, <http://dx.doi.org/10.1111/j.1365-294X.2010.04627.x>.
- Barsiene, J., Rybakovas, A., Garnaga, G., Andreikenaite, L., 2012. Environmental genotoxicity and cytotoxicity studies in mussels before and after an oil spill at the marine oil terminal in the Baltic Sea. *Environ. Monit. Assess.* 184 (4), 2067–2078, <http://dx.doi.org/10.1007/s10661-011-2100-0>.
- Bergström, P., Carlsson, M.S., Lindegarth, M., Petersen, J.K., Lindegarth, S., Holmer, M., 2017. Testing the potential for improving quality of sediments impacted by mussel farms using bioturbating polychaete worms. *Aquacult. Res.* 48 (1), 161–176, <http://dx.doi.org/10.1111/are.12870>.
- Bierne, N., Bonhomme, F., David, P., 2003. Habitat preference and the marine-speciation paradox. *Proc. R. Soc. Lond. B* 270 (1522), 1399–1406, <http://dx.doi.org/10.1098/rspb.2003.2404>.
- Bjork, M., Gilek, M., 1997. Bioaccumulation kinetics of PCB 31, 49 and 153 in the blue mussel, *Mytilus edulis* L as a function of algal food concentration. *Aquat. Toxicol.* 38 (1–3), 101–123, [http://dx.doi.org/10.1016/S0166-445X\(96\)00837-5](http://dx.doi.org/10.1016/S0166-445X(96)00837-5).
- Bonsdorff, E., Pearson, T.H., 1999. Variation in the sublittoral macrozoobenthos of the Baltic Sea along environmental gradients: a functional-group approach. *Austral. J. Ecol.* 24 (4), 312–326, <http://dx.doi.org/10.1046/j.1442-9993.1999.00986.x>.

- Bosch, S., Tyberghein, L., Deneudt, K., Hernandez, F., De Clerck, O., Sypard, A., 2018. In search of relevant predictors for marine species distribution modelling using the MarineSPEED benchmark dataset. *Divers. Distrib.* 24 (2), 144–157, <http://dx.doi.org/10.1111/ddi.12668>.
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45 (1), 5–32, <http://dx.doi.org/10.1023/A:1010933404324>.
- Breiman, L., Cutler, A., 2015. Package 'randomForest'. <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>.
- Cuéllar-Pinzón, J., Presa, P., Hawkins, S.J., Pita, A., 2016. Genetic markers in marine fisheries: types, tasks and trends. *Fish. Res.* 173, 194–205, <http://dx.doi.org/10.1016/j.fishres.2015.10.019>.
- Dabrowska, H., Kopko, O., Lehtonen, K.K., Lang, T., Waszak, I., Balode, M., Strode, E., 2017. An integrated assessment of pollution and biological effects in flounder, mussels and sediment in the southern Baltic Sea coastal area. *Environ. Sci. Pollut. Res.* 24 (4), 3626–3639, <http://dx.doi.org/10.1007/s11356-016-8117-8>.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquez, J.R.G., Gruber, B., Lafourcade, B., Leitao, P.J., Munkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schroder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36 (1), 27–46, <http://dx.doi.org/10.1111/j.1600-0587.2012.07348.x>.
- Dowd, W.W., Somero, G.N., 2013. Behavior and survival of *Mytilus* congeners following episodes of elevated body temperature in air and seawater. *J. Exp. Biol.* 216 (3), 502–514, <http://dx.doi.org/10.1242/jeb.076620>.
- Earl, D.A., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4 (2), 359–361, <http://dx.doi.org/10.1007/s12686-011-9548-7>.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, Jacob, McC., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29 (2), 129–151, <http://dx.doi.org/10.1111/j.2006.0906-7590.04596.x>.
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77 (4), 802–813, <http://dx.doi.org/10.1111/j.1365-2656.2008.01390.x>.
- Ericson, H., Thorsen, G., Kumblad, L., 2010. Physiological effects of diclofenac, ibuprofen and propranolol on Baltic Sea blue mussels. *Aquat. Toxicol.* 99 (2), 223–231, <http://dx.doi.org/10.1016/j.aquatox.2010.04.017>.
- Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14 (8), 2611–2620, <http://dx.doi.org/10.1111/j.1365-294X.2005.02553.x>.
- Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1, 47–50.
- Filipowicz, M., Burzynski, A., Smietanka, B., Wenne, R., 2008. Recombination in mitochondrial DNA of European mussels *Mytilus*. *J. Mol. Evol.* 67 (4), 377–388, <http://dx.doi.org/10.1007/s00239-008-9157-6>.
- Fraïsse, C., Belkhir, K., Welch, J.J., Bierne, N., 2016. Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Mol. Ecol.* 25 (1), 269–286, <http://dx.doi.org/10.1111/mec.13299>.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29 (1), 1189–1232, <http://dx.doi.org/10.1111/mec.13299>.
- Gardner, J.P., Zbawicka, M., Westfall, K.M., Wenne, R., 2016. Invasive blue mussels threaten regional scale genetic diversity in mainland and remote offshore locations: the need for baseline data and enhanced protection in the Southern Ocean. *Glob. Change Biol.* 22 (9), 3182–3195, <http://dx.doi.org/10.1111/gcb.13332>.
- Gilg, M.R., Hilbish, T.J., 2003. The geography of marine larval dispersal: coupling genetics with fine-scale physical oceanography. *Ecology* 84 (11), 2989–2998, <http://dx.doi.org/10.1890/02-0498>.
- Grabowska, M., Grzelak, K., Kukliński, P., 2015. Rock encrusting assemblages: structure and distribution along the Baltic Sea. *J. Sea Res.* 103, 24–31, <http://dx.doi.org/10.1016/j.seares.2015.05.003>.
- Hastie, T., Tibshirani, R., Friedman, J.H., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, New York, 744 pp.
- Heath, D.D., Rawson, P.D., Hilbish, T.J., 1995. PCR-based nuclear markers identify alien blue mussel (*Mytilus* spp.) genotypes on the west coast of Canada. *Can. J. Fish. Aquat. Sci.* 52 (12), 2621–2627, <http://dx.doi.org/10.1139/f95-851>.
- Hoarau, G., Rijnsdorp, A.D., Van Der Veer, H.W., Stam, W.T., Olsen, J.L., 2002. Population structure of plaice (*Pleuronectes platessa* L.) in northern Europe: microsatellites revealed large-scale spatial and temporal homogeneity. *Mol. Ecol.* 11 (7), 1165–1176, <http://dx.doi.org/10.1046/j.1365-294X.2002.01515.x>.
- Hoher, N., Kohler, A., Strand, J., Broeg, K., 2012. Effects of various pollutant mixtures on immune responses of the blue mussel (*Mytilus edulis*) collected at a salinity gradient in Danish coastal waters. *Mar. Environ. Res.* 75 (SI), 35–44, <http://dx.doi.org/10.1016/j.marenvres.2011.11.003>.
- Inoue, K., Waite, J.H., Matsuoka, M., Odo, S., Harayama, S., 1995. Interspecific variations in adhesive protein sequences of *Mytilus edulis*, *M. galloprovincialis*, and *M. trossulus*. *Biol. Bull.* 189 (3), 370–375, <http://dx.doi.org/10.2307/1542155>.
- Isæus, M., 2004. *Factors structuring Fucus communities at open and complex coastlines in the Baltic Sea*. University of Stockholm Print Center, Stockholm, Frescati, 40 pp.
- Jakobsson, M., Rosenberg, N.A., 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23 (14), 1801–1806, <http://dx.doi.org/10.1093/bioinformatics/btm233>.
- Johannesson, K., André, C., 2006. INVITED REVIEW: life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Mol. Ecol.* 15 (8), 2013–2029, <http://dx.doi.org/10.1111/j.1365-294X.2006.02919.x>.
- Kautsky, N., Evans, S., 1987. Role of biodeposition by *Mytilus edulis* in the circulation of matter and nutrients in a Baltic coastal ecosystem. *Mar. Ecol. Prog. Ser.* 38 (3), 201–212, <http://dx.doi.org/10.3354/meps038201>.
- Kijewska, A., Burzynski, A., Wenne, R., 2009. Molecular identification of European flounder (*Platichthys flesus*) and its hybrids with European plaice (*Pleuronectes platessa*). *ICES J. Mar. Sci.* 66 (5), 902–906, <http://dx.doi.org/10.1093/icesjms/fsp110>.
- Kijewska, A., Kalamarz-Kubiak, H., Arciszewski, B., Guellard, T., Petereit, C., Wenne, R., 2016. Adaptation to salinity in Atlantic cod from different regions of the Baltic Sea. *J. Exp. Mar. Biol. Ecol.* 478, 62–67, <http://dx.doi.org/10.1016/j.jembe.2016.02.003>.
- Kijewski, T., Śmietanka, B., Zbawicka, M., Gosling, E., Hummel, H., Wenne, R., 2011. Distribution of *Mytilus* taxa in European coastal areas as inferred from molecular markers. *J. Sea Res.* 65 (2), 224–234, <http://dx.doi.org/10.1016/j.seares.2010.10.004>.
- Kijewski, T., Wijsman, J.W.M., Hummel, H., Wenne, R., 2009. Genetic composition of cultured and wild mussels *Mytilus* from The Netherlands and transfers from Ireland and Great Britain. *Aquaculture* 287 (3–4), 292–296, <http://dx.doi.org/10.1016/j.aquaculture.2008.10.048>.
- Kijewski, T.K., Zbawicka, M., Väinölä, R., Wenne, R., 2006. Introgression and mitochondrial DNA heteroplasmy in the Baltic populations of mussels *Mytilus trossulus* and *M. edulis*. *Mar. Biol.* 149 (6), 1371–1385, <http://dx.doi.org/10.1007/s00227-006-0316-2>.

- Kopecka, J., Lehtonen, K.K., Barsiene, J., Broeg, K., Vuorinen, P.J., Gercken, J., Pempkowiak, J., 2006. Measurements of biomarker levels in flounder (*Platichthys flesus*) and blue mussel (*Mytilus trossulus*) from the Gulf of Gdansk (southern Baltic). *Mar. Pollut. Bull.* 53 (8–9), 406–421, <http://dx.doi.org/10.1016/j.marpolbul.2006.03.008>.
- Kotta, J., Oganjan, K., Lauringson, V., Parnoja, M., Kaasik, A., Rohtla, L., Kotta, I., Orav-Kotta, H., 2015. Establishing functional relationships between abiotic environment, macrophyte coverage, resource gradients and the distribution of *Mytilus trossulus* in a brackish non-tidal environment. *PLOS ONE* 10 (8), e0136949, <http://dx.doi.org/10.1371/journal.pone.0136949>.
- Larrain, M.A., Zbawicka, M., Aranedo, C., Gardner, J.P.A., Wenne, R., 2018. Native and invasive taxa on the Pacific coast of South America: impacts on aquaculture, traceability and biodiversity of blue mussels (*Mytilus* spp.). *Evol. Appl.* 11 (3), 298–311, <http://dx.doi.org/10.1111/eva.12553>.
- Larsen, P., Riisgård, H., 2016. Growth-prediction model for blue mussels (*Mytilus edulis*) on future optimally thinned farm-ropes in Great Belt (Denmark). *J. Mar. Sci. Eng.* 4 (3), 42, <http://dx.doi.org/10.3390/jmse4030042>.
- Larsson, J., Lind, E.E., Corell, H., Grahn, M., Smolarz, K., Lönn, M., 2017. Regional genetic differentiation in the blue mussel from the Baltic Sea area. *Estuar. Coast. Shelf Sci.* 195, 98–109, <http://dx.doi.org/10.1016/j.ecss.2016.06.016>.
- Larsson, J., Smolarz, K., Świeżak, J., Turower, M., Czerniawska, N., Grahn, M., 2018. Multi biomarker analysis of pollution effect on resident populations of blue mussels from the Baltic Sea. *Aquat. Toxicol.* 198, 240–256, <http://dx.doi.org/10.1016/j.aquatox.2018.02.024>.
- Lesser, M.P., 2016. Climate change stressors cause metabolic depression in the blue mussel, *Mytilus edulis*, from the Gulf of Maine. *Limnol. Oceanogr.* 61 (5), 1705–1717, <http://dx.doi.org/10.1002/lno.10326>.
- Li, S.G., Liu, C., Huang, J.L., Liu, Y.J., Zheng, G.L., Xie, L.P., Zhang, R.Q., 2015. Interactive effects of seawater acidification and elevated temperature on biomineralization and amino acid metabolism in the mussel *Mytilus edulis*. *J. Exp. Biol.* 218 (22), 3623–3631, <http://dx.doi.org/10.1242/jeb.126748>.
- Liaw, A., Wiener, M., 2002. Classification and regression by Random Forest. *R. News* 2/3, 18–22.
- Lockwood, B.L., Somero, G.N., 2011. Transcriptomic responses to salinity stress in invasive and native blue mussels (genus *Mytilus*). *Mol. Ecol.* 20 (3), 517–529, <http://dx.doi.org/10.1111/j.1365-294X.2010.04973.x>.
- Malachowicz, M., Kijewska, A., Wenne, R., 2015. Transcriptome analysis of gill tissue of Atlantic cod *Gadus morhua* L. from the Baltic Sea. *Mar. Genomics* 23, 37–40, <http://dx.doi.org/10.1016/j.margen.2015.04.005>.
- Meier, H.E.M., Hordoir, R., Andersson, H.C., Dieterich, C., Eilola, K., Gustafsson, B.G., Höglund, A., Schimanke, S., 2012. Modeling the combined impact of changing climate and changing nutrient loads on the Baltic Sea environment in an ensemble of transient simulations for 1961–2099. *Climate Dyn.* 39 (9–10), 2421–2441, <http://dx.doi.org/10.1007/s00382-012-1339-7>.
- Michaelidis, B., Portner, H.O., Sokolova, I., Tomanek, L., 2014. Advances in predicting the impacts of global warming on the mussels *Mytilus galloprovincialis* in the Mediterranean Sea. In: Goffredo, S., Dubinsky, Z. (Eds.), *Mediterranean Sea: Its History and Present Challenges*. Springer-Verlag, Berlin, 319–339, http://dx.doi.org/10.1007/978-94-007-6704-1_18.
- Nyberg, E., Faxneld, S., Danielsson, S., Eriksson, U., Miller, A., Bignert, A., 2015. Temporal and spatial trends of PCBs, DDTs, HCHs, and HCB in Swedish marine biota 1969–2012. *Ambio* 44 (Suppl. 3, SI), S484–S497, <http://dx.doi.org/10.1007/s13280-015-0673-5>.
- Ozoliņa, Z., 2017. Mussel farming and its potential in the Baltic Sea. *Econ. Bus.* 30 (1), 40–50, <http://dx.doi.org/10.1515/eb-2017-0004>.
- Peakall, R., Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 28 (19), 2537–2539, <http://dx.doi.org/10.1093/bioinformatics/bts460>.
- Pempkowiak, J., Pazdro, K., Kopecka, J., Perez, E., Sole, M., 2006. Toxicants accumulation rates and effects in *Mytilus trossulus* and *Nereis diversicolor* exposed separately or together to cadmium and PAHs. *J. Environ. Sci. Health Part A: Toxic/Hazard. Subst. Environ. Eng.* 41 (11), 2571–2586, <http://dx.doi.org/10.1080/10934520600927963>.
- Petersen, J.K., Hasler, B., Timmermann, K., Nielsen, P., Tørring, D. B., Larsen, M.M., Holmer, M., 2014. Mussels as a tool for mitigation of nutrients in the marine environment. *Mar. Pollut. Bull.* 82 (1–2), 137–143, <http://dx.doi.org/10.1016/j.marpolbul.2014.03.006>.
- Piwoni-Piórewicz, A., Kuklinski, P., Strekopytov, S., Humphreys-Williams, E., Najorka, J., Iglukowska, A., 2017. Size effect on the mineralogy and chemistry of *Mytilus trossulus* shells from the southern Baltic Sea: implications for environmental monitoring. *Environ. Monit. Assess.* 189 (4), 197, <http://dx.doi.org/10.1007/s10661-017-5901-y>.
- Pocwierz-Kotus, A., Bernas, R., Debowski, P., Kent, M.P., Lien, S., Kesler, M., Titov, S., Leliuna, E., Jespersen, H., Drywa, A., Wenne, R., 2014. Genetic differentiation of southeast Baltic populations of sea trout inferred from single nucleotide polymorphisms. *Anim. Genet.* 45 (1), 96–104, <http://dx.doi.org/10.1111/age.12095>.
- Potrykus, J., Albalat, A., Pempkowiak, J., Porte, C., 2003. Content and pattern of organic pollutants (PAHs, PCBs and DDT) in blue mussels (*Mytilus trossulus*) from the southern Baltic Sea. *Oceanologia* 45 (2), 337–355.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155 (2), 945–959.
- Protasowicki, M., Dural, M., Jaremek, J., 2008. Trace metals in the shells of blue mussels (*Mytilus edulis*) from the Poland coast of Baltic sea. *Environ. Monit. Assess.* 141 (1–3), 329–337, <http://dx.doi.org/10.1007/s10661-007-9899-4>.
- Qiu, J.-W., Tremblay, R., Bourget, E., 2002. Ontogenetic changes in hyposaline tolerance in the mussels *Mytilus edulis* and *M. trossulus*: implications for distribution. *Mar. Ecol. Prog. Ser.* 228, 143–152, <http://dx.doi.org/10.3354/meps228143>.
- Rawson, P.D., Hilbish, T.J., 1998. Asymmetric introgression of mitochondrial DNA among European populations of blue mussels (*Mytilus* spp.). *Evolution* 52 (1), 100–108, <http://dx.doi.org/10.1111/j.1558-5646.1998.tb05142.x>.
- Riddle, B.R., Dawson, M.N., Hadly, E.A., Hafner, D.J., Hickerson, M. J., Mantooth, S.J., Yoder, A.D., 2008. The role of molecular genetics in sculpting the future of integrative biogeography. *Prog. Phys. Geogr.* 32 (2), 173–202, <http://dx.doi.org/10.1177/0309133308093822>.
- Riginos, C., Wang, D., Abrams, A.J., 2006. Geographic variation and positive selection on M7 lysin, an acrosomal sperm protein in mussels (*Mytilus* spp.). *Mol. Biol. Evol.* 23 (10), 1952–1965, <http://dx.doi.org/10.1093/molbev/msl062>.
- Silliman, B.R., Bertness, M.D., Altieri, A.H., Griffin, J.N., Bazterrica, M.C., Hidalgo, F.J., Crain, C.M., Reyna, M.V., 2011. Whole-community facilitation regulates biodiversity on Patagonian rocky shores. *PLOS One* 6 (10), e24502, <http://dx.doi.org/10.1371/journal.pone.0024502>.
- Smietanka, B., Zbawicka, M., Sanko, T., Wenne, R., Burzynski, A., 2013. Molecular population genetics of male and female mitochondrial genomes in subarctic *Mytilus trossulus*. *Mar. Biol.* 160 (7), 1709–1721, <http://dx.doi.org/10.1007/s00227-013-2223-7>.
- Stuckas, H., Knobel, L., Schade, H., Breusing, C., Hinrichsen, H.H., Bartel, M., Langguth, K., Melzner, F., 2017. Combining hydrodynamic modelling with genetics: can passive larval drift shape the

- genetic structure of Baltic *Mytilus* populations? Mol. Ecol. 26 (10), 2765–2782, <http://dx.doi.org/10.1111/mec.14075>.
- Stuckas, H., Stoof, K., Quesada, H., Tiedemann, R., 2009. Evolutionary implications of discordant clines across the Baltic *Mytilus* hybrid zone (*Mytilus edulis* and *Mytilus trossulus*). Heredity 103 (2), 146–156, <http://dx.doi.org/10.1038/hdy.2009.37>.
- Sylvester, E.V.A., Bentzen, P., Bradbury, I.R., Clement, M., Pearce, J., Horne, J., Beiko, R.G., 2018. Applications of random forest feature selection for fine-scale genetic population assignment. Evol. Appl. 11 (2), 153–165, <http://dx.doi.org/10.1111/eva.12524>.
- Szefer, P., Frelek, K., Szefer, K., Lee, C.B., Kim, B.S., Warzocha, J., Zdrojewska, I., Ciesielski, T., 2002. Distribution and relationships of trace metals in soft tissue, byssus and shells of *Mytilus edulis trossulus* from the southern Baltic. Environ. Pollut. 120 (2), 423–444, [http://dx.doi.org/10.1016/S0269-7491\(02\)00111-2](http://dx.doi.org/10.1016/S0269-7491(02)00111-2).
- Szefer, P., Szefer, K., 1990. Metals in molluscs and associated bottom sediments of the Southern Baltic. Helgolander Meeresunters. 44 (3–4), 411–424, <http://dx.doi.org/10.1007/BF02365477>.
- Szefer, P., Wenne, R., 1987. Concentration of uranium and thorium in molluscs inhabiting Gdańsk Bay, Baltic Sea. Sci. Total Environ. 65, 191–202, [http://dx.doi.org/10.1016/0048-9697\(87\)90172-0](http://dx.doi.org/10.1016/0048-9697(87)90172-0).
- Szymczak-Zyla, M., Wawrzyniak-Wydrowska, B., Kowalewska, G., 2006. Products of chlorophyll a transformation by selected benthic organisms in the Odra Estuary (southern Baltic Sea). Hydrobiologia 554, 155–164, <http://dx.doi.org/10.1007/s10750-005-1016-5>.
- Telesca, L., Michalek, K., Sanders, T., Peck, L.S., Thyrring, J., Harper, E.M., 2018. Blue mussel shell shape plasticity and natural environments: a quantitative approach. Sci. Rep. 8, Art. No. 2865, 15 pp., <http://dx.doi.org/10.1038/s41598-018-20122-9>.
- Thomsen, J., Melzner, F., 2010. Moderate seawater acidification does not elicit long-term metabolic depression in the blue mussel *Mytilus edulis*. Mar. Biol. 157 (12), 2667–2676, <http://dx.doi.org/10.1007/s00227-010-1527-0>.
- Utermann, C., Parrot, D., Breusing, C., Stuckas, H., Staufenberger, T., Blumel, M., Labes, A., Tasdemir, D., 2018. Combined genotyping, microbial diversity and metabolite profiling studies on farmed *Mytilus* spp. from Kiel Fjord. Sci. Rep. 8, Art. No. 7983, 13 pp., <http://dx.doi.org/10.1038/s41598-018-26177-y>.
- Väinölä, R., Hvilson, M.M., 1991. Genetic divergence and a hybrid zone between Baltic and North Sea *Mytilus* populations (Mytilidae: Mollusca). Biol. J. Linn. Soc. 43 (2), 127–148, <http://dx.doi.org/10.1111/j.1095-8312.1991.tb00589.x>.
- Väinölä, R., Strelkov, P., 2011. *Mytilus trossulus* in Northern Europe. Mar. Biol. 158 (4), 817–833, <http://dx.doi.org/10.1007/s00227-010-1609-z>.
- Valladares, F., Matesanz, S., Guilhaumon, F., Araujo, M.B., Balaguer, L., Benito-Garzon, M., Cornwell, W., Gianoli, E., van Kleunen, M., Naya, D.E., Nicotra, A.B., Poorter, H., Zavala, M.A., 2014. The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. Ecol. Lett. 17 (11), 1351–1364, <http://dx.doi.org/10.1111/ele.12348>.
- Vuorinen, I., Hänninen, J., Rajasilta, M., Laine, P., Eklund, J., Montesino-Pouzols, F., Corona, F., Junker, K., Meier, H.E.M., Dippner, J.W., 2015. Scenario simulations of future salinity and ecological consequences in the Baltic Sea and adjacent North Sea areas – implications for environmental monitoring. Ecol. Indic. 50, 196–205, <http://dx.doi.org/10.1016/j.ecolind.2014.10.019>.
- Watson, J.R., Hays, C.G., Raimondi, P.T., Mitarai, S., Dong, C., McWilliams, J.C., Blanchette, C.A., Caselle, J.E., Siegel, D.A., 2011. Currents connecting communities: nearshore community similarity and ocean circulation. Ecology 92 (6), 1193–1200, <http://dx.doi.org/10.1890/10-1436.1>.
- Wenne, R., 2018. Single nucleotide polymorphism markers with applications in aquaculture and assessment of its impact on natural populations. Aquat. Living Resour. 31, Art. No. 2, 17 pp., <http://dx.doi.org/10.1051/alr/2017043>.
- Wenne, R., Bach, L., Zbawicka, M., Strand, J., McDonald, J.H., 2016. A first report on coexistence and hybridization of *Mytilus trossulus* and *M. edulis* mussels in Greenland. Polar Biol. 39, 343–355, <http://dx.doi.org/10.1007/s00300-015-1785-x>.
- Wennerström, L., Laikre, L., Ryman, N., Utter, F.M., Ab Ghani, N.I., André, C., DeFaveri, J., Johansson, D., Kautsky, L., Merilä, J., Mikhailova, N., Pereyra, R., Sandström, A., Teacher, A.G.F., Wenne, R., Vasemägi, A., Zbawicka, M., Johannesson, K., Primmer, C.R., 2013. Genetic biodiversity in the Baltic Sea: species-specific patterns challenge management. Biodiv. Conserv. 22 (13–14), 3045–3065, <http://dx.doi.org/10.1007/s10531-013-0570-9>.
- Yaroslavtseva, L.M., Sergeeva, E.P., 2005. Adaptivity of the bivalve *Mytilus trossulus* larvae to short and long-term changes in water temperature and salinity. Rus. J. Mar. Biol. 32 (2), 82–87, <http://dx.doi.org/10.1134/S1063074006020027>.
- Zbawicka, M., Burzynski, A., Wenne, R., 2007. Complete sequences of mitochondrial genomes from the Baltic mussel *Mytilus trossulus*. Gene 406, 191–198, <http://dx.doi.org/10.1016/j.gene.2007.10.003>.
- Zbawicka, M., Drywa, A., Śmietanka, B., Wenne, R., 2012. Identification and validation of novel SNP markers in European populations of marine *Mytilus* mussels. Mar. Biol. 159 (6), 1347–1362, <http://dx.doi.org/10.1007/s00227-012-1915-8>.
- Zbawicka, M., Saňko, T., Strand, J., Wenne, R., 2014a. New SNP markers reveal largely concordant clinal variation across the hybrid zone between *Mytilus* spp. in the Baltic Sea. Aquat. Biol. 21 (21), 25–36, <http://dx.doi.org/10.3354/ab00566>.
- Zbawicka, M., Skibinski, D., Wenne, R., 2003a. Doubly uniparental transmission of mitochondrial DNA length variants in the mussel *Mytilus trossulus*. Mar. Biol. 142 (3), 455–460, <http://dx.doi.org/10.1007/s00227-002-0969-4>.
- Zbawicka, M., Trucco, M.I., Wenne, R., 2018. Single nucleotide polymorphisms in native South American Atlantic coast populations of smooth shelled mussels: hybridization with invasive European *Mytilus galloprovincialis*. Genet. Sel. Evol. 50, Art. No. 5, 14 pp., <http://dx.doi.org/10.1186/s12711-018-0376-z>.
- Zbawicka, M., Wenne, R., Burzynski, A., 2014b. Mitogenomics of recombinant mitochondrial genomes of Baltic Sea *Mytilus* mussels. Mol. Genet. Genomics: MGG 289 (6), 1275–1287, <http://dx.doi.org/10.1007/s00438-014-0888-3>.
- Zbawicka, M., Wenne, R., Skibinski, D.O.F., 2003b. Mitochondrial DNA variation in populations of the mussel *Mytilus trossulus* from the Southern Baltic. Hydrobiologia 499, 1–12, <http://dx.doi.org/10.1023/A:1026356603105>.