LIMNOLOGY
and
OCEANOGRAPHY: METHODS

Limnol. Oceanogr.: Methods 12, 2014, 548–562

Evaluating and improving a semi-automated image analysis technique for identifying bivalve larvae

Jacob D. Goodwin, Elizabeth W. North, and Christine M. Thompson*
University of Maryland Center for Environmental Science, Horn Point Laboratory, Cambridge, MD 21613

Abstract

Knowledge of the distribution, abundance, and transport of bivalve larvae is limited due to their small size, similar morphologies between species, and lack of an automated approach for identification. The objective of this research is to evaluate and improve the accuracy of ShellBi, a novel supervised image classification method that uses birefringence patterns on the shells of bivalve larvae under polarized light to identify species. The performance of the ShellBi method was tested by rearing *Crassostrea virginica* (eastern oyster) larvae at different temperatures (21.3 and 27.5°C) and salinities (10.3, 14.1, 14.4, and 20.5). Differences in rearing temperatures resulted in differences in classification accuracy, as did large variations in salinity (≥10 units). Classification accuracies increased from 67–88% to 97–99% when training sets included images of larvae reared in conditions similar to those of the larvae being classified. Additional tests indicate that misclassification rates ranged from 0 to 13% for false positives and from 0 to 22% for false negatives, depending on the proportion of oyster larvae in the sample. Results suggest that this technique could be applied to field samples with high accuracy as long as the images that are used to make classifications include larvae that were reared in conditions that are similar to those in situ. In addition, these findings demonstrate that the ShellBi method can be used to measure and identify bivalve larvae in a different system than the one for which it was developed, suggesting that the method has broad applicability in marine and estuarine systems.

Understanding dispersal pathways and connectivity is important for effective fisheries management strategies (Fogarty and Botsford 2007). The larval stage of bivalves is the least understood aspect of their life history, but it is important to understand because it is the stage during which dispersal takes place, which in turn, influences population connectivity and gene flow (Kennedy 1996; Pineda et al. 2007; Dame 2012; Munroe et al. 2012). Species identification is important for understanding dispersal and its effect on the population connectivity of bivalves because larvae of different species can exhibit variations in behavior that may result in large divergences in transport (Shanks and Brink 2005; North et al. 2008). However, studies of bivalve larvae are difficult to conduct because of identification challenges, small sizes of individuals, high mortality rates, and spatial patchiness (Boicourt 1988; Garland and Zimmer 2002).

Many identification techniques of bivalve larvae are too time consuming or expensive to apply when conducting sampling on a large scale. Accordingly, specific pros and cons of identification techniques of bivalve larvae are reviewed in Garland and Zimmer (2002), Hendriks et al. (2005), and Thompson et al. (2012a). Identification can involve time-consuming methods that rely on morphological differences (Loosanoff et al. 1966; Chanley and Andrews 1971; Lutz et al. 1982). More rapid molecular techniques include multiplex PCR (Hare et al. 2000), quantitative PCR (Wight et al. 2009), and fluorescent in situ hybridization with DNA probes (Henzler et al. 2010). Although quantitative PCR can provide some insight into the quantity of bivalve larvae, it does not provide information on the sizes of those larvae. Furthermore, these methods can have high costs and limitations on sample volume.

An alternative method for rapid identification is ShellBi. ShellBi can be an accurate, cost-effective, and rapid approach for identifying and measuring bivalve larval shells once the initial effort to prepare this technique for use in a new system is complete. ShellBi is a semi-automated image-processing approach that uses birefringence patterns on the shells of larvae that appear when subjected to polarized light (Tiwari and Gallager 2003a, 2003b; Thompson et al. 2012a). Under polarized light, color and texture-based features are extracted from digital images of the larval shells by pattern recognition software. The algorithm used in this work, a Support Vector

*Corresponding author: E-mail: jgoodwin@umces.edu

Acknowledgments

Full text appears at the end of the article.

DOI 10.4319/lom.2014.12.548

Machine (SVM), generates decision boundaries that maximize differences between labeled categories (training images) and then applies the decision boundaries to classify new observations into those categories. For the ShellBi method, the categories are defined as groups of images of larval shells from known bivalve species (called 'training sets') and the observations are images of shells that need to be identified (called 'unknown sets'). In short, the classifier (the SVM) uses color- and texture-based features from the training set images to identify images of larval shells in the unknown set (Tiwari and Gallager 2003a, 2003b; Thompson et al. 2012a).

Thompson et al. (2012a) validated the ShellBi method with DNA and visual classification methods and improved it showing 98% identification accuracy for four hatchery-reared species *Argopecten irradians* (bay scallop), *Crassostrea virginica* (eastern oyster), *Mercenaria mercenaria* (quahog), and *Mya arenaria* (soft-shell clam). However, the species featured in their hatchery-reared training sets represented a simplified sample relative to field-caught larvae and larvae in situ may have had different growth rates due to environmental heterogeneities (Thompson et al. 2012a). Therefore, although obtained accuracies are high for identifying larvae reared in the hatchery, the effect of different growth conditions on shell formation between larvae reared in the hatchery and in the field may cause drops in accuracy. Therefore, improvements to the ShellBi method are needed when applied to field samples.

The overall objective of this research was to evaluate the use of the ShellBi method for identifying *C. virginica* bivalve larvae in the Choptank River, a tributary of Chesapeake Bay in Maryland, USA. Initially ShellBi was tested using bivalve species native to Cape Cod, Massachusetts, USA, and found in Waquoit Bay (Tiwari and Gallager 2003b; Thompson et al. 2012a). The bivalve species and physical characteristics of the mesohaline Choptank River differ from Waquoit Bay. Salinities near the surface of the Choptank River during the spawning season of oysters (May-October) are 0 to 14 and temperatures range from 17°C to 27°C (MDNR 2012). In contrast, Waquoit Bay water temperatures during May-October are 13°C to 26°C and salinities range from 28 to 32 (Thompson et al. 2012b). In addition to the overall objective of testing the ShellBi technique in a different system, the three specific objectives that guided this research were to 1) determine the influence of growth conditions on classification accuracy, 2) evaluate the influence of training set composition on classification accuracy, and 3) estimate misclassification rates of this method when applied to distinguish *C. virginica* larvae from other bivalve species found in the Choptank River.

Materials and procedures

Six bivalve species that are found in the Choptank River were spawned, their larvae were reared, and images of their shells were used to create training sets (Fig. 1). In addition, *C. virginica* larvae were reared in different growth conditions and imaged. A series of classification tests were conducted with the training sets and *C. virginica* images. Methods for spawning, rearing, imaging, and classifying larvae are described in this section.

Spawning and rearing bivalve larvae from the Choptank River

Six species of bivalve larvae were reared to obtain images for training sets: *C. virginica* (the target organism) and five other species that are abundant in the plankton along the mesohaline portion of Chesapeake Bay (Table 1). Adult specimens of the five species, *Ischadium recurvum* (hooked mussel), *Mulinia lateralis* (dwarf surf clam), *Mytilopsis leucophaeata* (dark false mussel), *Rangia cuneata* (Atlantic rangia), and *Tagelus plebeius* (razor clam) were collected from Choptank River field sites and brought to lab for spawning in 2009, 2010, 2011, and 2012. Some specimens of *M. lateralis* also were collected from the Corsica River (a tributary of Chesapeake Bay that is north of the Choptank River). Temperature fluctuation and strip spawning techniques were used to induce spawning (Chanley 1970; Kennedy et al. 1989). Larvae were raised at room temperature 23.0 ± 0.5°C (n = 30) (here and henceforth numbers after '±' are the standard deviation) and fed fresh *Isochrysis galbana* and *Thalassiosira pseodonana* (for D-stage and veliger larvae) and *Tetraselmis chui* (for pediveliger larvae). A subset of larvae was preserved in 80% ethanol buffered with sodium borate every two days from prodissoconch 1 through pediveliger stages so that different age/size classes for each species could be incorporated into training sets. The fixative was buffered to a target pH of 8.0 to inhibit dissolution of larval shells (Thompson unpubl. data).

In 2009, 2010, and 2011, multiple ages of *C. virginica* larvae (2-, 4-, 6-, 8-, 10-, 12-, 14- and 16-d old) were obtained from the Horn Point Oyster Hatchery where they had been reared at an average temperature of 25.9 ± 1.5°C (n = 30) and average salinity of 10.3 ± 0.9 (n = 30). These hatchery-reared *C. virginica* larvae were fed *Isochrysis galbana* and *Thalassiosira pseodonana* as D-stage larvae. For veliger stages, *Chaetoceros mulleri* was added. Pediveligers were fed *Tetraselmis chui* plus *Chaetoceros mulleri*. Algal concentrations averaged 5.7 × 10^4 cells mL^{-1} over the duration of the larval stages for hatchery-reared larvae. Larvae of *C. virginica* from 2009 were preserved in 80% ethanol buffered with sodium borate (Thompson et al. 2012a); larvae from 2010 and 2011 were preserved in 4% formalin buffered with sodium borate because larval shells stored in buffered ethanol began to crack after 2 years (Thompson and Goodwin unpubl. data). The preservative used to store larvae (formalin versus ethanol) did not interfere with the ability of ShellBi to classify bivalve larvae (Table 2).

In 2011, 1-d old D-stage *C. virginica* larvae were obtained from the hatchery and were reared at a mean temperature of 22.3 ± 0.4°C (n = 30) and mean salinity of 11.5 ± 0.3 (n = 30). Larvae were fed live cultures of *Isochrysis galbana* and *Thalassiosira pseodonana* (fed to D-stage and veliger larvae) and *Tetraselmis chui* (fed to pediveliger larvae) at an average concentration of 5.7 × 10^4 cells mL^{-1}. Subsets of larvae were preserved in 4% formalin buffered with sodium borate every 2 days up to day 20.

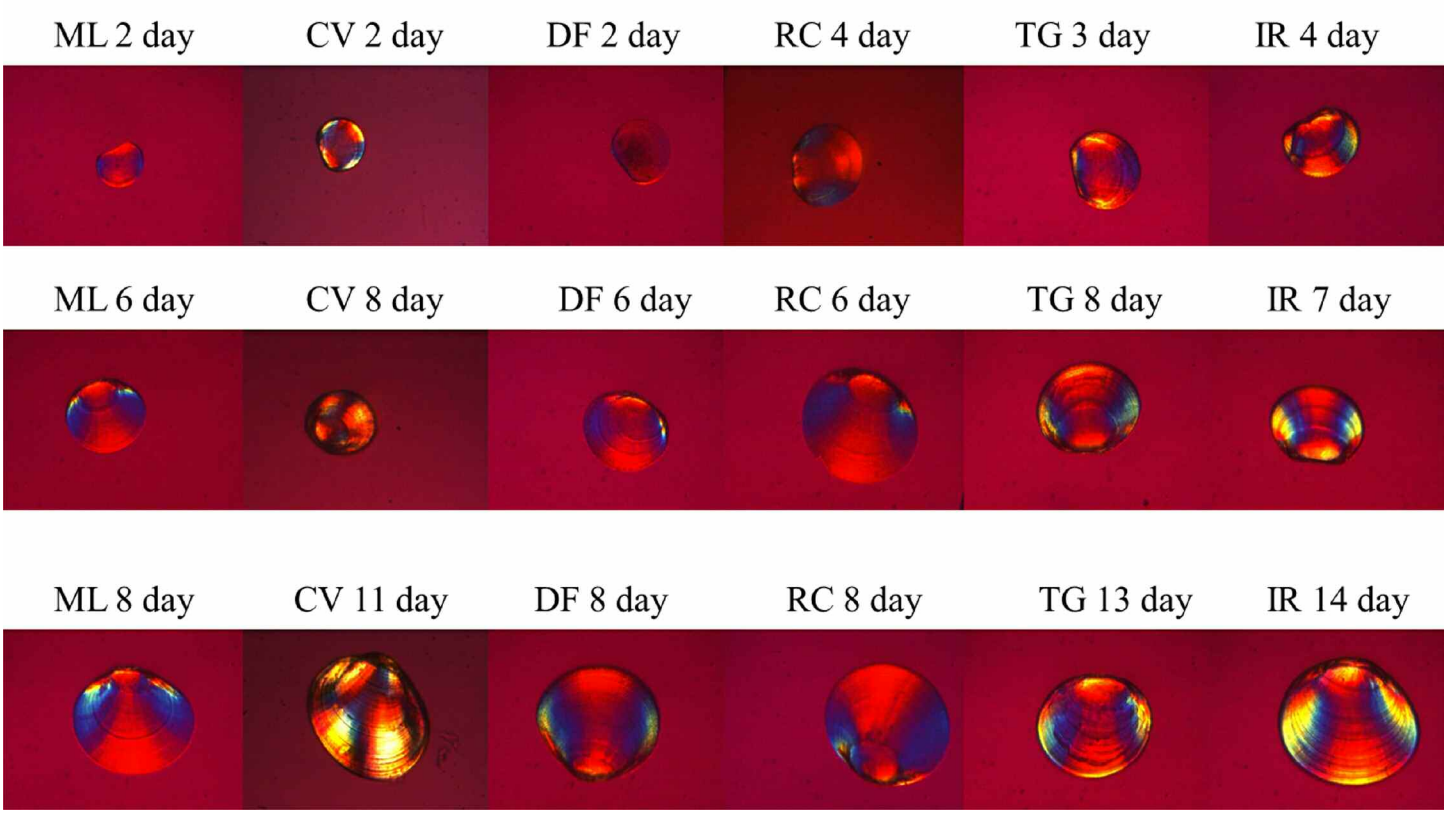

Fig. 1. Images under polarized light of the shells of six species of bivalve larvae used in the analysis ranging from early-stage veliger (top row, 2-4 d old) to late stage veliger (bottom row, 8-14 d old). Species pictured are *Mulinia lateralis* (ML), *Crassostrea virginica* (CV), *Mytilopsis leucophaeata* (DF), *Rangia cuneata* (RC), *Tagelus plebeius* (TG), and *Ischadium recurvum* (IR). Sizes of larvae range from 72-88 μm (top row), 95-155 μm (middle row), and 157-246 μm (bottom row).

Table 1. Spawning conditions for six species of bivalves that are found in the mesohaline region of the Choptank River.

Scientific name	Temperature	Salinity	Season
Ischadium recurvum	25-30°C	20	June-Nov
	(Chanley 1970)	(Chanley 1970)	(Chanley 1970)
Rangia cuneata	30°C	<15	late spring to early fall
	(Sundberg and Kennedy 1992)	(Sundberg and Kennedy 1992)	(Sundberg and Kennedy 1993)
Mytilopsis leucophaeata	30°C	0.5-18	Summer to fall
	(Kennedy 2011b)	(Kennedy 2011b)	(Kennedy 2011a)
Tagelus plebeius	30-32°C	10-30	June-Nov
	(Chanley and Castagna 1971)	(Chanley and Castagna 1971)	(Chanley and Castagna 1971)
Mulinia lateralis	28-30°C	20-30	May-Oct
	(Calabrese and Rhodes 1974)	(Calabrese and Rhodes 1974)	(Calabrese 1969)
Crassostrea virginica	28-30°C	12-27	Summer to fall
	(Kennedy 1996)	(Kennedy 1996)	(Kennedy 1996)

Rearing *C. virginica* larvae in different growth conditions

Larvae of *C. virginica* were reared at different temperatures, salinities, and food concentrations (parameters known to affect growth [Kennedy 1996]) to investigate how different growth conditions affect the classification accuracy of the ShellBi method.

Newly spawned *C. virginica* were obtained from Horn Point Oyster Hatchery and placed in 3-L glass rearing chambers within two temperature-controlled rooms. Water was collected from three sites within the Choptank River system (Tred Avon River, Harris Creek, and Choptank River at the Horn Point dock), and an external site (Chincoteague Bay) on the eastern shore of Maryland. Water was filtered to 1 μm in the field using a battery-operated pump (JABSCO model 50840-0012) and polypropylene cartridge system. Before rearing the larvae,

Table 2. Results of classification tests designed to determine if fixative type (ethanol versus formalin) influenced the classification accuracy of the ShellBi method. All fixatives for training sets and 'unknowns' were buffered with sodium borate. Training sets were composed of 250 images of the following species: *Crassostrea virginica*, *Ischadium recurvum*, *Mytilopsis leucophaeata*, and *Rangia cuneata.* Images of larvae in the training sets that were stored in either ethanol or formalin were used to classify images of *M. leucophaeata* that had been stored in either ethanol or formalin. Treatments denoted "ethanol & formalin" are composed of 100 images of *M. leucophaeata* that were stored in ethanol and 100 images of *M. leucophaeata* that were stored in formalin. The *M. leucophaeata* larvae were taken from the same cohort and stored in formalin or ethanol for an equal amount of time (11 months). All training sets had classification accuracies > 95%. Slightly lower accuracies were reported when training sets included images of shells stored in formalin (95% to 96%) compared with those stored in ethanol (97% to 98%). Based on the high classification accuracies for shells stored in both types of fixatives, it is concluded that the fixative used does not interfere with the ability of ShellBi to classify larvae.

Test Number	Fixative of training set	Fixative of unknown set	Percent classification accuracy
1	ethanol	ethanol	98.1
2	ethanol	formalin	95.2
3	ethanol	ethanol & formalin	97.3
4	formalin	ethanol	98.3
5	formalin	formalin	95.8
6	formalin	ethanol & formalin	96.1
7	ethanol & formalin	ethanol	97.1
8	ethanol & formalin	formalin	94.9
9	ethanol & formalin	ethanol & formalin	96.7

salinity was adjusted to provide a range of salinities that reflect conditions in situ in Chesapeake Bay. Salinity of the water collected at the Horn Point dock was raised to 10.3 and waters from the Tred Avon and Harris Creek were raised to 14.1 and 14.4, respectively, using Crystal Sea Marinemix (Marine Enterprises). The salinity of the Chincoteague Bay water was lowered to 20.5 using deionized (DI) water. Before starting this experiment, the water was filtered to 1 μm a second time.

The average water temperatures in the temperature-controlled rooms were 21.3 ± 1.0°C (n = 48) and 27.5 ± 0.6°C (n = 67). Each room contained 8 rearing chambers that held four salinity treatments (10.3 ± 0.7 [n = 58], 14.1 ± 0.7 [n = 63], 14.4 ± 0.6 [n = 53], and 20.5 ± 1.0 [n = 44]) using two chambers and two levels of food concentrations (high and low) within each salinity treatment. The concentration of algae fed to the larvae was based on the concentration of larvae in the containers (Helm et al. 2004), with low food treatments fed half the concentrations of the high food treatments. The ratio of larvae to algae in the high food treatments was on average 1:1.6 × 10^4, with the objective that the larvae would be fed to satiation. The average concentration of algae in the high and low food treatments were 9.2 × 10^4 cells mL^{-1} and 7.9 × 10^3 cells mL^{-1}, respectively. Algae were obtained from the Horn Point Oyster Hatchery and were composed of live cultures of *Isochrysis galbana* and *Thalassiosira pseodonana* (fed to D-stage and veliger larvae) and *Tetraselmis chui* (fed to pediveliger larvae). Subsets of larvae were preserved in 4% formalin buffered with sodium borate every 2, 4, 6, 8, 12, and 14 days in the warm chambers. In the cool conditions larvae took longer to develop to the pediveliger stage and were preserved every 2 days up to day 20.

Image acquisition for training and unknown sets

Images of all larval shells were taken by an Infinity 2.3C digital 8 megapixel camera mounted on a custom-built compound microscope fitted with a polarization filter and full wave compensation plate (λ). Larvae were first soaked in 40% bleach and 60% DI water buffered with sodium borate (hereafter referred to as buffered DI water) for a period of 15 min to remove tissue and break apart the valves of the shells. The larval shells were then sieved and rinsed with buffered DI water onto a Sedgewick Rafter slide. Digital images of individual shells were taken under 50× magnification at a resolution of 96 dpi. The microscope stage was moved manually or with a joystick attached to an automated stage to image one shell after another. Images were captured with shells at random orientations. A 12V 100W incandescent microscope bulb was used as a light source. Lumenera Analyze software (version 5.0.3 Lumenera Corporation) was used in conjunction with the digital camera to capture JPEG images. Settings on the software were adjusted so that they matched background color and cross polarization pattern as suggested in Thompson et al. (2012a) and kept constant between images. Major background color differences occurred throughout the day when a metal bracket was used for the full wave compensation plate which was near the light source of the microscope. Because these differences affected classification accuracies (results not shown), a plastic housing was used for the wave compensation plate to prevent background color drift.

To create a species category within a training set, 250 images of individual shells were selected for each species so that the images spanned the range of stages and sizes of the

larvae (prodissoconch-1 through pediveliger). Thompson et al. (2012a) found that at least 200 images should be used in a training set. Training sets were composed of different numbers of species. For example, a 6-species training set included 250 images of *C. virginica, I. recurvum*, *M. lateralis*, *M. leucophaeata*, *R. cuneata*, and *T. plebeius* for a total of 1500 images. All training sets were balanced: each species category had an equal quantity of images (250) with similar age representations of bivalve larvae.

Images of *C. virginica* shells from the experiment were used as unknown sets. The same imaging procedures that were used for the training sets were also used for *C. virginica* larvae reared in the growth experiment. There were 3288 images of larvae captured from the experiment. Those images were used to represent warm and cool conditions as well as four different salinity treatments.

Images were preprocessed before classification so that each larval shell, a region of interest (ROI), was defined and distinguished from its background (Thompson et al. 2012a) using MATLAB (version R2009a, Mathworks Inc.) and its image Processing Toolbox (version 6.3, Mathworks Inc.). The preprocessing (i.e., cropping) was performed using an automated ROI masking routine in MATLAB (Thompson et al. 2012a).

Image classification and analysis

Image classification was accomplished by extracting features from training sets, cross validating the training sets, extracting features from unknown images, and using the training features to classify unknown images (Thompson et al. 2012a). All images were processed using the Bivalve Larval Identification (BivLID) software implemented in MATLAB by C. Thompson based on algorithms used in Tiwari and Gallager (2003b) and Thompson et al. (2012a). Training set feature extraction and cross-validation were conducted before the classification of unknown images. The feature extraction process calculated 1104 Gabor texture features and 9 color-angle features for each image. A Principle Component Analysis (PCA) was then conducted using the Gabor texture features and color angles to isolate the 25 Gabor features that encompassed the most variability in the training set and to remove redundancy and noise (Zhao et al. 2010; Thompson et al. 2012a). After extracting and transforming features from the training set and unknown images, a Support Vector Machine (SVM) in BivLID was used for cross-validation and classification (Cawley 2000, http://theoval.cmp.uea.ac.uk/svm/toolbox/).

A leave-one-out cross validation procedure (LOO, Fukunaga and Hummels 1989) was run to assess performance of the training sets. This procedure left out one image from the training set, used features from the remaining images to classify the left-out image, and repeated this for all images to calculate cross validation accuracy for each category. Classification tests were also conducted. To classify an image, the SVM mapped the same features from the unknown image to the decision boundaries created with the training set using a one-to-one approach for each category (Lou et al. 2003). An "other" category was created so unknown images would not be classified as false positives, i.e., forced into a training set category to which they were not closely related (Davis et al. 2004). The output of the program indicates how many unknown images were classified into each training set category and the "other" category.

Larval shells were measured and statistical tests were performed to compare shell heights. To accomplish this, a script was created in MATLAB (version R2009b, Mathworks Inc.) to measure the maximum axis of a masked ROI of a larval shell as a measure of shell height. Nonparametric statistical tests were conducted because shell heights in all treatments were not normally distributed (Shapiro-Wilk, $\alpha = 0.05$, $P < 0.01$). Shell heights of *C. virginica* in the high and low food treatments were paired by salinity and temperature treatments for an even comparison (Sokal and Rohlf 1987). Median shell heights were not significantly different between larvae reared in high (95.9 μm, $n = 177$) and low (91.0 μm, $n = 177$) food treatments (Wilcoxon rank sum = 32750, Z = 1.39, $P < 0.17$, $n = 354$). Therefore images from high and low food treatments were pooled within each salinity and temperature treatment in further analyses. To determine if there was a difference in median shell heights between warm and cool treatments, a Wilcoxon rank sum test was employed with data pooled across salinity treatments. A Kruskal-Wallis one-way analysis of variance by ranks was used to test for differences in median shell heights between salinity treatments. After conducting the Kruskal-Wallis test, intergroup comparisons between salinity treatments were made using Mann-Whitley U tests. A Bonferroni adjustment was used to reduce type I error so that the P value for significance was set to 0.008 (Bland and Altman 1995). The number of larvae reared in warm and cool conditions was similarly represented across salinity treatments, and therefore, did not bias larval growth across salinity treatments for these tests. All statistical tests were performed using MATLAB (version R2012a, Mathworks Inc.).

Assessment

Tests were conducted to evaluate the influence of growth conditions on the classification accuracy of the ShellBi method, to determine the influence of training set composition on classification accuracy, and to estimate misclassification rates. A leave-one-out (Fukunaga and Hummels 1989) cross validation resulted in high cross validation classification accuracies (>90.8%) for all training sets except for a 6-species training set (74.7%) (Table 3).

The influence of growth conditions on classification accuracy

The effect of temperature on classification accuracy of a hatchery composed training set was tested using two training sets that contained *C. virginica* reared in warm conditions (a 3-species training set composed of 250 images each of *C. virginica*, *M. lateralis*, and *R. cuneata* and a 4-species training set

Table 3. Leave one-out (LOO) cross-validation accuracy of training sets for classifying *C. virginica*. The first column lists the analysis in which the training set was applied. The second column gives the two letter code of each species used in the training set (CV: *Crassostrea virginica*, RC: *Rangia cuneata*, ML: *Mulinia lateralis*, TG: *Tagelus plebeius*, IR: *Ischadium recurvum*, and DF: *Mytilopsis leucophaeata*). The third column lists the number of images in each training set. The fourth column gives the LOO percent accuracy for classifying *C. virginica.*

Analysis	Training set	Number of images	Percent cross validation accuracy
Temperature			
26.4 (hatchery 3-species)	CV, RC, ML	750	98.1
26.4 (hatchery 4-species)	CV, RC, ML, TG	1000	98.9
22.3 (cool Exp)	CV, RC, ML	750	97.1
Salinity			
10.3	CV, IR, RC	750	98.8
14.1	CV, IR, RC	750	99.6
14.3	CV, IR, RC	750	99.6
20.5	CV, IR, RC	750	99.2
Variation in growth conditions			
RC, ML, CV-2009	CV, RC, ML	750	99.6
RC, ML, CV-2009-2010	CV, RC, ML	750	98.5
RC, ML, CV-2009-2010-2011	CV, RC, ML	750	98.1
RC,ML,CV-2009-2010-2011-exp	CV, RC, ML	750	96.7
Larval stage			
Veliger	CV, ML, TG	750	99.0
D-stage	CV, ML, TG	750	96.4
Training set composition			
3-species	CV, RC, TG	750	95.6
3-species	CV, IR, RC	750	95.7
3-species	CV, ML, IR	750	92.4
3-species	CV, DF, RC	750	98.8
3-species	CV, DF, IR	750	95.6
3-species	CV, DF, TG	750	97.2
3-species	CV, IR, TG	750	94.8
3-species	CV, TG, ML	750	95.2
3-species	CV, ML, RC	750	98.1
4-species	CV, RC, TG, ML	1000	92.4
4-species	CV, RC, IR, ML	1000	92.4
4-species	CV, RC, IR, DF	1000	94.0
4-species	CV, DF, TG, IR	1000	94.0
4-species	CV, RC, DF, ML	1000	97.2
4-species	CV, DF, TG, ML	1000	94.0
4-species	CV, RC, TG, IR	1000	94.0
5-species	CV, RC, IR, TG, ML	1250	90.8
5-species	CV, RC, IR, TG, DF	1250	92.8
5-species	CV, RC, IR, ML, DF	1250	91.6
5-species	CV, RC, TG, ML, DF	1250	95.6
5-species	CV, IR, TG, ML, DF	1250	90.8
6-species	CV, RC, IR, TG, ML, DF	1500	74.7
6-species	CV*, RC, IR, TG, ML, DF	1500	92.1
3-category order-based (clams, oysters, mussels)	(CV), (IR), (RC, ML, DF,TG)	750	90.7
3-category order-based (clams, oysters, mussels)	(CV*), (IR), (RC, ML, DF,TG)	750	98.9

*Denotes that images of *C. virginica* larvae grown in different temperature and salinity treatments were added to the *C. virginica* training set category (Table 5).

that also included 250 images of *T. plebeius*). For both training sets, *C. virginica* larvae were reared in the hatchery at an average temperature of 25.9 ± 1.5°C (n = 30). The other species were reared in our laboratory at room temperature 23.0 ± 0.5°C (n = 30). The training sets contained images of larvae at similar age ranges (2-14 days old).

The 3- and 4-species training sets were used to conduct four classification tests in which the training sets remained the same and the "unknown" images of *C. virginica* shells from the experiment were varied. The two test sets were comprised of images of larvae reared in 1) the warm (27.5 ± 1.0°C, n = 67) treatment and 2) the cool (21.3 ± 1.0°C, n = 48) treatment. Each of these unknown sets included images of larval shells grown at all salinity levels and age ranges between 2-20 days old. The temperatures at which larvae were reared significantly influenced growth of the two treatments: larvae reared in cooler treatments had shorter median shell heights (77.0 μm, n = 365) than those reared in warm conditions (88.8 μm, n = 365) (Wilcoxon rank sum: 97903, Z = –12.7, P < 0.01, n = 730). The median shell height of larvae from the warm treatment was shorter, but not significantly, than the median shell height of the hatchery-reared *C. virginica* larvae in the training sets (114 μm, n = 916) (Wilcoxon rank sum: 107222, Z = -0.88, P = 0.39). On average, the accuracy of ShellBi for identifying *C. virginica* reared in the warm treatment was ~ 20% higher than the accuracy for identifying *C. virginica* reared in the cool treatment using 3-species and 4-species training sets (Fig. 2). In other words, the classification accuracy for *C. virginica* was highest when the temperature at which larvae in the unknown set were reared was similar to that of the training sets.

An additional analysis was conducted to test the effect of rearing temperature on classification accuracy using another training set composed of larvae reared in cool conditions. In this case, the training set was composed 250 images of each species reared in similar cool temperature conditions, *C. virginica* (22.3 ± 1.2°C, n = 58), and *Rangia cuneata* and *Mulinea lateralis* (23.0 ± 0.5°C, n = 30). This training set was used to classify *C. virginica* larvae from two treatments 1) warm (27.5°C, n = 1624) and 2) cool (21.3°C, n = 1664). The accuracy for identifying larvae from the cool treatment was 25% higher (91.0%) than the classification accuracy for larvae from the warm treatment (66.0%) (Fig. 2). Because shell heights differed between larvae grown in warm and cool conditions and because of the strong influence of temperature on classification accuracies, it is concluded that differences in temperature-dependent growth conditions between training sets and unknown sets influence the classification accuracy of the ShellBi method.

In addition to temperature, the effect of salinity on classification accuracy was tested using 3-species training sets composed of *C. virginica*, *R. cuneata*, and *I. recurvum*. The *C. virginica* used in the training sets and for the unknown sets were reared in the experiment at four salinities (10.3, 14.1, 14.4, and 20.5) and were pooled across temperatures. The images of *C. virginica* reared at the four salinities were used to create four different 3-species training sets. In addition to 250 images of *C. virginica*, each training set also had 250 images of *R. cuneata* and *I. recurvum* (reared in a salinity of 11.3). Each of the four training sets were then used to classify four unknown sets of 250 different *C. virginica* images from each of the three other salinity treatments. For example, the training set with *C. virginica* larvae raised in salinity of 10.3 was used to classify larvae from the three other treatments (14.1, 14.4, and 20.5). A total of 12 tests were conducted. High classification accuracies (>95%) occurred when training sets with larvae from low salinity treatments (10.3, 14.1, and 14.4) were used to identify "unknown" *C. virginica* larvae reared in the same low salinity treatments (Fig. 3). Accuracy dropped by 10% when these training sets were used to classify larvae raised in the higher salinity treatment (20.5) (Fig. 3). Training sets with larvae raised in the high salinity treatment (20.5) classified "unknown" larvae from the three lower salinity treatments with > 95% accuracy.

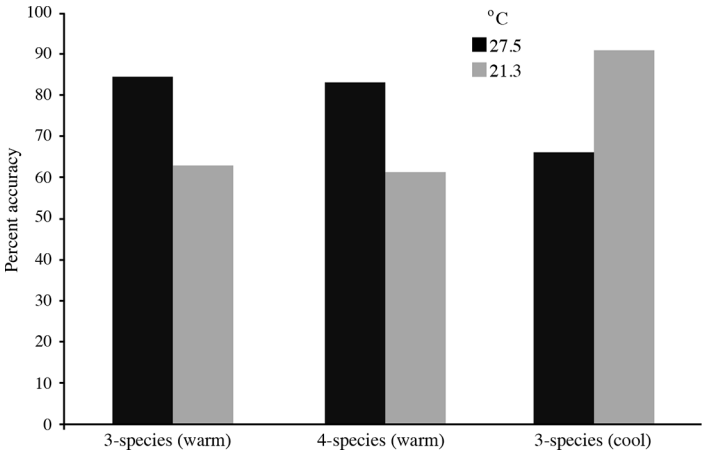

Fig. 2. Classification accuracy for *C. virginica* using two 3-species training sets (*C. virginica*, *M. lateralis*, and *R. cuneata)* and one 4-species training set (*C. virginica*, *M. lateralis*, *R. cuneata*, and *T. plebeius*). Images of shells of *C. virginica* were reared at 25.9°C for 'warm' training sets and at 23.3°C for the 'cool' training set. All three training sets were used to classify shells of *C. virginica* from warm (darker bars) and cool (lighter bars) treatments.

Median shell heights in treatments (n = 250 for each treatment) with salinities of 10.3, 14.1, 14.4, and 20.5 were 76.1 μm, 80.0 μm, 83.9 μm, and 98.3 μm, respectively. Shell heights were significantly different between the four treatments (Kruskal-Wallis test, df = 999, P < 0.01). Post-hoc pairwise comparisons were made using Mann-Whitney U tests. Salinity treatments were significantly different (P < 0.008, df = 499), except for salinity treatments 14.1 and 14.4 (P = 0.13, df = 499). Based on this and the results of the classification tests above, it is concluded that large (10 unit) differences in salinity-dependent growth conditions between training sets and unknown sets influence the classification accuracy of ShellBi.

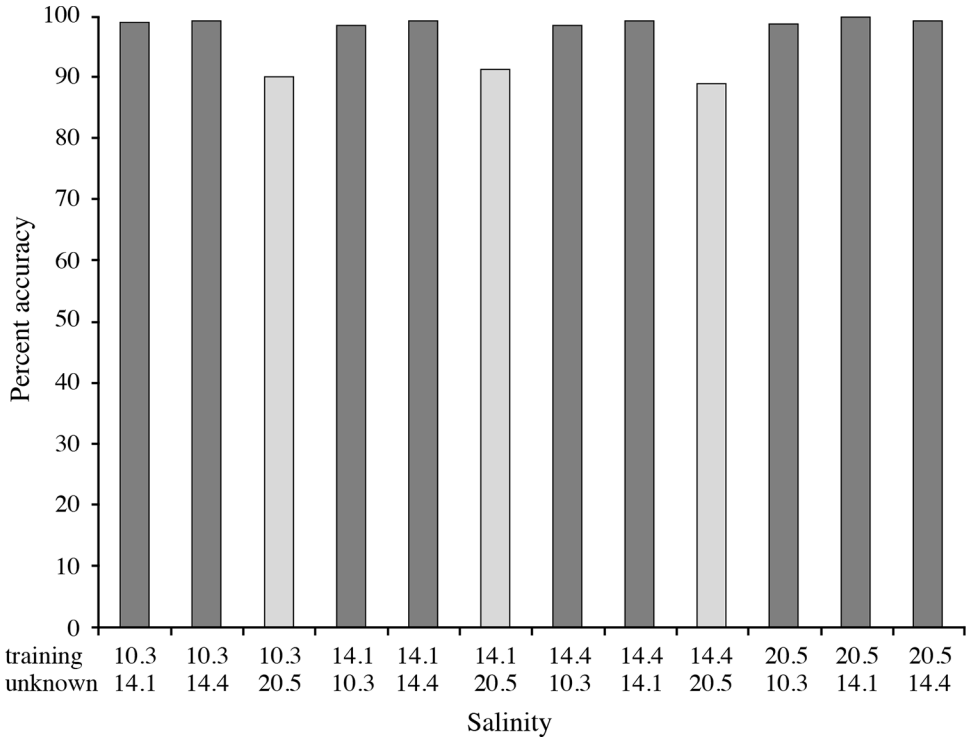

Fig. 3. Classification accuracies for shells of "unknown" *C. virginica* larvae raised in four different salinities (10.3, 14.1, 14.4, and 20.5) when classified with training sets composed of *R. cuneata*, *I. recurvum*, and *C. virginica* larvae, the latter of which were raised in the same four salinities. Numbers under each bar represent the salinity at which *C. virginica* were reared in the training set (upper number) and in the unknown set (lower number). Lighter bars indicate training sets in which larvae were reared at the lower three salinities and used to classify larvae in the high salinity treatment (20.5).

The influence of training set composition on classification accuracy

Three tests were conducted to determine if the composition of images in a training set influenced classification accuracy. (1) The first examined how changing the larval stage (D-stage versus veliger) within the *C. virginica* portion of the training set altered classification accuracy. (2) The second test was designed to identify how the number of categories in a training set influenced classification accuracy. (3) A third test was conducted to determine if increasing variation of growth conditions of larvae in the *C. virginica* portion of the training set affected classification accuracy.

(1) Larval images were broken down into 2 groups (1) D-stage larvae (comprised of larvae between 2-3 days old), and (2) veliger larvae (comprised of larvae between 6-20 days old). Two training sets composed of *C. virginica*, *M. lateralis*, and *T. plebeius* were created. All training sets contained the same images of *M. lateralis* and *T. plebeius*. Images in the *C. virginica* category were varied to form the two training sets that were comprised of (1) images of D-stage larvae raised in the hatchery and (2) images of veliger larvae raised in the hatchery. These training sets were used to classify unknown sets that were comprised of *C. virginica* images of (1) D-stage larvae from the hatchery, (2) D-stage larvae from the experiment, (3) veliger larvae from the hatchery, and (4) veliger larvae from the experiment. Results indicate that training sets containing images of D-stage *C. virginica* larvae classified "unknown" D-stage and "unknown" *C. virginica* veliger images with high accuracies (>98%). Training sets comprised of images of *C. virginica* veliger larvae and used to classify "unknown" D-stage *C. virginica* images had low accuracies (<29%). Based on these results, it is concluded that a training set should contain images of both D-stage and veliger larvae.

(2) Classification tests were conducted using training sets with various numbers of categories and the same set of unknown larvae. Images of *C. virginica*, *I. recurvum*, *T. plebeius*, *R. cuneata*, *M. lateralis*, and *M. leucophaeata* larvae were used to create nine 3-species training sets, seven 4-species training sets, five 5-species training sets, and one 6-species training set. These training sets were used to classify one unknown set comprised of *C. virginica* larvae from the warm and cool treatments of the experiment ($n = 998$). Results comparing the number of categories in a training set indicated that mean accuracies were 82% for 3-species categories ($n = 9$), 75% for 4-species categories ($n = 7$), 70% for 5-species categories ($n = 5$), and 67% for 6-species categories ($n = 1$) (Table!4). When the number of training set categories increased from 3 to 6, the accuracy of ShellBi dropped on average by 17% (Fig.!4). Within the 3-, 4-, and

Table 4. Percent classification accuracy of unknown *C. virginica* larvae from experiments (n = 3288) using training sets with different numbers and compositions of species. Training sets of 3-, 4-, 5-, and 6-species categories were comprised of *C. virginica* (CV), *R. cuneata* (RC), *T. plebeius* (TG), *I. recurvum* (IR), *M. lateralis* (ML), and/or *M. leucophaeata* (DF). 250 images were used for each category.

Training set	Percent classification accuracy	Number of images in training set
CV, RC, TG	69.5	750
CV, RC, IR	72.1	750
CV, ML, IR	82.1	750
CV, DF, RC	72.1	750
CV, DF, IR	99.8	750
CV, DF, TG	96.9	750
CV, IR, TG	97.1	750
CV, TG, ML	79.5	750
CV, ML, RC	71.8	750
CV, RC, TG, ML	66.7	1000
CV, RC, IR, ML	69.1	1000
CV, RC, IR, DF	72.2	1000
CV, DF, TG, IR	96.9	1000
CV, RC, DF, ML	69.2	1000
CV, DF, TG, ML	79.7	1000
CV, RC, TG, IR	69.5	1000
CV, RC, IR, TG, ML	66.6	1250
CV, RC, IR, TG, DF	69.6	1250
CV, RC, IR, ML, DF	69.3	1250
CV, RC, TG, ML, DF	66.8	1250
CV, IR, TG, ML, DF	79.6	1250
CV, RC, IR, TG, ML, DF	66.8	1500
CV*,RC, IR, TG, ML, DF	97.1	1500
order-based: (CV), (IR), (RC, TG, ML, DF)	87.8	750
order-based: (CV*), (IR), (RC, TG, ML, DF)	98.3	750

*Denotes that images of *C. virginica* larvae grown in different temperature and salinity treatments were added to the *C. virginica* training set category (Table 5).

5-species category training sets, classification accuracies varied by as much as 30% depending on which species combinations were used for each training set (Table 4). When the 6 species training set was grouped into a 3-category training set based on taxonomic order [1: Ostreoida, oysters (*C. virginica*), 2: Veneroida, clams (*M. lateralis*, *M. leucophaeata*, *R. cuneata*, *T. plebeius*), 3: Mytiloida, mussels (*I. recurvum*)], classification accuracy improved compared with the 6-species training set, from 66.8% to 87.8%. Therefore the number of categories in a training set and the species composition within them are important factors that affect the classification accuracy of *C. virginica* using the ShellBi approach.

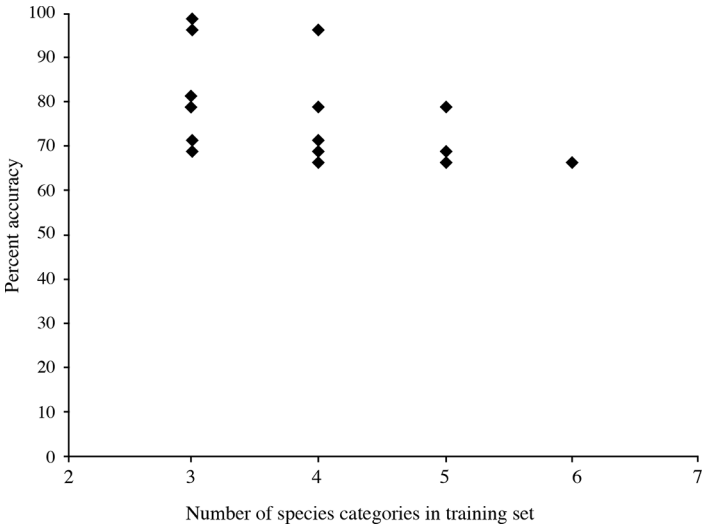

Fig. 4. Percent classification accuracy of ShellBi when classifying images of *C. virginica* shells using training sets with different numbers of species categories (see Table 4 for details). Training sets of 3-, 4-, 5-, and 6-species categories were comprised of hatchery-reared *C. virginica*, and the following species reared in the laboratory: *I. recurvum*, *M. lateralis*, *M. leucophaeata*, *T. plebeius*, and *R. cuneata*. Diamonds represent training sets, each with a different set of species comprising the categories in the training set.

(3) Four training sets composed of *C. virginica*, *M. lateralis*, and *T. plebeius* (250 images for each species) were created. All training sets contained the same (250) images of *M. lateralis* and *T. plebeius*. Images in the *C. virginica* category were varied to form the four different training sets, which were comprised of images of larvae raised: 1) in the hatchery in 2009, 2) in the hatchery in 2009 and 2010, 3) in the hatchery in 2009, 2010, and 2011, and 4) in the hatchery in 2009, 2010, and 2011 and images of *C. virginica* larvae from the warm and cool treatments of the experiment (Tables 3 and 5). The mean temperature and salinity at which the larvae were raised in each training set were 1) 25.4°C ± 1.6 and 10.6 ± 0.4 (n = 30), 2) 26.6°C ± 2.3 and 11.2 ± 0.4 (n = 60), 3) 25.9°C ± 1.1 and 9.1 ± 0.2 (n = 90), and 4) 25.3°C ± 2.3 and 13.2 ± 0.4 (n = 153), respectively. These training sets were used to classify the same unknown set, which was composed of images of *C. virginica* from the warm and cool treatments of the experiment (n = 424). Results indicate that as the variation in growth conditions increased within the *C. virginica* portion of the training set, classification accuracies increased from 76.7% to 98.5% (Table 6). In a second test, a 6-species training set and the 3-category training set based on taxonomic order (Ostreoida, Veneroida, Mytiloida) were used, with some (n = 100) of the *C. virginica* images replaced with those from the warm and cool treatments. These training sets were employed to classify the same unknown set used in the test in the previous experiment, which was composed of other images of *C. virginica* from

Table 5. The number of images of *C. virginica* larvae grown in different temperature and salinity treatments which were added to the *C. virginica* training set category denoted by CV* in Tables 3 and 4. Mean, standard deviation, and sample size for temperature and salinity measurements are reported.

Source	Temperature	Salinity	Number of images
Experimental chamber	27.9 ± 0.7 (*n* = 12)	10.3 ± 0.7 (*n* = 25)	8
Experimental chamber	27.7 ± 0.6 (*n* = 17)	14.1 ± 0.6 (*n* = 32)	27
Experimental chamber	27.5 ± 0.6 (*n* = 15)	14.4 ± 0.7 (*n* = 30)	12
Experimental chamber	27.6 ± 0.6 (*n* = 20)	20.5 ± 1.0 (*n* = 50)	20
Experimental chamber	21.1 ± 1.0 (*n* = 13)	10.3 ± 0.7 (*n* = 25)	19
Experimental chamber	20.9 ± 1.0 (*n* = 15)	14.1 ± 0.6 (*n* = 32)	34
Experimental chamber	21.4 ± 1.0 (*n* = 15)	14.4 ± 0.7 (*n* = 30)	6
Experimental chamber	22.7 ± 1.0 (*n* = 16)	20.5 ± 1.0 (*n* = 50)	19
Hatchery	25.9 ± 1.0 (*n* = 30)	10.3 ± 0.9 (*n* = 30)	105
TOTAL			250

Table 6. Percent classification accuracy using four training sets to identify "unknown" *C. virginica* larvae that were raised in the experiment. The training sets were composed images of *M. lateralis*, *T. plebeius*, and *C. virginica*, the latter of which were varied to incorporate larvae grown in different conditions: 1) in the hatchery in 2009 (CV-2009), 2) in the hatchery in 2009 and 2010 (CV-2009-2010), 3) in the hatchery in 2009, 2010, and 2011 (CV-2009-2010-2011), and 4) in the hatchery in 2009, 2010, and 2011, and in the temperature-controlled experiment (CV-2009-2010-2011-exp).

Training set	Percent accuracy
CV-2009	76.7
CV-2009-2010	76.8
CV-2009-2010-2011	84.7
CV-2009-2010-2011-exp	98.5

the warm and cool treatment of the experiment (*n* = 424). When larvae from the experiment were added to the *C. virginica* portion of the training set, classification accuracy with the 6-category training set improved from 66.8% to 97.1%. Classification accuracies with the 3-category training set were slightly higher than those with the 6-category training set, improving from 87.8% to 98.3% when images of larvae from the experiment were included in the training set. Based on these findings, it is recommended that the images of larvae used to create training sets be representative of the growth conditions of larvae in need of identification, especially in terms of temperature and salinity.

Estimating misclassification rates

Classification tests were performed to determine how well the ShellBi method could identify the target species *C. virginica* given various proportions in a sample. Two training sets were used: a 6-species training set composed of 250 images each of *C. virginica*, *M. lateralis*, *T. plebeius*, *R. cuneata*, *M. leucophaeata*, and *I. recurvum* larvae, and a 3-category order-based training set, using the same 6 species categorized by taxonomic order [1: Ostreoida, oysters (*C. virginica*), 2: Veneroida, clams (*M. lateralis*, *M. leucophaeata*, *R. cuneata*, *T. plebeius*), 3: Mytiloida, mussels (*I. recurvum*)]. Both training sets contained images of larvae from warm and cool treatments of the experiment to ensure wide variation in growth conditions within the training sets (Tables 3 and 5). Three different groups of unknown sets were classified: 1) *C. virginica*, *T. plebeius*, and *M. lateralis*, 2) *C. virginica*, *T. plebeius*, and *I. recurvum*, and 3) *C. virginica*, *R. cuneata*, and *M. lateralis*. Each group contained 7 sets of 100 images of "unknown" larvae in which the percentage of images of *C. virginica* varied (2%, 10%, 25%, 33%, 50%, 75%, and 90%), with the remaining percentages comprised of equal number of images of two other species. Indices of classifier performance were calculated based on the actual number of *C. virginica* images and on true positives, false positives, and false negatives for *C. virginica*. A true positive occurs when an image of *C. virginica* is classified as *C. virginica*. A false positive occurs when an image of a species other than *C. virginica* is classified as *C. virginica*. A false negative occurs when an image of *C. virginica* is misclassified as any other species. Probability of detection (i.e., the probability that the classifier will identify images correctly, P_D = true positive counts/[true positive counts + false negative counts] [Hu and Davis 2006]), specificity (i.e., the probability that the classifier's prediction is correct for each category, SP = true positive counts/[true positive counts + false positive counts] [Baldi and Brunak 2001]), and the ratios of false positives and false negatives to the actual number of *C. virginica* images (e.g., if a sample had 2 images of *C. virginica* and 4 images of mussels were classified as *C. virginica*, then the false positive ratio would be 4:2 or 2.0) were calculated. All indices of classifier performance (P_D, SP, false positive and false negative ratios) were calculated for the 3-category and 6-species training sets, which were applied to each of the unknown groups.

Use of the order-based training set resulted in a similar number of misclassifications as the 6-species training set,

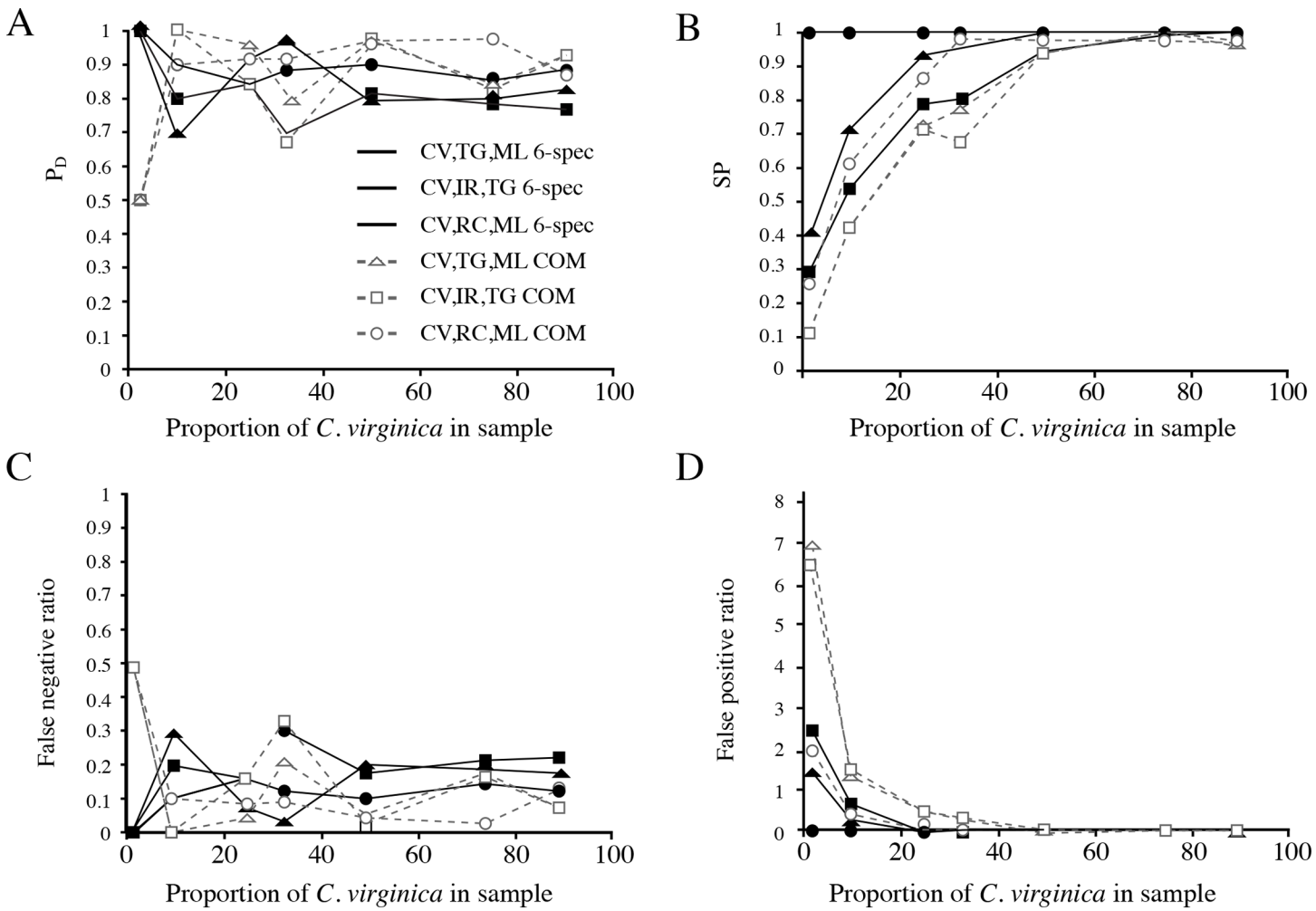

Fig. 5. Misclassification metrics versus the proportion of *C. virginia* (CV) images in a sample: A) probability of detection (P_D), B) specificity (SP), C) the ratio of false negatives to actual *C. virginica* images, and D) the ratio of false positives to actual *C. virginica* images. For all panels, two training sets were used to classify 3 groups of unknown larvae in different proportions. A 6-species training set (6-spec, solid lines) was composed of six categories, each for a separate species: *C. virginica*, *I. recurvum*, *M. lateralis*, *M. leucophaeata*, *R. cuneata*, and *T. plebeius*. A second training set (order-based, dotted lines) contained images of these species grouped by order (clams: *M. lateralis*, *M. leucophaeata*, *R. cuneata*, *T. plebeius*; oyster: *C. virginica*, mussel: *I. recurvum*). These training sets were used to classify three different groups of images of "unknown" larvae: 1) *C. virginica*, *T. plebeius*, and *M. lateralis* (CV, TG, ML), 2) *C. virginica*, *T. plebieus*, and *I. recurvum* (CV, TG, IR), and 3) *C. virginica*, *R. cuneata*, and *M. lateralis* (CV, RC, ML). Each group contained "unknown" sets of images in which the percentage of *C. virginica* in the set ranged from 2% to 90%.

except when the proportion of images of *C. virginica* in a sample was very low (Fig.!5). The probability of detection (P_D) was generally equal or higher for classifications by the order-based training set than for the 6-species training set except when the proportion of images of *C. virginica* comprised 2% of the sample (Fig. 5A). Specificity increased for both training sets as the proportion of images of *C. virginica* in a sample increased, with the 6-species training set performing slightly better when the number of *C. virginica* was high (Fig. 5B). False negative ratios did not exceed 0.33 except for the order-based training set when it was used to classify low percentages of *C. virginica* (2%) (Fig. 5C). The ratio of false positives to actual numbers was higher with the order-based training set when there were relatively few images of *C. virginica* in a sample (Fig. 5D), but this corresponded to a low number of misclassified images (3-8). These metrics show that higher proportions of *C. virginica* in a sample will result in greater classification accuracy, particularly with the order-based training set.

The highest number of false positive and false negative misclassifications from each training set was used to construct confidence intervals that depict the misclassifications that can be expected for different proportions of *C. virginica* in a sample (Fig.!6). The actual *C. virginica* images present plus the highest number of false positives was used to construct the upper line of the interval and the actual *C. virginica* minus the highest number of false negatives was used to construct the lower line of the interval. The confidence interval for the 6-species training set varied from < 5% error at low percentages

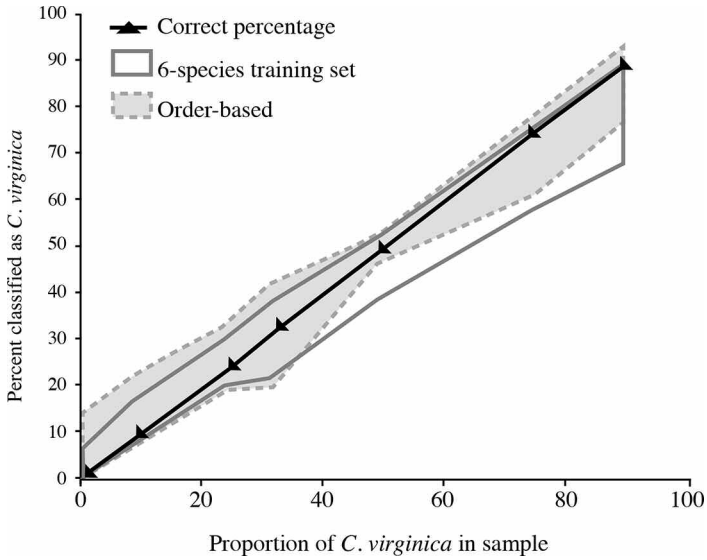

Fig. 6. Classification confidence intervals for the 6-species (no fill with solid gray line) and order-based (gray shading with dashed gray line) training sets. Confidence intervals were constructed around the correct percentage of *C. virginica* classified in a sample (solid line with triangles) using the highest number of false positives and false negatives from tests summarized in Fig. 5. False positives were added to the correct number of *C. virginica* images to construct the top lines and false negatives were subtracted from the correct number of *C. virginica* images to construct the bottom lines. The closer the gray lines are to the black line, the smaller the classification error, which ranged from 5% to 21% for the 6-species training set and from 1% to 22% for the 3-category order-based training set.

(2% *C. virginica* larvae) to < 21% error at higher percentages (90% *C. virginica* larvae). The higher misclassifications at higher percentages are a result of more *C. virginica* being classified as other bivalves (i.e., false negatives) (Fig. 6). The confidence interval for the 3-category order-based training set varied from < 1% error at low percentages (2% *C. virginica* larvae) to < 22% error at medium percentages (33% *C. virginica* larvae) to < 11% at the highest percentages (90% *C. virginica* larvae). The highest error for the 3-category order-based training set is a combined effect of increased false positives and false negatives in the middle ranges (33% *C. virginica*). Based on these results, it is expected that misclassification rates will be within 5% to 21% for the 6-species training set and within 1% to 22% for the 3-category order-based training set depending on the proportion of *C. virginica* in a given sample.

Discussion

Our evaluation shows that the ShellBi technique can be applied with success to distinguish *C. virginica* larvae from the larvae of other bivalve species that are found in the Choptank River, indicating that this approach has application to different species and systems than the one in which it was developed (Waquoit Bay). Results indicate that 1) classification accuracies can increase by as much as 30% when training sets include images of larvae grown in conditions similar to those that are being classified, 2) accuracies can increase by 69% when larvae of different stages (both D-stage and veligers) are included in training sets, and 3) average accuracies are 15% higher when the number of categories within a training set is three compared with six. Although the first two points are novel and specific to this method, the third point has been shown in other image-processing methods that are used to identify plankton (Davis et al. 2004; Grosjean et al. 2004). Finally, misclassification rates were estimated for our target species *C. virginica*, which suggest that this technique can be applied with error rates from 1% to 22% when proportions of the target organisms in the sample range from 2% to 90% (Fig. 6). Results indicate that further methods development aimed at reducing false positive and negative classification rates is a priority.

Differences in growth conditions based on salinity and temperature influenced median shell heights as well as the accuracy of classifying *C. virginica*. Higher temperatures and salinities correspond to faster growth in *C. virginica* (Kennedy 1996) and influence growth in other bivalve larvae (Chanley 1970; Sundberg and Kennedy 1992). Shell heights of *C. virginica* in warm treatments were larger than those in cool treatments, but were shorter than those of hatchery-reared larvae grown at similar warm temperatures. This could be due to the lower assortment of algae fed to the experimental treatments compared with the diet of hatchery *C. virginica* (Langdon and Newell 1996). Regardless of the cause of variation, our results indicate that using images in training sets of larvae that were grown in similar conditions as the unknown sets resulted in higher classification accuracies. This suggests that differences in growth conditions may influence the formation of the shells of bivalve larvae, and hence alter birefringence patterns and classification accuracies. However, potential changes in shell structure and birefringence patterns under different growth conditions warrants further investigation.

The number of categories in a training set and the composition of species in a training set altered the classification accuracy of *C. virginica*. As the number of training set categories increased from 3 to 6, the average accuracy dropped by ~15%, which is consistent with previous studies (Davis et al. 2004; Grosjean et al. 2004; Thompson et al. 2012a). A training set in which 6 species were grouped into 3 categories based on taxonomic order increased classification accuracy of *C. virginica* from 66.8% to 87.8%. These findings suggest that ShellBi would perform well in systems with low numbers of bivalve species in the plankton at any given time (e.g., a system in which 3 species spawn during spring) or in systems where non-target species can be aggregated into a few (≤3) categories.

The composition of the training set was also important. When used to identify the same unknown set, a training set composed of *C. virginica*, *R. cuneata*, and *T. plebeius* had 69.5% accuracy, whereas one of *C. virginica*, *M. leucophaeata*, and *I. recurvum* had 99.8% accuracy (Table 4). This may be explained, to some degree, because smaller *C. virginica* appear to have

similar colors as later stage *T. plebeius* (Fig. 1). This suggests that some species of bivalves at different stages may have birefringence patterns that are similar, resulting in lower classification accuracies, whereas others have patterns that are more distinct, resulting in higher classification accuracies. Although further investigation is needed to determine how shell patterns compare between species throughout development and influence classification accuracies, grouping similar species into a small number of categories can help improve classification accuracies and could be optimized through a machine-learning technique (Fernandes et al. 2009).

The confidence range for misclassifications that can be expected for different proportions of *C. virginica* in a sample may be a conservative estimate. The training sets used in this study were balanced (contain equal numbers of images in each species category) and the SVM classifier assumes that the unknown set contains equal representations of each category (Provost 2000; Lin et al. 2002), but the proportion of *C. virginica* in our unknown sets was varied. Adjusting the cost function (C parameter) of an SVM can help avoid false positives (Sun et al. 2007) and could result in narrower confidence intervals. Future directions to improve ShellBi include adjusting the cost function given different percentages of target species (*C. virginica*) in a sample.

Although the initial set up of ShellBi requires time and effort, ShellBi is the fastest way to both identify and measure different species of bivalve larvae to date once training sets are established. Microscope techniques require a significant time investment while many molecular techniques require time and expense to set up primer or antibody designs or to sequence adult DNA (Garland and Zimmer 2002; Hendriks et al. 2005). When compared with multiplex PCR, ShellBi is less expensive and time consuming for bivalve larvae because individual larvae do not have to be isolated (Thompson et al. 2012a). Although quantitative PCR can provide some insight into the quantity of bivalve larvae, it does not provide information on the sizes of those larvae, which ShellBi does. Another promising technique is fluorescence *in situ* hybridization with DNA probes (Henzler et al. 2010), but the costs are currently prohibitive for large sampling efforts.

Results of this study suggest that ShellBi has broad applicability for the study of size-specific changes in the distribution and abundance of bivalve larvae in estuarine and marine systems. ShellBi has been used successfully to identify larvae in Waquoit Bay (Thompson et al. 2012b) and is being used to help enhance current understanding of *C. virginica* larval dispersal and connectivity in the Choptank River (Goodwin unpubl. data). This technique could be applied to other ecologically and economically important bivalves, both in the laboratory with samples collected from sediment-laden estuaries or in flow-through systems for underway identification of early stage bivalves in marine waters (the tissues of early-stage larvae do not impede resolving birefringent patterns allowing flow-through imaging under field conditions) (S. Gallager pers. comm.). Furthermore, ShellBi may provide insight into the dynamics of other calcareous organisms with shells that show birefringent patterns under polarized light (e.g., pteropods, Goodwin unpubl. data). Finally, because this image-based approach has the potential to be fully automated, it has promise to radically expand our knowledge of the dynamics of bivalve larvae via in situ monitoring platforms and gliders.

Comments and recommendations

Based on the experiments carried out in this study, several improvements are recommended for future applications and research. The first is to establish training sets with several ages of bivalve larvae reared in a range of environmental conditions similar to the system of study. In addition, we recommend the use of the fewest number of categories in a training set as possible. We found that a 3-category training set based on taxonomic order was slightly more accurate at classifying oyster larvae than a 6-category training set in which each category represented a separate species. It is possible that the species grouped by order (e.g., clam larvae) could be distinguished with a second classification test using categories that correspond to species (e.g., *R. cuneata*, *T. plebeius*, *M. lateralis*, *M. leucophaeata*).

Another recommendation is to ensure that the microscope and camera image capture settings are configured so that the background color in all images is uniform for both training and unknown sets. Thompson et al. (2012a) found that training sets created with different microscope settings were not compatible. We found that major background color differences could negatively affect classification accuracies (results not shown), but that minor background color differences (see Fig. 1) for tests conducted in this manuscript did not result in in poor classification accuracies. To avoid major background color variations, we recommend against using metal brackets for polarizers or full wave compensation plates when they are near the light source of the microscope. Changes in temperature due to heating by the light source can lead to large differences in the background color of images when using metal housings. A nonmetal or plastic housing for a polarizer or wave compensation plate near the light source offers more stable conditions that provide similar background colors between images.

The next step for improving the ShellBi method is to increase the speed of image acquisition, ROI extraction, and classification. For the tests presented here, the microscope stage was moved manually or with a joystick attached to an automated stage before an image was taken. A person can image about 100 larval shells per hour with this approach. Currently, efforts toward automation have been made using an automated camera and stage system that will automatically image an entire slide in 46 min (regardless of the number of shells per slide). With this system, 50% of the larvae in a field sample are being imaged in 46 min (half of two slides), which

is faster and more likely to detect rare species than manual identification, which most often relies on subsamples much smaller than half of the sample. In addition, efforts are underway to automate post processing of the bivalve images with automatic ROI detection, ROI cropping, and classification steps, with care taken to assess and minimize errors that can be introduced by subsampling and automation of image analysis (Bachiller et al. 2012). As these enhancements improve how we apply the ShellBi method, so will our ability to rapidly process samples and to conduct field studies with greater spatial and temporal resolution, thereby increasing our understanding of the occurrence and patterns in the presence of bivalve larvae in the field.

Acknowledgments

We thank Victor Kennedy for helping us spawn and rear bivalve larvae and Scott Gallager for use of equipment and training on the ShellBi method. We also thank Aidan Fischer, Kaley Hanrahan, Kayla Hinson, Adam Schlenger, Tom Wazniak, and Rachel Woodward for assistance with collecting, rearing, spawning, and imaging bivalve larvae. Special thanks to Lisa Guy, Melissa Radcliffe, Stephanie Tobash-Alexander, Donald Meritt, and the Horn Point Oyster Hatchery crew for providing oyster larvae, algae and space in the hatchery for our brood stock. In addition, we thank Eneko Bachiller, the Associate Editor, and an anonymous reviewer for their helpful comments and suggestions that have greatly improved this paper. This project was supported by the National Science Foundation (OCE-0829512) and The Bailey Wildlife Foundation. This is UMCES HPL contribution number 4904.

References

Bachiller, E., J. A. Fernandes, and X. Irigoien. 2012. Improving semiautomated zooplankton classification using an internal control and different imaging devices. Limnol. Oceanogr. Methods. 10:1-9 [doi:10.4319/lom.2012.10.1].

Baldi, P., and S. Brunak. 2001. Bioinformatics-the machine learning approach. MIT Press.

Boicourt, W. C. 1988. Recruitment dependence on planktonic transport in coastal waters, p. 183-202. *In* B. J. Rothschild [ed.], Toward a theory on biological-physical interactions in the world ocean. Kluwer Academic Publishers [doi:10.1007/978-94-009-3023-0_10].

Bland, J. M., and D. G. Altman. 1995. Multiple significance tests: the Bonferroni method. BMJ 310:170 [doi:10.1136/bmj.310.6973.170].

Calabrese, A. 1969. *Mulinia lateralis*: Molluscan fruit fly? Proc. Nat. Shellfish. Assoc. 59:65-66.

———, and E. W. Rhodes. 1974. Culture of *Mulinia lateralis* and *Crepidula fornicate* embryos and larvae for studies of pollution effects. Thalassia Jugosl. 10:89-102.

Cawley, G. C. 2000. (MATLAB) Support vector machine toolbox (v0.55\beta). Univ. of East Anglia. <http://theoval.cmp.uea.ac.uk/svm/toolbox/>

Chanley, P. 1970. Larval development of the hooked mussel, *Brachiodontes recurves* Rafinesque (Bivalvia: Mytilidae) including a literature review of larval characteristics of the Mytilidae. Proc. Nat. Shellfish. Assoc. 60:86-94.

———, and J. D. Andrews. 1971. Aids for identification of bivalve larvae of Virginia. Malacologia 11:45-119.

———, and M. Castagna. 1971. Larval development of the stout razor clam, *Tagelus plebeius* solander (solecurtidae: Bivalvia). Ches. Sci. 12:167-172 [doi:10.2307/1350776].

Dame, R. F. 2012. Population processes, p. 75-103. *In* P. Petrailia, and H. Linna [eds.], Ecology of marine bivalves an ecosystem approach, 2nd ed. CRC Press Taylor Frances Group.

Davis, C. S., Q. Hu, S. M. Gallager, X. Tang, and C. J. Ashjian. 2004. Real-time observation of taxa-specific plankton distributions: an optical sampling method. Mar. Ecol. Prog. Ser. 284:77-96 [doi:10.3354/meps284077].

Fernandes, F. A., X. Irigoien, G. Boyra, J. A. Lozano, and I. Inza. 2009. Optimizing the number of classes in automated zooplankton classification. J. Plankton Res. 31(1):19-29 [doi:10.1093/plankt/fbn098].

Fogarty, M. J., and L.W. Botsford. 2007. Population connectivity and spatial management of Mar. Fish. Oceanogr. 20(3):112-123 [doi:10.5670/oceanog.2007.34].

Fukunaga, K., and D. M. Hummels. 1989. Leave-one-out procedures for nonparametric error estimates. IEEE Trans. Pattern Anal. Mach. Intellig. 11:421-423 [doi:10.1109/34.19039].

Garland, E. D., and C. A. Zimmer. 2002. Techniques for the identification of bivalve larvae. Mar. Ecol. Prog. Ser. 225:299-310 [doi:10.3354/meps225299].

Grosjean, P., M. Picheral, C. Warembourg, and G. Gorsky. 2004. Enumeration, measurement, and identification of new zooplankton samples using the ZOOSCAN digital imaging system. J. Mar. Sci. 61:518-525.

Hare, M. P., S. R. Palumbi, and C. A. Butman. 2000. Single-step species identification of bivalve larvae using multiplex polymerase chain reaction. Mar. Biol. 137:953-961 [doi:10.1007/s002270000402].

Helm, M., N. Bourne, and A. Lovatelli. 2004. Hatchery culture of bivalves: A practical manual. FAO Fisheries Technical Paper 471.

Hendriks, I. E., L. A. van Duren, and P. M. J. Herman. 2005. Image analysis techniques: a tool for the identification of bivalve larvae? J. Sea Res. 54:151-162 [doi:10.1016/j.seares.2005.03.001].

Henzler, C. M., E. A. Hoaglund, and S. D. Gaines. 2010. FISH-CS—a rapid method for counting and sorting species of marine zooplankton. Mar. Ecol. Prog. Ser. 410:1-11 [doi:10.3354/meps08654].

Hu, Q., and C. Davis. 2006. Accurate automatic quantification of taxa-specific plankton abundance using dual classification with correction. Mar. Ecol. Prog. Ser. 306:51-61 [doi:10.3354/meps306051].

Kennedy, V. S. 1996. Biology of larvae and spat, p. 371-421. *In* V. S. Kennedy, R. E. Newell, and A. F. Eble [eds.], The Eastern oyster *Crassostrea virginica*. Maryland Sea Grant.

———. 2011a. The invasive dark false mussel *Mytilopsis leucophaeata* (Bivalvia: Dreissenidae): A literature review. Aqua. Ecol. 45:163-183 [doi:10.1007/s10452-010-9344-6].

———. 2011b. Biology of the uncommon dreissenid bivalve *Mytilopsis leucophaeata* (Conrad 1831) in central Chesapeake Bay. J. Mull. Stud. 77:154-164 [doi:10.1093/mollus/eyr002].

———, R. A. Lutz, and C. A. Fuller. 1989. Larval and early postlarval development of *Macoma mitchelli* Dall (Bivalvia: Tellinidae). Veliger 32:29-38.

Langdon, C. J., and I. E. Newell. 1996. Digestion and nutrition in larvae and adults, p. 231-269. *In* V. S. Kennedy, R. E. Newell, and A. F. Eble [eds.], The Eastern oyster *Crassostrea virginica*. Maryland Sea Grant.

Lin, Y., Y. Lee, and G. Wahba. 2002. Support Vector Machines for classification in nonstandard situations. Mach. Learn. 46(1-3):191-202 [doi:10.1023/A:1012406528296].

Loosanoff, V. L., H. C. Davis, and P. E. Chanley. 1966. Dimensions and shapes of larvae of some marine bivalve mollusks. Malacologia 4:351-435.

Lou, T., K. Kramer, D. Goldgof, L. O. Hall, S. Samson, A. Remsen, and T. Hopkins. 2003. Learning to recognize plankton, p. 888-893. *In* Proceedings of the IEEE International Conference on Systems, Man and Cybernetics.

Lutz, R. J., and others. 1982. Preliminary observations on the usefulness of hinge structures for identification of bivalve larvae. J. Shell. Res. 2(1):65-70.

[MDNR] Maryland Department of Natural Resources. 2012. Fixed station monthly monitoring buoy EE1.1. <http://mddnr.chesapeakebay.net/bay_cond/bay_cond.cfm?param=wt&station=EE11>

Munroe, D. M., J. M. Klinck, E. E. Hofmann, and E. N. Powell. 2012. The role of larval dispersal in metapopulation gene flow: local population dynamics matter. J. Mar. Res. 70:441-467 [doi:10.1357/002224012802851869].

North, E., Z. Schlag, R. Hood, M. Li, L. Zhong, T. Gross, and V. S. Kennedy. 2008. Vertical swimming behavior influences the dispersal of simulated oyster larvae in a coupled particle tracking and hydrodynamic model of Chesapeake Bay. Mar. Ecol. Prog. Ser. 359:99-115 [doi:10.3354/meps07317].

Pineda, J., J. A. Hare, and S. Sponaugle. 2007. Larval transport and dispersal in the coastal ocean and consequences for population connectivity. Oceanography 20(3):22-39 [doi:10.5670/oceanog.2007.27].

Provost, F. 2000. Machine learning from imbalanced data sets 101, p. 1-3. *In* N. Japkowicz [ed.], Learning from imbalanced data sets—Papers from the AAAI workshop. AAAI Press.

Shanks, A. L., and L. Brink. 2005. Upwelling, downwelling, and cross-shelf transport of bivalve larvae: test of a hypothesis. Mar. Ecol. Prog. Ser. 302:1-12 [doi:10.3354/meps302001].

Sokal, R., and J. Rohlf. 1987. Non parametric methods in lieu of ANOVA introduction to biostatistics, p. 225-227. *In* R. Sokal and J. Rohlf [eds.], W.H. Freeman and Company.

Sun, Y., M. M. Karnel, A. K. C. Wong, and Y. Wang. 2007. Cost-sensitive boosting for classification imbalanced data. Patt. Recogn. 40:3358-3378 [doi:10.1016/j.patcog.2007.04.009].

Sundberg, K., and V. S. Kennedy. 1992. Growth and development in larval and post-metamorphic *Rangia cuneata* (Sowerby 1831). J. Shell. Res. 11:9-12.

——— and ———. 1993. Larval settlement of Atlantic Rangia, *Rangia cuneata* Bivalvia: Mactridae). Estuaries 16(2):223-228 [doi:10.2307/1352493].

Thompson, C. M., M. P. Hare, and S. M. Gallager. 2012a. Semi-automated image analysis for the identification of bivalve larvae from a Cape Cod estuary. Limnol. Oceanogr. Methods 10:538-554 [doi:10.4319/lom.2012.10.538].

———, R. H. York, and S.M. Gallager. 2012b. Species-specific abundance of bivalve larvae in relation to biological and physical conditions in a Cape Cod estuary: Waquoit Bay, Massachusetts (USA). Mar. Ecol. Prog. Ser. 469:53-69 [doi:10.3354/meps09998].

Tiwari, S., and S. M. Gallager. 2003a. Optimizing multiscale invariants for the identification of bivalve larvae. Proceedings of the 2003 IEEE International Conference on Image Processing, Barcelona, Spain, September 14-17, 2003.

———, and ———. 2003b. Machine learning and multiscale methods in the identification of bivalve larvae. Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, October 14-17, 2003 [doi:10.1109/ICCV.2003.1238388].

Wight, N. A., J. Suzuki, B. Vadopalas, and C. S. Friedman. 2009. Development and optimization of quantitative PCR assays to aid *Ostrea lurida* Carpenter 1984 restoration efforts. J. Shell. Res. 28(1): 33-41 [doi:10.2983/035.028.0108].

Zhao, F., F. Lin, and H. S. Sea. 2010. Binary SIPPER plankton image classification using random subspace. Neurocomputing 73:1853-1860 [doi:10.1016/j.neucom.2009.12.033].

Submitted 7 November 2013
Accepted 27 June 2014
Revised 24 July 2014