TITLE: Multiple drivers of interannual oyster settlement and recruitment in the lower Chesapeake Bay
JOURNAL: Conservation Genetics
AUTHORS: Brendan Turley, Kimberly Reece, Jian Shen, Jeong-Ho Lee, Ximing Guo, Jan McDowell
CORRESPONDING AUTHOR: Jan McDowell, Virginia Institute of Marine Science, mcdowell@vims.edu

Electronic Supplementary Material 1

SNP Panel Development

The SNP panel was developed using the SNP-type chemistry available from Fluidigm Corporation (San Francisco, CA) for use on the Fluidigm Biomark. SNP-type assays incorporate one of two alternate fluorescence molecules for discrimination of SNP alleles per locus. The Biomark is a high-throughput platform running a modified polymerase chain reaction (PCR) that adds fluorescence molecules to the newly synthesized DNA strands and records an image of the resulting products. The native software processes the images to determine relative fluorescence values for each sample. Scatterplots of the fluorescence for each locus are used to designate heterozygous and homozygous individuals.

The SNP primers were designed by Fluidigm to amplify the genomic region containing the SNP from previously determined flanking sequences. Candidate SNPs were chosen based upon sequence conformance to Fluidigm's design restrictions: (1) sequences needed to be a minimum of 60 base pairs in length, (2) a maximum of 250 base pairs are needed on either side of a SNP position, (3) SNP targets must be bi-allelic (i.e., only have two alternative states), (4) there cannot be another SNP within 20 base pairs on either side of the target SNP, and finally, (5) sequences should not have guanine-cytosine content greater than 65%. Upon submission, Fluidigm designs sets of primers for the desired SNP target for use in the Biomark. In total, 95 SNP sequences were submitted and tested on the Fluidigm Biomark located at the Virginia Institute of Marine Science.

A test panel of 48 oysters collected throughout much of *C. virginica's* range was constructed to evaluate the SNP-type primers. Oysters from the Lafayette River, Elizabeth River, James River, Maryland, Eastern Shore of Virginia, Long Island Sound, Florida, Alabama, and Louisiana were used. The test panel oysters were genotyped using the recommended SNP-type protocol. SNP Genotyping Analysis ver. 4.1.2 software developed by Fluidigm was used to cluster the processed fluorescence images. K-means clustering with a confidence level of 65% was used to define clusters and designate heterozygote and homozygote genotypes. The confidence level is the distance from a cluster center where a percentage of fluorescence values are located. The clusters were checked by eye and given a quality score ranging from zero to four to account for polymorphism and how well defined the fluorescence clusters appeared. A quality score of zero was assigned to scatterplots with no cluster pattern or complete failure of the marker with relative fluorescence below 0.3. A quality score of one was assigned to monomorphic scatterplots indicated by only one cluster. A quality score of two was assigned to polymorphic scatterplots with two clusters visible. A quality score of three was assigned to polymorphic scatterplots with three clusters visible but the clusters were not well defined and with some ambiguous calls. A quality score of four was assigned to polymorphic scatterplots with three well-defined clusters and few ambiguous calls. Loci were chosen because polymorphism was observed in the test panel while monomorphic loci were excluded. Polymorphic loci are desirable because it increases the likelihood that individuals and populations can be distinguished from each other. Forty-eight total SNP loci were chosen for genotyping. Only 41 loci were used in population genetic analyses. The loci myc-326 and sab-145 were removed because they had no fluorescence across the majority of samples. The mtDNA loci was removed because it had monomorphic expression across individual samples. The locus at-473 was removed because it was a duplicate of locus pt-473 serving as an analytical control. The loci fer-116 and rpl4-493 were removed because there was low fluorescence across the majority of individual samples. The locus myc-317 was removed because it had a monomorphic expression across individual samples. The STA primers were used for the initial PCR normalization runs (Electronic Supplementary Material 2), and the ASP primers are the assay specific primers used in the Biomark for fluorescence acquisition.

The SNP loci were tested for evidence of selection using the program BayeScan ver. 2.1 (Foll and

Gaggiotti 2008). The BayesScan identifies outlier loci potentially indicative of locus specific selection. The false discovery rate Q-value was set at 0.05. The reference samples were grouped per geographic location and analyzed for population structure. Pairwise F_{ST} values were calculated using Weir and Cockerham (1984) method implemented in Arlequin ver. 3.5 (Excoffier and Lischer 2010).

Significance correction

Briefly, a type-I error rate (α) of 0.05 was divided by the sum of one over one through the number of tests being performed to determine an adjusted α (Benjamini and Yekutieli 2001).

Model description

A hydrodynamic connectivity model (Sisson and Shen 2012) to simulate oyster larval dispersal and settlement originating in the Lafayette River was developed at the Virginia Institute of Marine Science (VIMS) under contract with the Chesapeake Bay Foundation (CBF). This model is a modification of the VIMS three-dimensional Hydrodynamic Eutrophication Model (HEM-3D, Sisson et al. 1997). The James and the Elizabeth Rivers were included in the model because their hydrodynamics influence that of the Lafayette River and the model's lateral resolution within the Lafayette River is about 50-100 m. Freshwater inputs into the James River used in the model were obtained from USGS gauges for the James, Appomattox, and Chickahominy Rivers. The Elizabeth River freshwater inputs come from urban runoff and fluctuate with precipitation. Salinity information was obtained from the Chesapeake Bay model based upon hourly reports for the summer of 2008. Within the Lafayette River, eleven oyster reef locations were selected by the CBF and modeled as oyster spawning sites. Reef size was set at 4047 square meters (one acre) and at each reef a million larvae were released per square meter. Larvae were released from the bottom and a daily mortality rate was applied at 0.18 day^{-1} (Sisson and Shen 2012). Simple oyster larvae behaviors were incorporated in which larvae swim upwards for a predetermined time after a change in salinity threshold is detected as per North et al. (2008). The simulation was run for 222 virtual days and the larval component ran the last 21 days equaling a 14-day planktonic phase and a 7-day settlement phase. Additionally, the model was run with and without runoff using averaged rain gauge measurements for 2008 as a typical year. Most oyster larvae produced in the Lafayette River was predicted to be retained in the system until settlement (Figure S3).

Literature Cited

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* 29(4):1165–1188

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10(3):564–567

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* 180(2):977–993

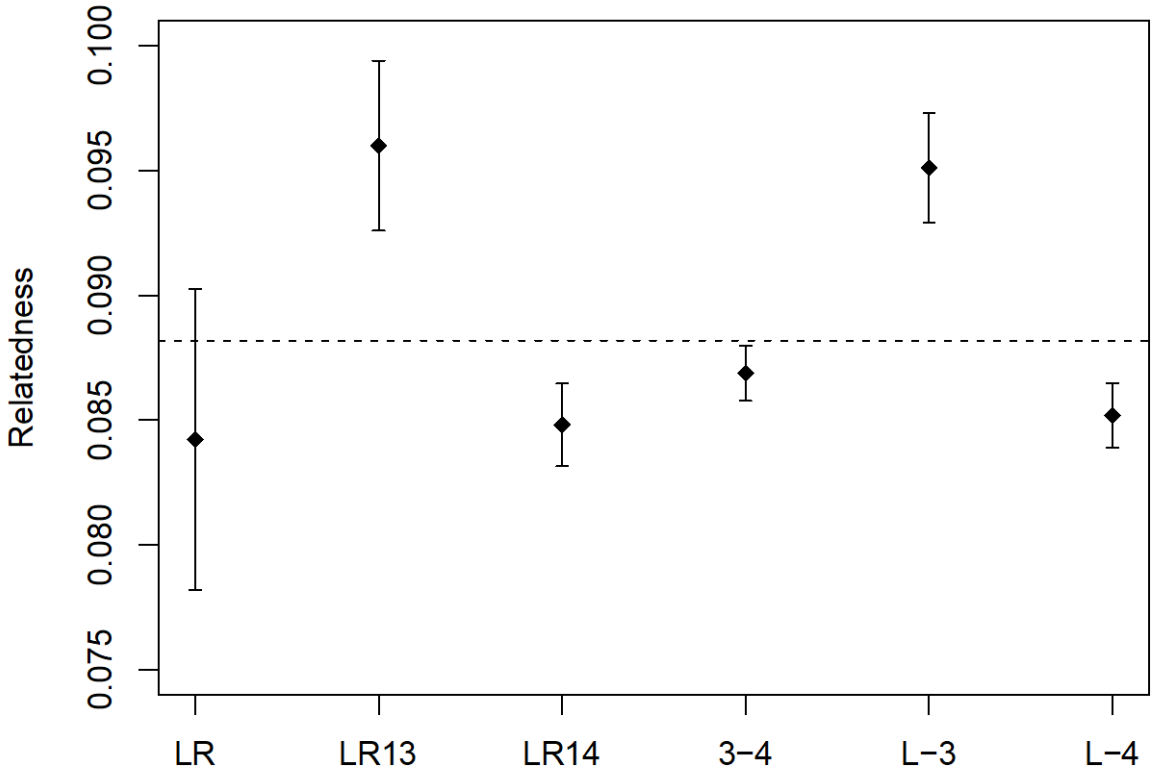
North EW, Schlag Z, Hood RR, Li M, Zhong L, Gross T, Kennedy VS (2008) Vertical swimming behavior influences the dispersal of simulated oyster larvae in a coupled particle-tracking and hydrodynamic model of Chesapeake Bay. *Marine Ecology Progress Series* 359:99–115

Sisson GM, Shen J, Kim SC, Boon JD, Kuo AY (1997) *VIMS three-dimensional hydrodynamic-eutrophication model (HEM-3D): application of the hydrodynamic model to the York River system* (Special Report in Applied Marine Science and Ocean Engineering No. 341)

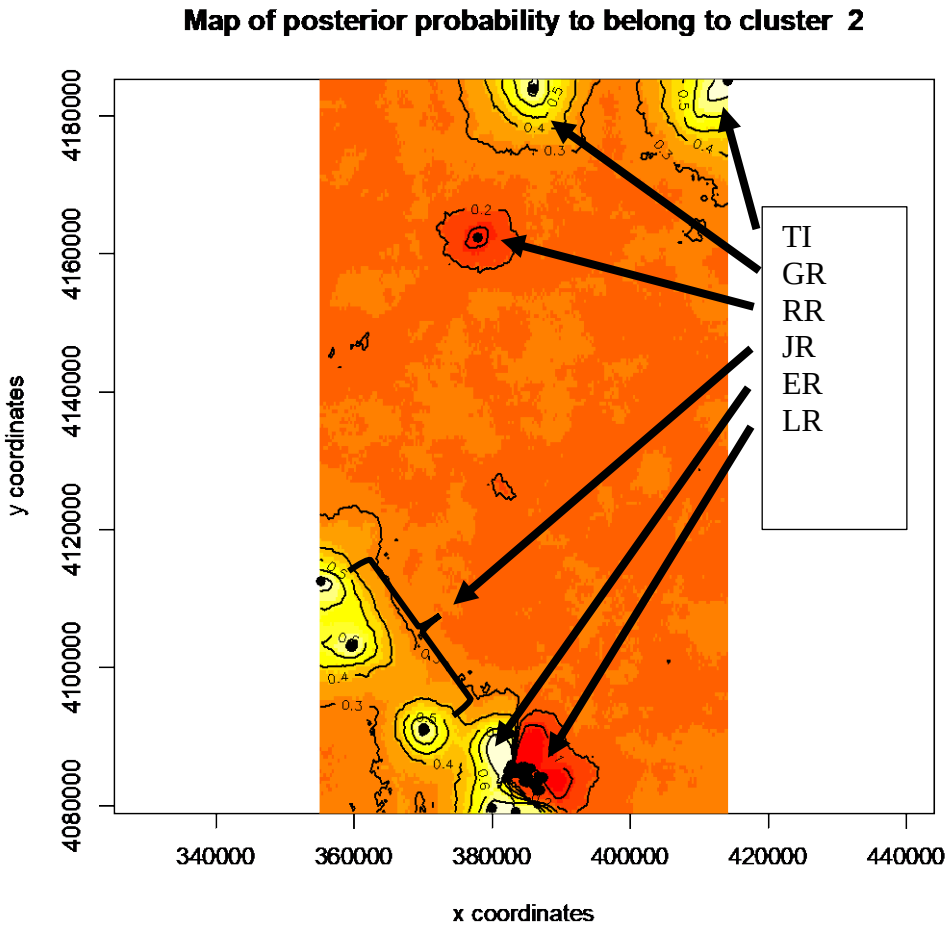
Sisson GM, Shen J (2012) *Modeling of Oyster Larval Connectivity for CBF in Support of NOAA ' S Community-Based Restoration Program & Restore America ' s Estuaries Oyster and Reef Balls on Sanctuary Reefs in MD and VA - Phase Three* (Special Report No. 433 In Applied Marine Science and Ocean Engineering)

Weir BS, Cockerham, CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38(6):1358–1370

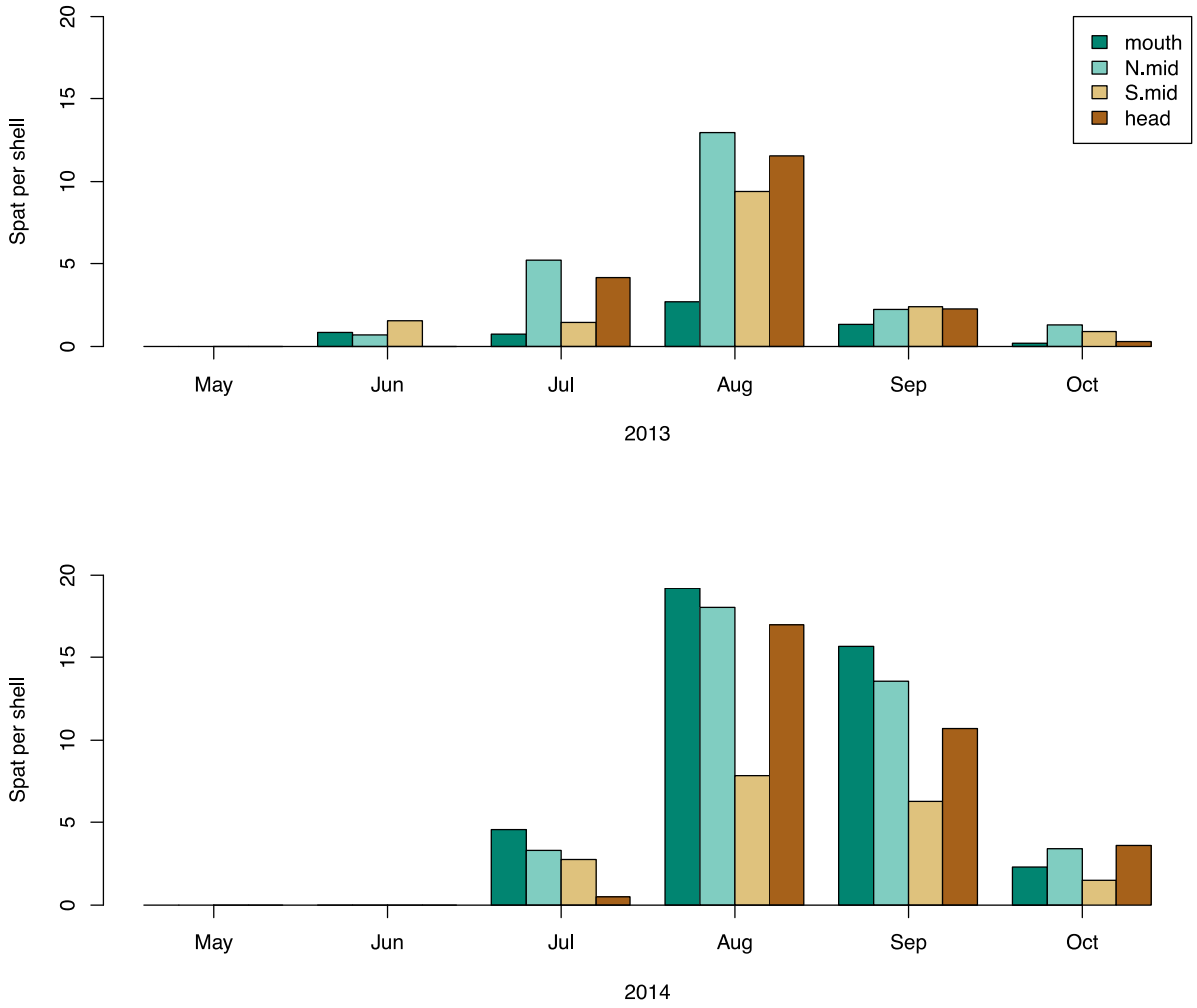
Supplementary Figures



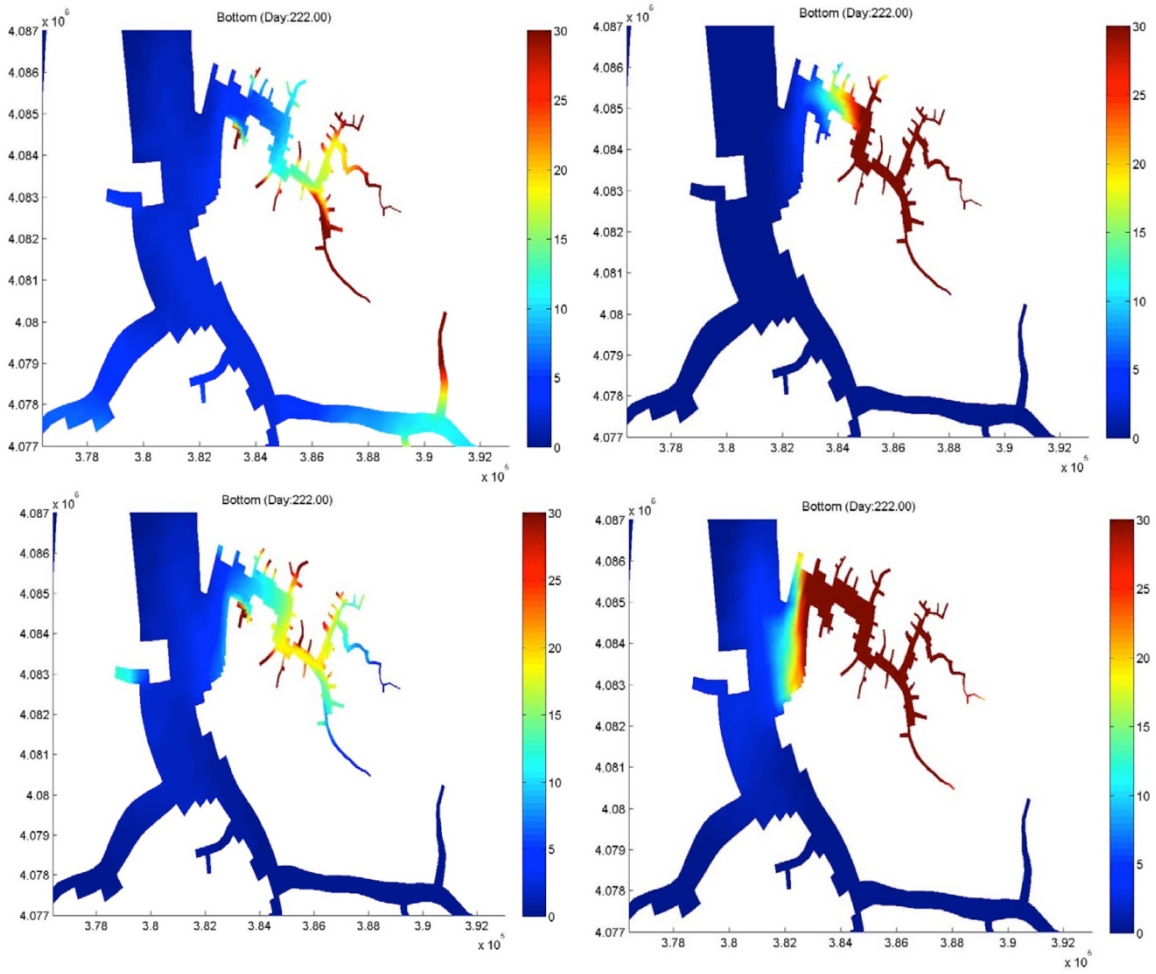
Supplementary Figure (S1): Mean relatedness values estimated using the Wang (2002) method in related (Pew et al. 2015) implemented in R. Oysters from the Lafayette River (LR) reference sample, 2013 spat (LR13), and 2014 spat (LR14) were compared. Also, values for pooled samples LR13 and LR14 (3-4), LR and LR13 (L-3), and LR and LR14 (L-4) were estimated. The whiskers are 95% confidence intervals and the dotted line in the mean value estimated for all samples pooled together.



Supplementary Figure (S4): Posterior GENELAND assignment probabilities. The sampling locations were TI – Tangier Island; GR – Great Wicomico River; RR – Rappahannock River; ER – Elizabeth River; LR – Lafayette River. This figure displays the separate assignment of RR from TI due to in the inclusion of spatial data.



Supplementary Figure (S3): Mean spat per shell from shell-string survey results for 2013 and 2014. Data were used in ANOVA to test significant interactions between locations, years, and months.



Supplementary Figure (S4). Larval oyster connectivity model output for the Lafayette and Elizabeth Rivers. The left column contains results for the larval release location within the Lafayette nearest to the mouth and the right column contain results for the larval release location within the Lafayette nearest to the head. Top row are results without runoff and bottom row are results with runoff. The color scale is related to oyster larvae density with red as the highest and blue as the lowest. Image modified from Sisson and Shen 2012.

Supplementary tables.

Supplementary Table (S1) Global statistics: Observed heterozygosity (Ho); within population gene diversity or expected heterozygosity (He); individual inbreeding coefficient relative to total population (Fit); fixation index (Fst); individual inbreeding coefficient relative to subpopulation(Fis); and minor allele frequency; F-statistics and HWE exact p-values calculated in R package pegas v. 0.9 (Paradis, 2010); corrected p-value = 0.01

SNP name	Ho	He	Fit	Fst	Fis	MAF	HWE p.exact
arp_133	0.16	0.16	0.01	0.04	-0.04	0.09	1.000
ba_83	0.23	0.44	0.49	0.01	0.48	0.33	**0.000**
bt_558	0.10	0.10	0.03	0.01	0.02	0.05	1.000
bty_288	0.32	0.42	0.25	0.06	0.20	0.31	**0.000**
cm4_346	0.47	0.50	0.06	0.00	0.06	0.45	0.175
dap_330	0.40	0.43	0.07	0.05	0.02	0.30	0.141
fab_160	0.27	0.28	0.06	0.09	-0.03	0.18	0.282
fad_185	0.54	0.49	-0.09	0.01	-0.11	0.42	0.015
glup_354	0.13	0.13	-0.02	0.01	-0.02	0.07	0.754
hsp27_122	0.19	0.33	0.46	0.08	0.42	0.20	**0.000**
hsp6_205	0.15	0.14	0.09	0.24	-0.20	0.07	0.791
hsp70_237	0.30	0.28	-0.04	0.01	-0.05	0.18	0.181
hsp70_450	0.02	0.02	-0.01	0.03	-0.04	0.01	1.000
idh_313	0.41	0.43	0.05	0.06	-0.01	0.31	0.385
ldl_185	0.50	0.50	0.00	0.01	-0.01	0.46	1.000
mac_449	0.19	0.19	0.00	0.01	-0.01	0.10	0.409
mt_iii_465	0.01	0.02	0.61	0.01	0.61	0.01	**0.000**
myc_80	0.47	0.50	0.09	0.05	0.04	0.40	0.107
myc1_194	0.19	0.19	0.00	0.01	0.00	0.10	1.000
myc7_373	0.43	0.46	0.15	0.10	0.06	0.34	0.128
mych_289	0.00	0.02	1.00	0.00	1.00	0.01	**0.000**
not1_322	0.52	0.50	0.04	0.10	-0.06	0.43	0.294
nss_417	0.27	0.23	-0.10	0.35	-0.69	0.17	**0.000**
nss1_228	0.26	0.49	0.49	0.00	0.49	0.43	**0.000**
nss2_198	0.48	0.48	0.02	0.01	0.00	0.38	0.930
pl_514	0.38	0.41	0.17	0.13	0.05	0.27	0.120
prp_198	0.00	0.00	0.92	0.01	0.92	0.00	**0.005**
pt_473	0.50	0.49	0.00	0.03	-0.02	0.42	0.938
rpl_176	0.31	0.32	0.05	0.02	0.03	0.20	0.430
rpl13a_183	0.16	0.24	0.40	0.03	0.38	0.13	**0.000**
rpl19_537	0.30	0.33	0.10	0.03	0.07	0.21	0.018
rpl7_234	0.39	0.38	-0.01	0.04	-0.05	0.25	0.548
rpl9_451	0.34	0.28	-0.16	0.25	-0.55	0.19	**0.000**
rpo_422	0.27	0.45	0.41	0.02	0.40	0.33	**0.000**
rpp2_171	0.51	0.46	-0.10	0.01	-0.11	0.36	**0.006**
rps15_301	0.06	0.06	0.00	0.01	-0.02	0.03	1.000
rps23_327	0.34	0.38	0.16	0.05	0.11	0.26	**0.002**
stp_470	0.28	0.28	0.02	0.01	0.01	0.16	0.888
tf_393	0.40	0.39	-0.02	0.00	-0.02	0.25	0.704
unk_399	0.13	0.16	0.20	0.01	0.19	0.09	**0.000**
upp_263	0.28	0.30	0.07	0.01	0.07	0.19	0.096

Supplementary Table (S2): Linkage analysis results: rd values from index of association calculation in Poppr ver.2.4.1 implemented in R ver.3.4.0; significant p-values are in bold and underlined; corrected p-value = 0.007

Sample	rd	p-value
LR	0.008	**0.002**
LR13	0.003	0.035
LR14	0.001	0.274
ER	0.001	0.336
JR	0.004	0.097
RR	0.004	0.107
GR	0.006	0.377
TI	0.017	**0.001**
NEH	0.015	**0.001**

Supplementary Table (S3) Results of analysis of variance on shell-string data from the Lafayette River collected in the summer spawning seasons of 2013 and 2014. Analyses performed in R.

Variables	Degrees of freedom	Sum Sq	Mean Sq	F value	p-values
month	1	5161	5161	125.591	< 2e-16 ***
year	1	1957	1957	47.629	9.39e-12 ***
site	1	29	29	0.694	0.41
month:year	1	2310	2310	56.202	1.49e-13 ***
month:site	1	12	12	0.296	0.59
year:site	1	588	588	14.31	0.00017 ***
month:year:site	1	163	163	3.974	0.046 *
Residuals	952	39125	41		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’