ELECTRONIC SUPPLEMENTS

Supplement 1. Hydrodynamic model validation

The hydrodynamic model was calibrated using temperature and salinity profile measurements from 2010 at 7 stations distributed in the western parts of Limfjorden (Fig. S1). The profile data was obtained from the Danish national database on aquatic monitoring data (ODA) at https://odaforalle.au.dk. Results from the hydrodynamic model were validated with measurements from the year 2009, 2011, 2012 and 2017 using temperature and salinity data from the same 7 profile stations and 10 minute resolved time series of water level data from 8 positions distributed throughout the western Limfjorden. The water level data was obtained from the Danish meteorological institute. The agreement between modelled and measured water level is very good at all 8 positions, with an overall bias of less than 3 cm and a root mean square error (RMSE) of 13.5 cm (Table S3). No direct current measurements have been available for validation, but the good representation of water level is a strong indicator that the currents are also well represented by the hydrodynamic model (Fig. S2). Temperature is also well represented in the model with an overall bias of less than 0.5 °C and a RMSE of 1.4 °C (Fig. S3) (Table S1). In the inner parts of Limfjorden, the deeper water tends to lag the yearly temperature variation. Salinity has a strong spatial variation, but a weak temporal variation. It is well represented by the model in the bulk of the Limfjorden (Table S2). However, in Skive Fjord and Lovns Basin (areas 13 and 14 Fig.1. in the main text) which also receive the largest fresh water inputs, the model underestimates salinity (Fig. S4) and also fails to reproduce the stratification. However, the water level is well represented in Skive, implying realistic water velocities. The lack of high saline bottom water is, therefore most likely due to an insufficient description of the vertical turbulent mixing. The overall salinity bias is 2.4 PSU and the RMSE 3.1 PSU.

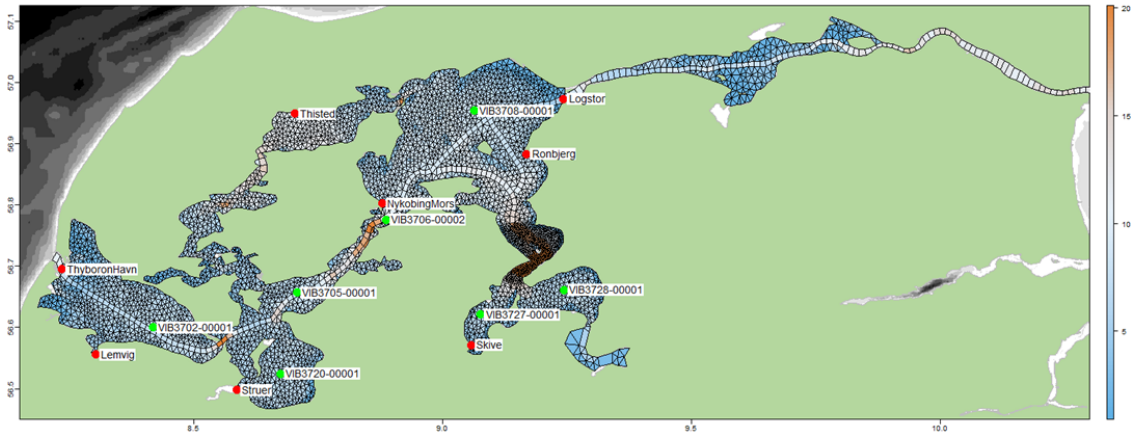

Fig. S1. Limfjorden model mesh and validation stations and positions. Green dots correspond to ODA validation stations for temperature and salinity. Red dots correspond to the Danish Meteorological Institute (DMI) validation stations for water level. Color bar indicates depth in meters.

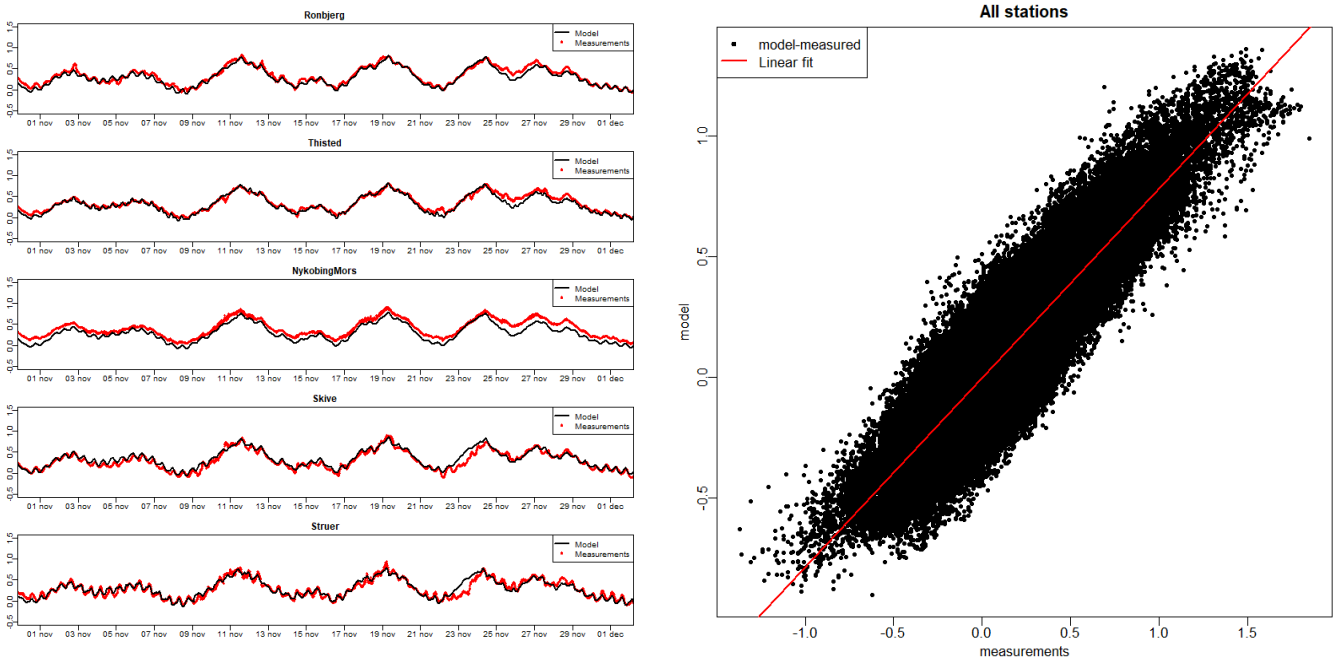

Fig. S2. Water level. Model (black) vs. Measurement (red) validation.

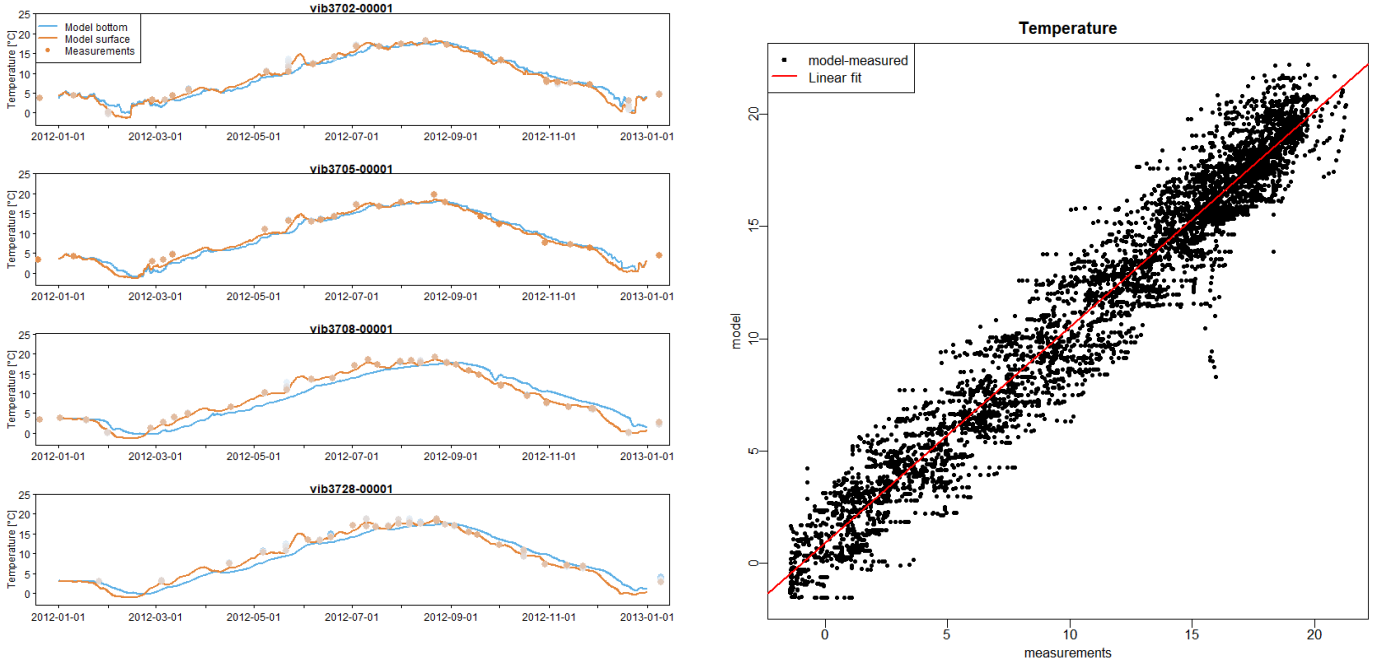

Fig. S3. Temperature. Model (lines) vs. Measurement (points) for validation.

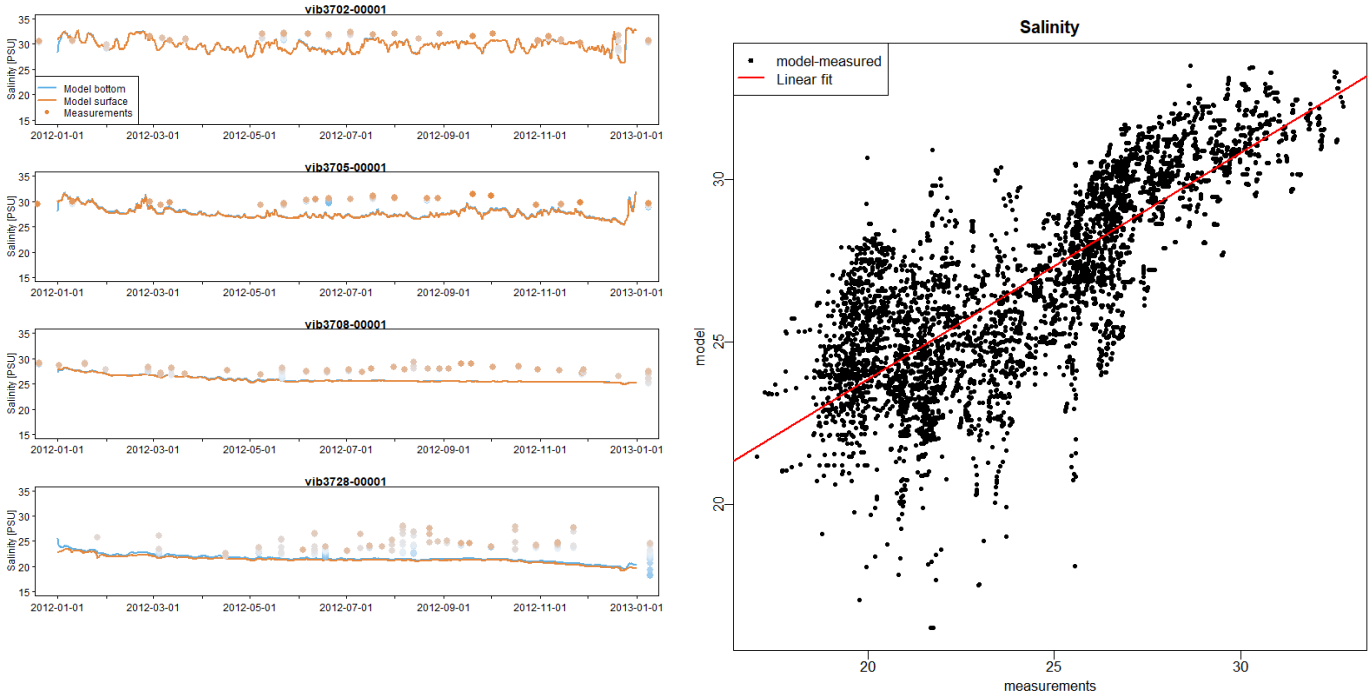

Fig. S4. Salinity. Model (lines) vs. Measurement (points) for validation.

Table S1. Temperature values from the model vs. Measurement for validation. . Number of measurements (n), mean difference between model and measurement (bias), root mean square error between model and measurement (rms), ratio of model standard deviation to measurement standard deviation (nsd), R-Squared (r2) and P-value (pval) of linear fit.

Station	n	bias	rms	nsd	r2	pval
VIB3702-00001	742	-0.29	1.17	0.99	0.96	0
VIB3705-00001	574	-0.35	1.18	1.01	0.96	0
VIB3706-00002	803	-0.67	1.47	1.02	0.95	0
VIB3708-00001	1106	-0.31	1.79	0.99	0.93	0
VIB3720-00001	228	-0.41	1.35	0.93	0.94	0
VIB3727-00001	786	-0.84	1.51	1.01	0.96	0
VIB3728-00001	474	-0.50	1.34	0.98	0.93	0
All	6187	-0.46	1.44	1.01	0.95	0

Table S2. Salinity values for the model vs. Measurement for validation. Number of measurements (n), mean difference between model and measurement (bias), root mean square error between model and measurement (rms), ratio of model standard deviation to measurement standard deviation (nsd), R-Squared (r2) and P-value (pval) of linear fit.

Station	**n**	**bias**	**rms**	**nsd**	**r2**	**pval**
VIB3702-00001	742	-1.18	1.71	1.17	0.20	<1E-20
VIB3705-00001	574	-2.47	2.74	0.87	0.13	<1E-10
VIB3706-00002	803	-2.44	2.79	0.53	0.06	<1E-10
VIB3708-00001	1106	-1.33	2.04	0.43	0.00	NS
VIB3720-00001	228	-1.98	2.17	0.82	0.04	0.0026
VIB3727-00001	786	-5.03	5.34	0.60	0.07	<1E-10
VIB3728-00001	474	-1.97	2.60	0.25	0.06	<1E-6
All	6187	-2.40	3.13	1.15	0.64	0

Table S3. Water level values for the model vs. Measurement for validation. Mean difference between model and measurement (bias), root mean square error between model and measurement (rms), ratio of model standard deviation to measurement standard deviation (nsd), R-Squared (r2) and P-value (pval) of linear fit.

Position	**bias**	**rms**	**nsd**	**r2**	**pval**
Ronbjerg	-0.08	0.15	0.84	0.84	0
Løgstør	-0.01	0.12	0.84	0.85	0
Thisted	0.03	0.10	0.90	0.87	0
NykobingMors	-0.06	0.11	0.88	0.90	0
Skive	0.01	0.11	0.90	0.86	0
ThyboronHavn	0.01	0.17	0.85	0.74	0
Lemvig	-0.11	0.18	0.99	0.75	0
Struer	0.03	0.11	1.07	0.85	0
All	-0.03	0.14	0.88	0.80	0

Supplement 2. Sensitivity analysis

The sensitivity was investigated using a one-parameter- at-a-time analysis in which each model parameter was separately varied in three ways high, medium and low change in the parameter and compared to a control scenario while the other parameters were kept unchanged (Gomes et al. 2017). The choice of the parameter range was made to cover all possible scenarios but still within the biological limits of the species. The parameters tested in the sensitivity analysis were: number of agents released per station (with a total of 207 stations), pelagic larval duration (PLD), number of releases per month during May, number of stations used for the release, and diffusivity coefficient (Table S4). The parameter values for the sensitivity analysis were carefully chosen after a literature review on the same species and related studies. For the number of agents, extra tests were made choosing a maximum number of 1000 agents per site (released 4 times) and reducing it by a factor of 10. In this paper we examine 1000, 10 and 1 agents released per site. For the PLD, simulations were made at an early model stage with different pelagic larval durations from 11 to 46 days (PLD for bivalves adjusted from Moksnes et al. 2014) with a weekly increase in the number of days. The pelagic larval duration of 21 days was set as our control, the reported value for mussel larvae (Widdows 1991, Riisgård et al. 2015). . Release events were tested in 3 different ways, 1 release, 15 releases, and 30 releases within the same spawning. All scenarios were conducted with May as the main spawning month. The number of stations included in the study was also tested. We started with all stations and reduced them randomly by 50%. Finally, diffusivity added to the larvae was tested by a factor of ±10. The dispersal was calculated from the dispersion coefficients used in similar studies by Treml et al. (2015). The formula for calculating the dispersion using a random walk was obtained from (Hansen et al. 2015)

Table S4. Model parameters used in the sensitivity analysis for the Limfjorden Agent Based Model. Values are shown as low, medium and high change in the parameter. Selected values were applied in the final simulation.

Parameter description	Values	Selected value
Number of agents released per station	[1,10,1000]	400
Pelagic larval duration (days)	[11,34,46]	21
Number of releases per month	[1,4,30]	30
Number of stations	[104,207]	207
Diffusivity (m s^{-1})	[0.01,1,10]	0.01

Once the parameter was modified, we run the ABM model and stored the end positions of the larvae. We then calculated the downstream connectivity probability (Eq.1 in the main text) and obtained 17x17 matrices per scenario. These results were evaluated calculating a simple sensitivity index (SI) based on Pethybridge et al. 2013 equation:

$$SI = \frac{1}{n}\sum_{t=1}^{n}\frac{|C_t^0 - C_t^1|}{C_t^0} * 100(\%) \qquad \text{(Eq.S1)}$$

Where n is the number of simulated days, C_t^0 is the downstream connectivity probability matrix resulted from the control and C_t^1 the downstream connectivity probability matrix resulted from the scenario with the change is the parameter we want to test. This way we obtained a second connectivity matrix with the sensitivity index (SI) calculated in each of the cells. The main results from the sensitivity tests were plotted as simple bars containing mean, and value range from all the matrix connections. Fig. S5 shows the results from all of the tested parameters and their difference to the control (tests with number of stations not shown). The control scenario is the zero value in the y-axis. The control scenario was set up to: 100 agents per site, 21 days of PLD, four releases in May, and diffusivity of 0.01 m s^{-1}.

Model results revealed to be not sensitive to the use of 1000 agents per site. However, a threshold was found when using 10 agents per site and less, were the results became highly sensitive to the control (Fig. S5). The number of agents released per site was set to 400. During the first tests with PLDs, 11, 18 and 21 days produced a similar result in terms of proportion of larvae settling in most of the areas. However, when we weekly increased the PLD from 21 days to 46 days we observed a decrease of larvae in most subareas (not shown) and the pelagic larval duration became a very sensitive parameter. The PLD was set to 21 days, the minimum reported value for mussel larvae (Widdows 1991, Riisgård et al. 2015). The number of releases per simulation was very sensitive when there was only one release per month compared to the control value of four releases per month. It was decided to use 30 releases, one per day at random times. The model also appeared to be very sensitive to changes in the diffusivity coefficient. It was therefore decided to use a low speed (Treml et al. 2015) for larval movements to be primarily dominated by hydrodynamic processes.

The sensitivity analysis shows the relative importance with respect to the change on each model parameter compared to the control. The results of the tests were considered when choosing the parameter values as input in the model (Table S4).

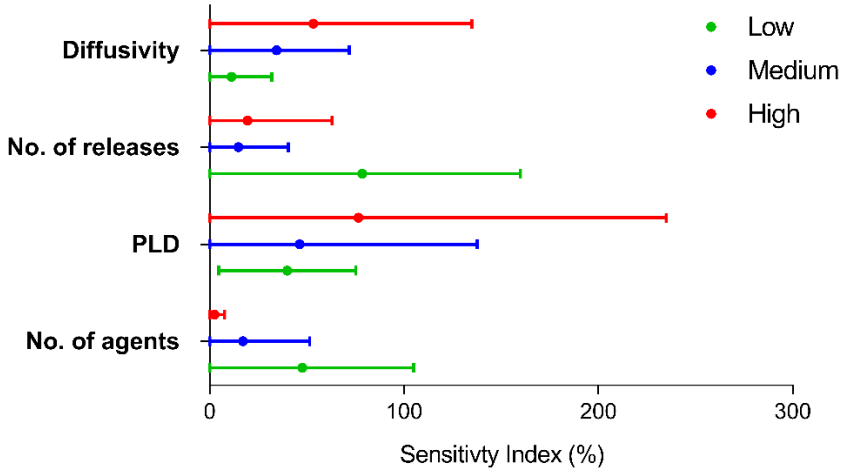

Fig. S5. Sensitivity analysis results for downstream connectivity matrices with a change in parameter. Zero value on x axis indicates the control scenario where the rest of the parameters are compared to. Mean values of the 17x17 matrices are represented as dots. Value range is represented with the bars. The color indicates the value of the parameter (low, medium or high) See Table S4. The larger the bar, the larger the difference in the results compared to the control and vice versa. Control is set up as: Diffusivity (0.01 m^2 s^{-1}), No. of agents (100 per site), No. of releases (4), PLD (21 days).

Supplement 3. Genetic analysis

In this section we show supplementary figures and tables from the genetic analysis conducted in the study.

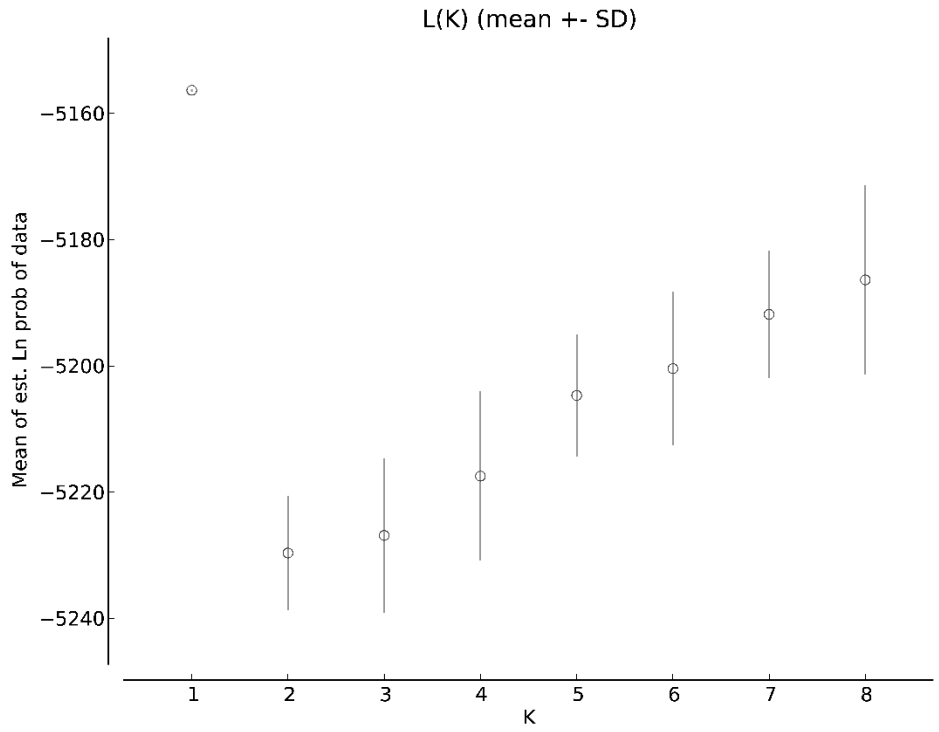

Fig. S6. Mean log likelihood of the data in Structure for the 20 runs for each K between 1 and 8. The error bars represent the standard deviation (SD). This plot was produced by the online tool Structure harvester (Earl and vonHoldt, 2012).

Table S5. Fst values between all pairs of sampling sites. They are labeled as: VE (Venøsund), EE (East of Venøsund), DR (Dråby Vig), SA (Sallingsund), TE (Løgstør Basin), SK (Skive Fjord) and FO (Lovns Basin). Fst values and their confidence intervals (0.025 and 0.975 quantiles) are presented above the diagonal. Corrected *p*-values using the Holm-Bonferroni method are presented below the diagonal.

	VE	EE	SA	DR	TE	SK	FO
VE		-0.004 [-0.008, 0.000]	-0.002 [-0.009, 0.007]	0.002 [-0.005, 0.011]	0.007 [-0.002, 0.017]	0.004 [-0.004, 0.012]	-0.001 [-0.007, 0.007]
EE	1		0.006 [-0.009, 0.028]	0.000 [-0.009, 0.010]	0.005 [-0.009, 0.022]	0.010 [-0.001, 0.020]	0.001 [-0.005, 0.009]
SA	1	1		0.013 [-0.005, 0.040]	0.014 [-0.005, 0.043]	0.010 [-0.009, 0.044]	-0.003 [-0.009, 0.005]

DR	1	1	0.48		0.004 [-0.010, 0.021]	0.011 [-0.004, 0.028]	0.002 [-0.006, 0.011]
TE	1	1	0.88	1		0.008 [-0.004, 0.021]	-0.002 [-0.009, 0.007]
SK	1	0.43	0.58	0.93	1		0.003 [-0.006, 0.014]
FO	1	0.58	1	1	1	1	

Table S6. Expected and observed mean heterozygosities for each sampling site. They are labeled as: VE (Venøsund), EE (East of Venøsund), DR (Dråby Vig), SA (Sallingsund), TE (Løgstør Basin), SK (Skive Fjord) and FO (Lovns Basin). Heterozygosities were computed without the mitochondrial marker.

Genetic site	Expected	Observed
DR	0.304946	0.303968
EE	0.311329	0.28927
FO	0.32698	0.335955
SA	0.285955	0.275456
SK	0.306393	0.303041
TE	0.316938	0.33042
VE	0.308149	0.298975

LITERATURE CITED

Earl DA and vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4(2):359-361

Gomes VM, Assis AO, Matias CA, Saraiva JP, Gomes FA, Wainer GA, Lima BS, Magalhaes AS, Calixto WP, Flores PHR, Bulhoes JS (2017) Analytical method for calculating the sensitivity index of system parameters. CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), Pucon, p 1-6

Hansen FT, Potthoff M, Uhrenholdt T, Vo HD, Linden O, Andersen JH (2015) Development of a prototype tool for ballast water risk management using a combination of hydrodynamic models and agent-based modeling. WMU Journal of Maritime Affairs, 14:219-24

Moksnes PO, Jonsson P, Jacobi MN, Vikström K (2014) Larval Connectivity and Ecological Coherence of Marine Protected Areas (MPAs) in the Kattegat-Skagerrak Region. Swedish Institute for the Marine Environment. Report No 2014:2

Pethybridge H, Roos D, Loizeau V, Pecquerie L, Bacher C (2013) Responses of European anchovy vital rates and population growth to environmental fluctuations: an individual-based modeling approach. Ecological Modelling, 250: 370–383

Riisgård HU, Lundgreen K, Pleissner D (2015) Environmental Factors and Seasonal Variation in Density of Mussel Larvae (*Mytilus edulis*) in Danish Waters. Open Journal of Marine Science, 5:280-289

Treml EA, Ford JR, Black KP, Swearer SE (2015) Identifying the key biophysical drivers, connectivity outcomes, and metapopulation consequences of larval dispersal in the sea. Movement Ecology, 3(1): 1–16

Widdows J (1991) Physiological ecology of mussel larvae. Aquaculture, 94(c):147–163