Heredity (2012) 108, 616–625
© 2012 Macmillan Publishers Limited All rights reserved 0018-067X/12
www.nature.com/hdy

Heredity

ORIGINAL ARTICLE

Identifying and reducing AFLP genotyping error: an example of tradeoffs when comparing population structure in broadcast spawning versus brooding oysters

H Zhang and MP Hare

Phylogeographic inferences about gene flow are strengthened through comparison of co-distributed taxa, but also depend on adequate genomic sampling. Amplified fragment length polymorphisms (AFLPs) provide a rapid and inexpensive source of multilocus allele frequency data for making genomically robust inferences. Every AFLP study initially generates markers with a range of locus-specific genotyping error rates and applies criteria to select a subset for analysis. However, there has been very little empirical evaluation of the best tradeoff between culling all but the lowest-error loci to minimize overall genotyping error versus the potential for increasing population genetic signal by retaining more loci. Here, we used AFLPs to compare population structure in co-distributed broadcast spawning (*Crassostrea virginica*) and brooding (*Ostrea equestris*) oyster species. Using existing methods for almost entirely automated marker selection and scoring, genotyping error tradeoffs were evaluated by comparing results across a nested series of data sets with mean mismatch errors of 0, 1, 2, 3, 4 and >4%. Artifactual population structure was diagnosed in high-error data sets and we assessed the low-error point at which expected population substructure signal was lost. In both species, we identified substructure patterns deemed to be inaccurate at average mismatch error rates ⩽2 and >4%. In the species comparison, the optimum data sets showed higher gene flow for the brooding oyster with more oceanic salinity tolerances. AFLP tradeoffs may differ among studies, but our results suggest that important signal may be lost in the pursuit of 'acceptable' error levels and our procedures provide a general method for empirically exploring these tradeoffs.

Heredity (2012) **108,** 616–625; doi:10.1038/hdy.2011.132; published online 25 January 2012

Keywords: mismatch error; homoplasy; larval dispersal; gene flow; *Crassostrea virginica*; *Ostrea equestris*

INTRODUCTION

In non-model organisms with few genomic resources, amplified fragment length polymorphisms (AFLPs) provide an efficient means to assay a much more genomically representative sample of polymorphic loci (100s) compared with microsatellites (10s). AFLP markers are efficient because a pair of PCR reactions can be used to simultaneously amplify fragments from multiple chromosomal loci (Vos *et al*., 1995). Electrophoresis of the resulting amplicons generates a fragment profile or 'fingerprint' from which a subset of bands (loci) are deemed reproducible, that is, relatively free from genotyping error. The more genomically representative sample of markers obtainable with AFLPs comes at a price; however: AFLPs are dominant marker data (presence/absence of amplified fragments) with lower information content per locus than in codominant genotypes. Furthermore, coalescent methods of analysis are currently inapplicable with AFLP data for lack of a suitable mutation model.

The control of genotyping error, the primary focus of this study, is a ubiquitous and typically opaque aspect of published AFLP analyses with potentially strong effects on AFLP utility. Genotyping error is detected as a mismatch between the multilocus fingerprints in replicate genotypes from an individual (Meudt and Clarke, 2007). Some sources of genotyping error, such as co-migrating fragments from two or more loci (homoplasy), are well characterized and largely avoidable by scoring only fragments of longer length (Vekemans *et al*., 2002). Other molecular sources of error relate to the sensitivities of PCR to DNA quality and quantity, or artifacts from fragment visualization. These technical considerations partially determine AFLP data quality (Bonin *et al*., 2004). It is often difficult to identify that a technical problem has occurred without comparing replicate AFLP fingerprints, underscoring the importance of using good molecular technique. Even with technically pristine data, genotyping error rate is influenced by the criteria applied at three crucial analytical steps: (1) deliniating the fragment length size boundaries for discrete markers within fingerprints (marker bins, interpreted as different loci), (2) selecting a subset of marker bins to analyze data from (ignoring other fragments deemed uninformative or too prone to scoring error), and (3) scoring fragments as present or absent within analyzed bins so that false positives (from background noise) and low-signal false negatives are minimized. These steps and their influences on genotyping error apply to manual and automated scoring of gels or electropherograms, but the criteria tend to be more subjective and the steps less discrete with manual procedures. These multiple factors affecting error suggest that in most applications, AFLP genotyping error can be minimized but rarely eliminated without severely limiting the number of loci and information content in the resulting data.

Department of Natural Resources, Cornell University, Ithaca, NY, USA
Correspondence: Dr H Zhang, Department of Natural Resources, Cornell University, 208 Fernow Hall, Ithaca, NY 14853, USA.
E-mail: zhbiocas@gmail.com

Received 18 May 2011; revised 21 November 2011; accepted 6 December 2011; published online 25 January 2012

It is undeniable that in principle, genotyping error should be minimized. In practice, however, after technical error sources have been minimized as much as possible, control on overall error is often exerted primarily by ignoring fragment bins deemed to be error prone. There is some risk that perceived or real publication pressures will constrain authors to analyze only low-error data sets, without even evaluating whether additional signal was available for a given genotyping effort by including more loci in the analyzed data. Thus, there are likely to be pragmatic tradeoffs between the increased signal of population structure obtained by including more loci (despite higher average genotyping error) versus more error-free data sets with fewer loci and potentially less power to detect population structure. These tradeoffs have barely been explored or quantified with respect to analyzing population structure. Our goal here is to explore methods for evaluating this tradeoff.

Defining binset boundaries, and particularly optimizing those boundaries is the initial analytic step most frequently done manually, and arguably the step that most limits transferability of AFLPs across labs. The manual binsetting approach most commonly employed entails visual comparison of all sample chromatographs in an overlay or pile-up fashion. Even if initial bin boundaries are determined by a parameterized automated procedure such as available in GeneMapper (Applied Biosystems, Foster City, CA, USA), optimization of bin boundaries is then subjective, laborious, and becomes dramatically more laborious if large numbers of samples are genotyped, especially for species with large genomes (thereby producing relatively dense fragment traces). We are aware of only two programs that apply an optimization criterion to initial binset creation: Peakmatcher (DeHaan *et al.*, 2002) minimizes differences between replicated samples while applying multiple criteria to choose among many overlapping bin candidates. RawGeno (Arrigo *et al.*, 2009) evaluates fingerprints from all samples, starts with one massive bin and repeatedly applies rules for subdividing and retaining bins such that bin widths are minimized (not to exceed a defined value) and intervals between bins are maximized. Here, with the expectation that AFLP analysis steps two and three can 'clean up' a binset containing many error-prone bins (that is, we need not 'trust' the optimization, but simply use it as a starting point), one of the goals of this study was to apply automated binset optimization in anticipation of analyzing larger data sets in the future.

Recently, two methods were published for reducing AFLP genotyping error at the marker selection and fragment calling stages (steps 2 and 3 above) by minimizing false positives and negatives. Herrmann *et al.* (2010) established locus elimination criteria and peak height (that is, signal amplitude) fragment calling thresholds for each locus independently based on the peak height distribution across all samples. Whitlock *et al.* (2008) used the mean of each peak height distribution per locus to eliminate low-signal loci and evaluate genotyping error rates resulting from a range of absolute or relative peak height fragment calling thresholds. Both methods use replicate genotypes to guide the optimization in terms of minimizing genotyping error, while maximizing the number of loci retained. In our experience using these methods, it is rarely clear which thresholds are more desirable, those giving lower average error and smaller data sets or those yielding higher error and larger data sets, but they make it possible to explore trade-offs using objective and repeatable criteria. Two studies that have used these tools to compare population genetic informativeness of low and high stringency data sets found that the latter, with fewer scored loci, seemed to provide slightly more explanatory power with respect to population structure (Herrmann *et al.*, 2010; Crawford *et al.*, 2011).

In an important review on tracking and assessing genotyping errors for any marker type, Bonin *et al.* (2004) generalized that the potential consequences of error for inferences is inversely proportional to the scale of inference; potentially severe bias can result when the unit of analysis is individuals within a pedigree or individual loci within a population (for example, genome scan outliers, pairwise linkage disequilibrium), whereas lesser effects are expected at the population or among-population level. Effects on estimation of genetic diversity and population structure have mostly been hypothesized (Bonin *et al.*, 2004), demonstrated with simulations (Caballero *et al.*, 2008), and only recently explored empirically (Arrigo *et al.*, 2009; Herrmann *et al.*, 2010; Crawford *et al.*, 2011).

Bonin *et al.* (2004) hypothesized that 0.1% average genotyping error among loci may provide little improvement for population genetic inferences compared with 2%. It is important to know whether the same can be said of 3 and 4% average error among loci because if so, most studies would be able to include many more loci with potentially greater power to detect subtle substructure. For population genetic inferences, empirical error rates in published plant and animal studies have been <5% (Bonin *et al.*, 2004) but rarely much over 2%. The optimum tradeoff is not expected to be the same across study systems, but until the empirical population genetic consequences of these trade-offs are compared for a number of systems, choosing an acceptable genotype error rate will be arbitrary.

Here, we use AFLPs to analyze the population structure of two broadly distributed oyster species in the western North Atlantic: the eastern oyster, *C. virginica* (Gmelin 1791) and the crested or horse oyster, *O. equestris* (Say 1834). These two species co-exist along most coasts of the southeastern United States but our comparison focuses on the biotically rich central Atlantic Florida coast where both species occur in a string of lagoons. A well-characterized step cline in *C. virginica* (Reeb and Avise, 1990; Karl and Avise, 1992) centered near Cape Canaveral, Florida (Hare and Avise, 1996) anchors our AFLP error rate comparisons to a known phylogeographic pattern that should be observed in representative data. AFLP markers previously were used to compare an eastern oyster sample from just north of the Cape Canaveral cline with a sample from Port Charlotte, southwestern Florida, and strong differentiation was found as expected (Murray and Hare, 2006). However, population comparisons directly across the step cline have not been made previously with AFLPs. For crested oysters, mitochondrial DNA sequence homogeneity has been reported between Atlantic and Gulf of Mexico populations, but the small and geographically sparse samples provided little power to test for structure along the Atlantic coast (Kirkendale *et al.*, 2004).

This species comparison tests the generality of biotic and/or abiotic processes generating coastal population structure in co-distributed species with larval dispersal but somewhat different larval biology. The crested oyster broods its larvae for several days before release to the plankton, and oceanic salinities are most favorable to post-settlement growth and reproduction (Hoese, 1960). The eastern oyster has external fertilization and fully planktotrophic larvae that enjoy higher viability within estuarine salinities (Shumway, 1996). Thus, despite larval brooding by *O. equestris*, we hypothesized that it experiences higher dispersal success than *C. virginica* among eastern Florida lagoons by means of advection along the continental shelf, and therefore would have lower population substructure among sampled lagoon populations.

MATERIALS AND METHODS

Sample collection

For a larger study focused on *C. virginica* in 2007, nylon mesh bags full of clean oyster shell were deployed on the intertidal shore at 30 sites along eastern Florida to collect recently settled juvenile oysters (spat). Spat samples were

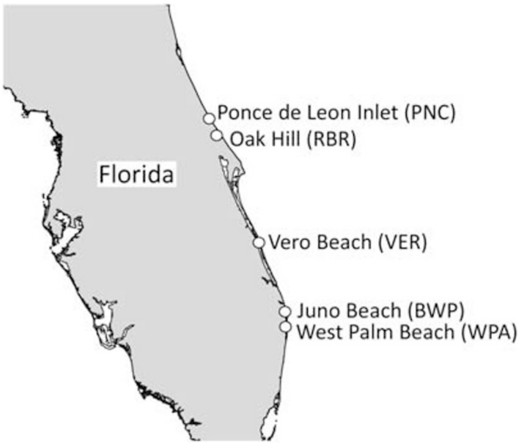

Figure 1 Sample localities for *C. virginica* and *O. equestris* with acronym abbreviations.

collected and fresh shell bags deployed monthly from May to July. Because *C. virginica* and *O. equestris* are difficult to distinguish at the spat stage, all spat up to 100–150 individuals were collected at each site and initially regarded as oyster spp. Spat tissues were dissected to avoid the gastrointestinal tract, and preserved in 95% ethanol. Total genomic DNA was extracted from entire specimens (1–20 mg of tissue) using the DNeasy 96 Tissue kit (QIAGEN Inc., Valencia, CA, USA) following the manufacturer's protocol for animal tissues. DNA was eluted in 100 μl elution buffer and diluted to 20 ng/μl working stocks based on spectrophotometric measurements.

For the purposes of this study, we wanted to focus on a small set of populations with known population structure (in *C. virginica*), but also wanted the AFLP binsets to be based on as many samples as possible. We obtained spat at 18 of the 30 sampling sites and genotyped a total of 573 oysters from the 18 sites, including all available spat from five focal sites (Figure 1) where both species proved to be abundant (see below). To avoid ascertainment biases during automated binset optimization, some spat from the other 13 sites were also included and duplicate genotyped.

Species identification

Species-specific primers were designed to amplify different-sized fragments from the nuclear 28S rRNA gene of each species. Each multiplex PCR reaction of 15 μl total volume included 0.025 units of *Taq* (Invitrogen, Carlsbad, CA, USA), 1× buffer (Invitrogren), 0.2 mg ml^{-1} BSA, 0.125 mM of each dNTP, 1.0 mM MgCl$_2$, 0.2 μM each of Cv28s-L (5′-AGAACCGGGGAGAGGTGC-3′) and Cv28-R (5′-AGGGACGAGAGCGGAAGG-3′), 0.1 μM each of Oe28s-2L (5′-AAGAGCCGGGGGAAGGTAT-3′) and Oe28s-R (5′-ACCGAGGATCCCA CCTAGA-3′), and 1.0 μl DNA template. An MJ Research PTC-225 was used for PCR with 95 °C for 2 min, 11 touchdown cycles (94 °C for 30 s, annealing for 20 s at 68 °C to 63 °C, −0.5 °C per cycle, 72 °C for 30 s), then 20 additional cycles with annealing at 63 °C. PCR products were visualized in 1.5% agarose gels. The expected band sizes were 150 bp for *C. virginica* and 400 bp for *O. equestris.*

AFLP data collection

AFLP fragments were generated according to a modified version of the procedure outlined by Vos *et al.* (1995). Primers (one fluorescently labeled), adaptors, enzymes and PCR conditions were the same as in Murray and Hare (2006) except for several modifications made after exploring their effects on reproducibility between duplicate reactions: the digestion included 100 ng of genomic DNA in a 50 μl total volume with 10 U *Mse*I; pre-selective PCR included 0.1 μg μl^{-1} BSA in 10 μl reactions; selective PCR included 0.0024 μg μl^{-1} BSA in 10 μl reactions. The four selective primer pairs used, *Eco*RI-ACT/*Mse*I-CAA, *Eco*RI-ACT/*Mse*I-CAC, *Eco*RI-ACT/*Mse*I-CAG, and *Eco*RI-ACT/*Mse*I-CAT, were previously selected as less error prone from a larger set examined in *C. virginica* (Murray and Hare, 2006). Two negative controls and one positive control were included on each plate. AFLP fragments were run with an internal LIZ size standard on an ABI PRISM 3730 genetic analyzer (Applied Biosystems) at the Cornell University Life Sciences Core Laboratories Center.

Creating primary binsets

Replicate AFLP fingerprints, including independent restriction digestions of genomic DNA, were collected for 117 *C. virginica* and 68 *O. equestris* individuals from across all 18 collection sites (32% of all genotyped samples) to measure genotyping repeatability and create binsets for each selective primer pair in each species. Those duplicate fingerprints ('duplicates' hereafter) in which one or both had obviously poor amplification (to an extent visually detectable from the electropherogram trace) were removed from further analyses. Finally, 110 to 115 duplicates in *C. virginica* (variation among the four selective primers), and 68 duplicates in *O. equestris* were used for binset creation. Binsets for each species were created separately, but because the primer pairs were chosen for ease of reproducible scoring in *C. virginica*, it is possible that ascertainment biases could affect some aspects of the *O. equestris* data.

Bin definitions were initially determined from comparison of duplicates using Peakmatcher software (DeHaan *et al.*, 2002). This program analyzes a matrix of fragment sizes (bp) output by Genemapper 4.0 (Applied Biosystems) from multiple duplicate fingerprints. Peakmatcher optimizes marker bins by evaluating many possible overlapping bins while maximizing duplicate fingerprint similarity and applying rules to exclude bins with error-prone attributes. It was shown by the authors to be highly consistent with manual methods, but more definable and repeatable (DeHaan *et al.*, 2002). Peakmatcher was applied to each species separately, and for each selective primer pair in turn, using the 'autobin' feature of Genemapper to generate comprehensive (75–500 bp) fragment size tables from duplicates using a relative fluorescent units minimum of 200. Peakmatcher settings were: 'category range'=0.3–0.5, 'category increment'=0.1, 'minimum repeatability'=75%, 'minimum peaks present'=60%.

The bin definitions from Peakmatcher were imported into Genemapper where each bin was visually checked across all duplicates. Bin borders were manually adjusted to center the main peak density, and bins (loci) were removed if it appeared they could be error prone upon inclusion of additional samples (that is, non-duplicated samples). For example, some loci were rejected because of adjacent peaks crowding a bin.

The resulting 'modified Peakmatcher binset' was used in Genemapper for automated calling with a minimum signal intensity of 200 relative fluorescent units. The resulting binary fragment presence/absence matrix was exported to Microsoft Excel where error rate for each locus was calculated in two ways:

(1) Mismatch error rate (Bonin *et al.*, 2004):

$$\frac{N_{(0,1)}+N_{(1,0)}}{N_{(1,0)}+N_{(0,1)}+N_{(1,1)}+N_{(0,0)}} \qquad (1)$$

For each locus, $N_{(1,1)}$ represents the number of duplicates where both genotypes have a fragment, $N_{(0,0)}$ represents the number of duplicates where neither genotype has a fragment, $N_{(0,1)}$ and $N_{(1,0)}$ represent the number of mismatches where one genotype of a duplicate has a fragment and the other does not.

(2) Jaccard error rate (Holland *et al.*, 2008):

$$\frac{N_{(0,1)}+N_{(1,0)}}{N_{(1,0)}+N_{(0,1)}+N_{(1,1)}} \qquad (2)$$

The Jaccard error rate, by discounting (0,0) occurrences, will be inflated relative to the mismatch rate when the band-present phenotype is at low frequency, but in these cases it could be argued that the mismatch rate estimation is too insensitive to errors. Thus, the two estimation methods are complementary. Arbitrary thresholds were applied such that bins with mismatch error rate >15% or Jaccard error rate >30% were removed to create the smaller 'primary binsets' for subsequent automated fragment calling (Supplementary Figure 1). On the basis of the four primer pairs, the primary binset for *C. virginica* included a total of 229 loci with a 5.46% average mismatch error rate, and for *O. equestris* included 187 loci with a 5.87% average mismatch error rate.

Species assignment confirmation

The 28S species diagnostic was only applied to a subset of specimens and among these a few disagreements were found with a first-generation mitochondrial diagnostic (data not shown). To resolve these ambiguities we used preliminary AFLP data from all specimens to perform assignment tests, classifying known-species reference sets based on unanimous mtDNA identifications and assigning the rest as unknowns. For each species' primary binset in turn, Genemapper was used to generate a binary AFLP matrix for all 573 individuals. Assignment tests were conducted using AFLPOP v. 2.0 (Duchesne and Bernatchez, 2002) applied to each matrix separately. The assignment likelihood threshold was set to 0, that is, an individual was allocated to the species in which it had the highest likelihood.

Marker selection and fragment calling

Further marker selection for each species was accomplished in two ways. First, ScanAFLP (Herrmann *et al*., 2010) was applied to all samples (including both duplicates) to either eliminate bins or apply locus-specific signal intensity thresholds based on qualities of the fragment signal intensity distribution. Because the whole point of ScanAFLP is to optimize low-signal calling thresholds, we used the primary binset to export fragment data from each species using Genemapper with minimum relative fluorescent units of 50. The resulting ScanAFLP output is a binary data matrix referred to here as MatrixB for consistency with Herrmann *et al*. (2010), although we implemented only steps (2) and (3) from their procedure. For the second step, ScanAFLP output matrices were imported into Excel, bins were sorted by locus-specific mismatch error, and then the highest mismatch error bins (based on duplicates) were iteratively removed until a desired average error rate was achieved among the remaining loci. Nested data sets having zero to 4% average mismatch error, in addition to matrixB, were created by randomly removing one fingerprint from each duplicate. Degree of polymorphism was ignored during binset creation and bin culling, so monomorphic loci were present in all analyses.

Genetic diversity and population structure comparisons

To evaluate the effect of genotyping error on population genetic inferences, the genotype matrices with different average error rates were used to estimate population diversity and infer population structure in the two species. Intrapopulation genetic diversity of each species was measured from all loci in each data set as band richness after rarefaction to $n=29$, Br(29), and percentage of polymorphic loci at 1% level for a standardized sample size, PLP1%(29), calculated using Aflpdiv 1.1 (Coart *et al*., 2005). Band richness is the average number of band-present phenotypes expected at each AFLP locus in each population in a specified sample size (Coart *et al*., 2005). Expected heterozygosity, H_e, was estimated using AFLP-SURV 1.0 (Vekemans *et al*., 2002) using the Bayesian method with non-uniform prior for allele frequency estimation and assuming Hardy–Weinberg equilibrium. We distinguish between band-present *phenotype* frequency, which is a simple band counting calculation, from band-present *allele* frequency that when estimated using Bayesian procedures, depends on the inferred allele frequency spectrum (Bonin *et al*., 2007). To test for differences in genetic diversity among populations, one-way ANOVA was applied using calculations in Excel (McDonald, 2009).

Multilocus pairwise F_{ST} within species was calculated according to the Bayesian method of AFLP-SURV with statistical significance evaluated relative to a null distribution of values based on 10 000 permutations of individuals among populations. Observed F_{ST} values were considered significant if they were greater than the 99th percentile of the null distribution (Vekemans *et al*., 2002). Individual locus F_{ST} was calculated as the standardized variance of weighted average allele frequencies (Hedrick, 2005, p489) and tested for a correlation with mismatch error rate using SPSS11.5. To test for hierarchical population differentiation, analysis of molecular variance (AMOVA) was carried out using GenAlEx 6.3 (Peakall and Smouse, 2006) to estimate ϕ_{PT} based on a Euclidian band-sharing genetic distance.

The number of distinct populations represented in the data from each species was estimated using STRUCTURE 2.3 (Pritchard *et al*., 2000). For each species, STRUCTURE was run 20 times each for $K=1$–5, using the correlated allele frequencies and admixture model, RECESSIVEALLELES set to 1 and sampling locations as a prior (LOCPRIOR=1). Runs used 400 000 MCMC iterations after a burn-in of 100 000. Using Structure Harvester v. 0.6.7 (http://taylor0.biology.ucla.edu/struct_harvest/) for calculations, the best supported value of K was judged by comparing estimated Ln P(X|K) averaged over the 20 runs at each K. Structure results were plotted using DISTRUCT 1.1(Rosenberg, 2004).

Some substructure patterns are indicative of technical artifacts and we looked for these in each of the data sets. Care was taken to make sure there was not a one to one correspondence between PCR plate composition and sample composition from a locality. Then, plate boundary artifacts (plate effects) were identified with STRUCTURE bar charts by determining whether inferred clusters (different colors in the output for $K>1$) corresponded with plate boundaries as opposed to locality of origin.

RESULTS

Characterizing AFLP profiles and error

Our automated procedures generated candidate bins (loci) with error rates as high as 90% for Jaccard and 25% for mismatch error before arbitrary truncation to 30% and 15%, respectively, produced the primary binset as a starting point for (further) marker selection and calling using ScanAFLP (Supplementary Figure 1). First, each species-specific primary binset was used to generate a genotype matrix including all individuals (that is, both species combined). In cases where the 28S diagnostic had been ambiguous because of disagreement with earlier mtDNA assays, AFLP species assignments with each of these data sets were consistent with 28S, confirming its accuracy. Two specimens showed a conflict between AFLP and 28S identifications, but these had the lowest assignment likelihoods and poor quality (low signal) AFLP data. After removal of these two specimens, data sets contained 246 individuals assigned to *C. virginica* and 197 individuals to *O. equestris*.

For each species, five nested (that is, non-independent) data sets were created from matrixB with incrementally less average mismatch error based on duplicates (Table 1). For *C. viginica,* some AFLP data from the CAT-selective primer showed plate effects (Supplementary Figure 2) as described below and was repeated from the archived pre-selective reaction (15 individuals). The matrixB mismatch error rate was 4.7% in *C. virginica* and 5.7% in *O. equestris*. The number of loci in the six data sets (0–4% error plus matrixB) ranged from 28 to 206 for *C. virginica*, and 26 to 167 for *O. equestris*.

Mismatches can be tallied within duplicates across all loci, or for each locus across all duplicates. Comparing duplicate genotypes in matrixB, the percent of mismatches varied from 0.49 to 16.02% and had a median of 3.88% in *C. virginica* versus 1.80 to 21.56% in *O. equestris* with a median of 4.49%. Non-normalized chromatographs in duplicate pairs with the highest error were checked for signs of low or high signal intensity and no distinctions were found. Locus-specific mismatch error rate distributions also were right skewed in all data sets (not shown) with a minimum of zero and maximum of 21.74% for *C. virginica* and 25.00% for *O. equestris* in the matrixB data sets.

Homoplasy was evident from a negative correlation between fragment size and band-present frequency, which was significant when based on band phenotypes ($r=-0.49$ to -0.87; $P<0.01$) and on Bayesian band-present allele frequencies ($r=-0.41$ to -0.87; $P<0.01$; Figure 2). Decomposing the patterns that produced this correlation, both species had roughly *J*-shaped frequency distributions of band-present phenotype frequencies, with the 0–0.1 class (fixed or nearly fixed for band-absent allele) and the 0.9–1.0 class (fixed or nearly fixed for band-present allele) being the second and first most abundant classes, respectively, in every data set (Supplementary Figure 3). However, whether examining band phenotype frequency or Bayesian-estimated allele frequencies, the frequency distribution was

Table 1 Number of duplicate samples (DS) and loci for AFLP data sets in each species, listed by the selective *Eco*RI and *Mse*I nucleotides

	No. of DS	*Number of loci*							
		0% Error	*1% Error*	*2% Error*	*3% Error*	*4% Error*	*MatrixB*	*Primary binset*	*Peakmatcher output*
C. virginica									
*Eco*RI-ACT/*Mse*I-CAA	115	6	24	33	39	44	49	52	217
*Eco*RI-ACT/*Mse*I-CAC	115	9	26	37	43	51	52	59	211
*Eco*RI-ACT/*Mse*I-CAG	110	7	28	38	46	51	53	64	231
*Eco*RI-ACT/*Mse*I-CAT	113	6	14	30	38	47	52	54	210
Total		28	92	138	166	193	206	229	869
O. equestris									
*Eco*RI-ACT/*Mse*I-CAA	68	5	12	17	21	24	31	33	151
*Eco*RI-ACT/*Mse*I-CAC	68	9	23	31	42	53	61	68	176
*Eco*RI-ACT/*Mse*I-CAG	68	4	9	16	19	27	35	42	197
*Eco*RI-ACT/*Mse*I-CAT	68	8	19	26	33	38	40	44	178
Total		26	63	90	115	142	167	187	702

The average mismatch error rate (four primer pairs) of peakmather output, primary binset, matrixB are 8.30%, 5.46%, 4.72% in *C. virginica*, and 9.67%, 5.87%, 5.72% in *O. equestris*, respectively.

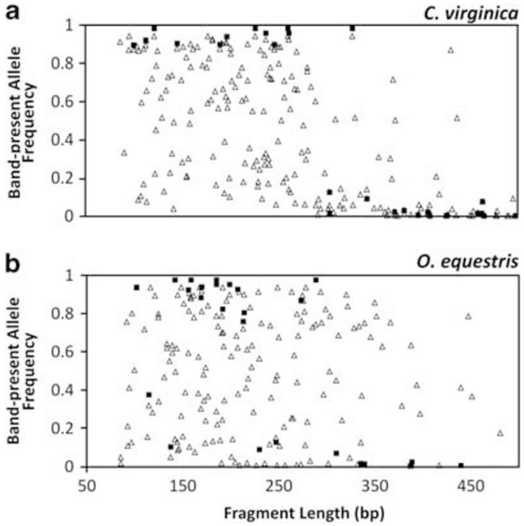

Figure 2 Average frequency of band-present allele across the five populations relative to fragment size (bp) for (**a**) *C. virginica* and (**b**) *O. equestris* based on the matrixB (open triangles) and 0 error (filled squares) data sets. The 0 error data are a subset of the matrixB data after culling high-error loci. Allele frequencies were estimated from all individuals using AFLPsurv with Bayesian non-uniform priors.

distinct for fragment lengths below and above 300 bp (Figure 2). As expected from homoplasy (comigrating fragments) in smaller fragments, band-present allele frequencies were skewed-high for fragments <300 bp whereas frequencies were skewed-low, with many fewer moderate frequency loci, for fragments >300 bp (Figure 2). Furthermore, in both species, as higher error loci were discarded to produce smaller data sets with lower average error, the loci with moderate band-present frequencies were disproportionately eliminated (Figure 2; Supplementary Figure 3), strengthening the disparity in allele frequency spectrum for fragments <300 and >300 bp in length. Homoplasy was not restricted to small fragments; even fragments >300 bp showed a significant negative correlation between fragment size and band-present frequency for *C. virginica* in matrixB (r=−0.60, $P<0.01$) and for *O. equestris* in the three largest data sets ($P<0.05$).

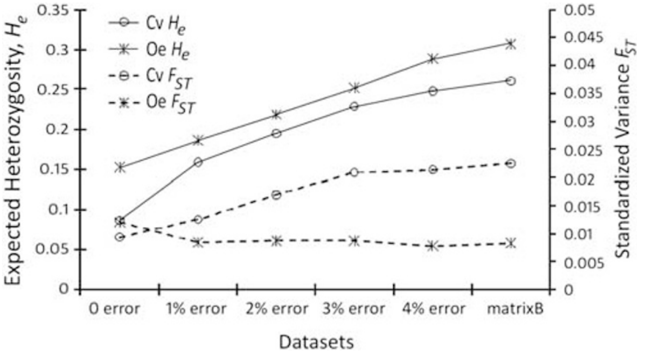

Figure 3 Overall expected H_e(left scale and solid lines) and overall F_{ST} (right scale, dashed lines) plotted for *C. virginica* (open circles) and *O. equestris* (hatch marks) across the six data sets. Both parameters were estimated with AFLPsurv using Bayesian non-uniform priors.

Mismatch error rate relative to H_e and F_{ST}

It is important to know the degree to which genotyping error is associated with observed levels of genetic diversity or degree of population differentiation. Not surprisingly, gene diversity (H_e) was substantially higher in data sets of both species containing more loci and higher average mismatch error (Figure 3). All else being equal, this should lead to lower levels of F_{ST} in the higher-error data sets, as has been shown previously for AFLP data (Herrmann *et al*., 2010), but that was not the case here. Instead, in *C. virginica* overall F_{ST} increased roughly in parallel with H_e, showing its highest value in matrixB where mismatch error was greatest, whereas F_{ST} varied little across the *O. equestris* data sets (Figure 3). Indeed, locus-specific F_{ST} was positively associated with mismatch error rates in the matrixB data ($P<0.01$; r=0.37) and all smaller data sets except 0% error.

Effect of error rate on population structure inference

For *C. virginica* data with average error rates of 2% or less, STRUCTURE failed to detect any population structure (strongest support for K=1, Figure 4a). For data sets with mismatch error ⩾3%, STRUCTURE results indicated support for K=2 or 3 in *C. virginica* (Figure 4a). With K=2 assumed, STRUCTURE divided the five samples into two clusters consistent with the phylogeographical

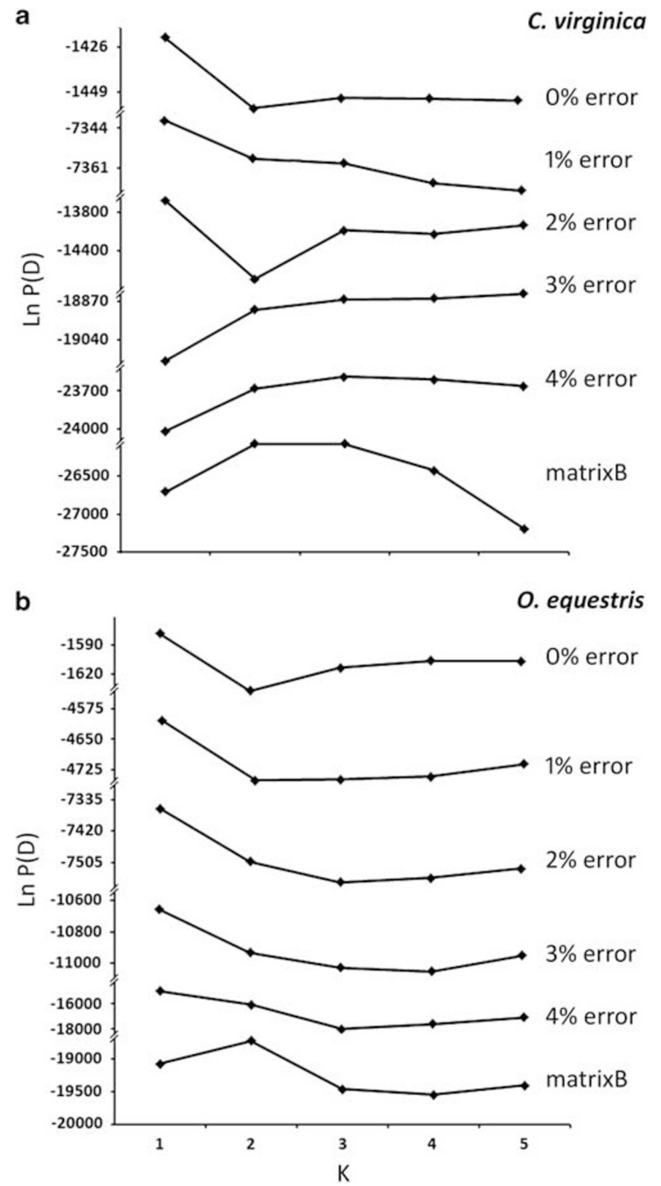

Figure 4 Mean Ln *P*(*D*) values from 20 STRUCTURE runs evaluating *K*=1–5 in the six data sets for (**a**) *C. virginica* and (**b**) *O. equestris*. Note breaks in the *Y* axis scale; absolute probabilities vary across data sets because of their different number of loci, but it is the pattern of relative support across different K in different data sets that is of interest.

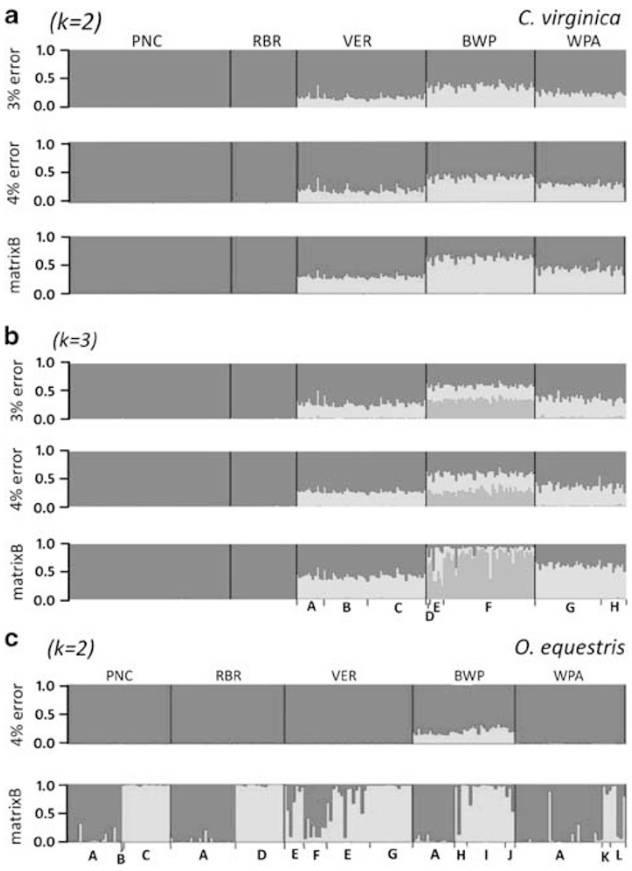

Figure 5 Bar plots from STRUCTURE analysis using different data sets from *C. virginica* and assuming (**a**) *K*=2 or (**b**) *K*=3. Data set labels are to the left of each plot, population samples are demarcated with thin vertical black lines and labeled above with acronym names, and each individual genotype is represented as a vertical bar partitioned into segments of different shades representing the posterior probabilities of membership in *K* inferred clusters. In (**b**), membership on experimental genotyping plates is labeled at the bottom with capital letters A – H and plate boundaries are shown with black tick marks to highlight possible artifacts. (**c**), Bar plots from STRUCTURE analysis using 4% error and matrixB data sets from *O. equestris* with *K*=2. Bar plots and legends organized as in (**a**) and (**b**). Note genetic cluster boundaries in the matrixB result align with genotyping plates in some cases, indicating plate artifacts that are not evident in the 4% error data.

break at Cape Canaveral, FL (Hare and Avise, 1996). In the north, samples PNC and RBR were clustered together into a population with no admixture, whereas south of Cape Canaveral VER, BWP, WPA formed a cluster with similar admixture composition across the samples (Figure 5a). Clustering with *K*=3 either made admixture patterns more complex in every southern individual (3% error) or distinguished the BWP sample with a different admixture pattern (4% error and matrixB; Figure 5b). MatrixB data showed plate effects within the BWP sample.

For *C. virginica* data sets with inferred $K>1$, that is, 3% error to matrixB, we evaluated tradeoffs between signal and noise based on the magnitude of differentiation. One statistic that appeared informative was mean genetic distance among clusters, calculated from the matrix of Bayesian posteriors output from STRUCTURE. For the highest likelihood *K*=3 runs with each data set, the mean genetic distance among the three inferred clusters increased from 3% error to 4% error then decreased with matrixB (0.041, 0.052, 0.020, respectively), as might be expected if population substructure signal was initially being added with additional loci but eventually became overwhelmed by noise. Similarly, focusing on the 3rd cluster that occurs only in BWP (Figure 5b), the F_{ST} analog reported by STRUCTURE increased from 0.42 to 0.52 between 3% error and 4% error, then dropped to 0.09 for the matrixB data set.

A slightly different tradeoff pattern was found using AMOVA to compare data sets for the strength of support for two-region hierarchical structure in *C. virginica*. The genetic break between regions was significant in all data sets ($P\leqslant0.001$) and explained up to 5.14% of the variance in the 3% error data but decreased in smaller and larger data sets (Supplementary Figure 4). The proportion of variation among populations within the regions increased with average error to a high of 2.85% in matrixB data, all with $P\leqslant0.003$ except that 0 error was not significant (Supplementary Figure 4). In summary, STRUCTURE results showed maximum differentiation of clusters in 4% error data, whereas AMOVA detected maximum regional variance with 3% error, but in both cases similar patterns would lead to identical conclusions with both 3% and 4% data sets.

In *O. equestris*, no geographical structure (K=1) was found among the five populations in all data sets except matrixB (Figure 4b), but if K=2 was assumed then matrixB data showed cluster boundaries aligned with plate boundaries (Figure 5c). Interestingly, in the 4% error result for *O. equestris,* the second 'cluster' distinguished BWP as an admixed population, similar to results in *C. virginica* with K=3. AMOVA was used to test all possible regional breaks (single partitions) among the five populations analyzed. 4% error data showed a significant regional break between VER and BWP (0.49% variation between regions, $P\leqslant 0.01$), and the same break was marginally significant with 3% error data ($P\leqslant 0.048$).

Population genetic comparison between species

Quantitative comparisons are reported for the 4% error data set but the 3% error data showed consistent patterns except where noted. All diversity measures, *Br*(29), PLP1%(29) and Bayesian H_e, had higher average values in *O. equestris*, but the differences were only significant for H_e ($P<0.05$, paired t-test across populations). In addition, *C. virginica* showed higher variation in diversity among populations than *O. equestris*, as indicated by the coefficient of variation across samples (Table 2). Planned comparisons between the two northern and three southern samples in *C. virginica* showed significantly lower diversity in the north for Br(29) (ANOVA, $F_{(1,775)}$=6.94, P=0.009) and H_e ($F_{(1,770)}$=5.40, P=0.02).

Higher pairwise F_{ST} values were found in *C. virginica* compared with *O. equestris*, even for pairs within one or the other *C. virginica* region (Table 3). All pairwise F_{ST} values in *C. virginica* were different from zero at the 0.01 level whereas in *O. equestris* this was true only for the VER-BWP comparison. On the basis of non-hierarchical AMOVA analyses, more variation was distributed among populations in *C. virginica* (4.90%) than in *O. equestris* (0.71%).

DISCUSSION

The effects of genotyping error on population genetic inferences

Relatively high genotyping errors are a well-known weakness of AFLP data; they can be minimized but rarely eliminated. In highly polymorphic species with large genomes, even selective amplification using three random nucleotides will produce complex AFLP fragment profiles with inevitable genotyping error (Althoff *et al.*, 2007).

Given that many population genetic inferences depend primarily on increased genomic sampling for accuracy and statistical power, there is likely to be a tradeoff whereby larger data sets (more loci)

Table 3 Pairwise F_{ST} values (lower left) and *P* values from permutation tests of significance (upper right) in *C. virginica* and *O. equestris*

	PNC	*RBR*	*VER*	*BWP*	*WPA*
C. virginica					
PNC		0.002	0.000	0.000	0.000
RBR	0.011		0.000	0.000	0.000
VER	0.039	0.038		0.000	0.000
BWP	0.081	0.068	0.024		0.000
WPA	0.057	0.042	0.011	0.022	
O. equestris					
PNC		0.862	0.022	0.046	0.254
RBR	0.000		0.199	0.025	0.091
VER	0.009	0.004		0.007	0.013
BWP	0.009	0.010	0.011		0.503
WPA	0.004	0.007	0.010	0.002	

Table 2 Collection coordinates, month of collection in 2007, average March–May 2007 salinity (parts per thousand), sample size (*N*) and genetic diversity statistics for *C. virginica* and *O. equestris* at five collection sites

Sample ID	*Lat. (°N)*	*Long. (°W)*	*Collect. month*	*Salinity (p.p.t.)*	*N*	*Br(29)*	*PLP1% (29)*	H_e	*s.e. (H_e)*
C. virginica									
PNC	29.08	80.93	06	33.9	72	1.664	0.777	0.2219	0.0121
RBR	28.90	80.85	05,06	35.5	29	1.627	0.627	0.2207	0.0123
VER	27.65	80.37	06	32.5	57	1.735	0.808	0.2567	0.0121
BWP	26.87	80.07	05,06,07	32.8	48	1.715	0.767	0.2488	0.0126
WPA	26.68	80.05	06	31.7	40	1.712	0.751	0.2476	0.0118
					Mean	1.691	0.746	0.2391	0.0122
					s.d.	0.044	0.070	0.0166	
					CV	2.608	9.343	6.9626	
O. equestris									
PNC	29.08	80.93	06	33.9	37	1.731	0.761	0.2883	0.0140
RBR	28.90	80.85	05,06	35.5	40	1.726	0.768	0.2969	0.0142
VER	27.65	80.37	06	32.5	45	1.710	0.761	0.2791	0.0143
BWP	26.87	80.07	05,06,07	32.8	36	1.738	0.768	0.2987	0.0139
WPA	26.68	80.05	06	31.7	39	1.676	0.711	0.2786	0.0142
					Mean	1.716	0.754	0.2883	0.0141
					s.d.	0.025	0.024	0.0095	
					CV	1.441	3.208	3.2944	

Abbreviation: CV, coefficient of variation.

Diversity statistics estimated from 4% error AFLP data sets include band richness after rarefaction to N=29, Br(29); Percent polymorphic loci based on band phenotypes after standardization to N=29 with 1% threshold, PLP(29)1%; expected heterozygosity and its s.e. (H_e). The mean, s.d. and CV of each diversity statistic were also presented.

accompanied by larger mean mismatch error yield greater information about population differentiation than error-free data sets including fewer loci. Improvements in performance gained by including some error-prone loci are likely to be limited, however, if error-generated noise ultimately overwhelms the signal in the data. Most AFLP studies report average mismatch error rates around 2%, so the most common protocol is not to use error-free data, but to arbitrarily determine a threshold for data to analyze and interpret.

Our goal was to devise methods for empirically identifying genotyping error and to evaluate its effects on population genetic inferences. The known genetic cline in *C. virginica* provided some expectations for evaluating error effects, making it informative to compare results using nested data sets with different levels of mean among-locus mismatch error. Our results suggested that both high- and low-error data sets gave biased views of population structure. Although evaluating the consequences of error will be more subjective in the absence of *a priori* knowledge about substructure patterns, our methods for minimizing genotyping error and evaluating signal/error tradeoffs involved two approaches that can be applied in any study: (i) test for substructure patterns that are not expected in nature (for example, plate effects) and (ii) evaluate trends in population genetic summary statistics across nested data sets with a range of mean error rates measured in duplicates. We suggest that a subjective and transparent analysis of genotyping error effects is better than no analysis at all because it makes uncertainty in the results more explicit.

Plate-level effects can be caused by factors out of the researcher's control such as slight band shifts in individual runs of a genetic analyzer. Plate effects in both *C. virginica* and *O. equestris* seemed to be manifest in data sets having average mismatch error near 5% or more, whereas all population substructure patterns which were found using data with 3 and 4% average mismatch error were consistent with biological expectations or interpretations. Thus, our first error detection and evaluation strategy yielded a clear upper bound for the average mismatch error rates needed to avoid plate effects.

To create data sets differing in overall genotyping error (using our 'primary binset' as a starting point), we applied the locus selection and phenotype-calling criteria in scanAFLP (Herrmann *et al.*, 2010) before progressively culling loci with the highest mismatch error as measured in duplicate samples. This approach contrasts with AFLPSCORE (Whitlock *et al.*, 2008) where the interaction of locus selection and phenotype scoring thresholds are plotted with respect to average genotyping error in replicate samples, but no further culling of loci is perfomed after choosing a set of thresholds. We suspect that the skewed distribution of locus-specific error rates observed here is not unusual, compromising average genotyping error as a sole criterion for choosing a final data set. Others have noted the same thing, for example, by removing 'singleton' loci from the optimum data set as judged with AFLPSCORE (Crawford *et al.*, 2011).

The progressive culling of high-error loci ultimately produced data sets with relatively low information content. The lowest error *C. virginica* data sets (0–2% error) failed to detect known population structure using STRUCTURE, although the regional break was evident from AMOVA even in the 0 error data. This low power to detect regional substructure with assignment tests is at least partly attributable to a limited number of loci being included (28, 92 or 138 loci). However, the positive association between mismatch error and F_{ST} among individual loci suggests that low-error data sets may have been biased for low F_{ST}, at least in *C. virginica*. Given that mismatch error was the primary metric used for reducing the among-locus average error, and with this metric moderate allele frequencies tended to show higher error, culling high-error loci also biased the allele frequency spectrum. The resulting low-error data sets were depleted of loci with moderate average allele frequencies. This might ordinarily be expected to increase inferred population structure, not decrease it, for the same reasons that apply to SNP ascertainment biases (Morin *et al.*, 2004). Indeed, the zero error data set for *O. equestris* shows slightly higher mean F_{ST} relative to other data sets. However, with balanced sampling across an allele frequency cline as in *C. virginica*, moderate allele frequencies within the combined samples are expected to be associated with high F_{ST}. In addition, skewing allele frequency distributions away from moderate frequency (high H_e) loci in low-error data sets will lower the power of assignment tests beyond what is expected based on the same number of unbiased loci. Thus, achieving low mismatch error by locus culling does not necessarily lead to accurate inferences, but can introduce a bias on F_{ST} in either direction.

Homoplasy is a well-documented artifact of scoring dense fragment profiles (Vekemans *et al.*, 2002) with stronger impacts expected on smaller fragments (Althoff *et al.*, 2007; Caballero *et al.*, 2008). Simulated metapopulations with $\bar{p} = 0.1$ studied by Caballero *et al.* (2008) showed homoplasy effects to include a downward bias on F_{ST} and an upward bias on both H_e and $\bar{p}$, each bias monotonically increasing with shorter fragment size. In this study, a negative correlation between fragment size and band-present phenotype frequency indicated some homoplasy effects in all fragment size classes. For unknown reasons band-present frequency distributions transitioned abruptly between fragments <300 versus >300 bp in *C. virginica* (Figure 2). The Caballero *et al.* (2008) predictions for homoplasy effects were seen much more strongly in loci <300 versus >300 bp in 4% error *C. virginica* data; the loci <300 bp had $\bar{p}$ five times larger and H_S three times larger than loci >300 bp. In contrast to expectations, however, among-locus average F_{ST} was twice as high for short fragments (0.0409 vs 0.0212, both significantly different than zero at $P<0.001$). In fact, this F_{ST} discrepancy between short and long fragments was an order of magnitude for 0 error (0.0114 vs 0.002, $P<0.006$ for both) and 1% error data, and smallest for matrixB (0.0529 vs 0.0424). The differences in average F_{ST} do not appear to be a function of genomic sampling because the number of loci in short (11–131) vs long (17–122) fragment subsets was similar across the data sets. The *O. equestris* data showed the same differences between short and long fragments and the same bias trends, although in this case the fragments >300 bp in each data set had F_{ST} that was not statistically different from zero, whereas the shorter fragments had F_{ST} similar to that shown in Figure 3. We do not have an explanation for this disagreement with simulation-based predictions, but note that many of the highest F_{ST} loci among small fragments showed clinal patterns in agreement with previous codominant data (data not shown). These loci clearly have useful information content despite indications of strong homoplasy effects at that fragment size class. In fact, excluding loci <300 bp for fear of homoplasy effects would have removed all but one locus with $F_{ST} > 0.1$.

By relying on duplicate samples for automated binset optimization and subsequent marker selection, this study used AFLP analysis methods that can facilitate scaling up to much larger samples. Although this approach is atypical, the application of Peakmatcher here produced a similar number of initial loci (229 in the primary binset) as previously achieved by manual binning (226 loci) of the same four primer pairs in different Florida samples of *C. virginca* (Murray and Hare, 2006). The use of duplicate fingerprints for binset optimization, rather than requiring that bins be defined based on all samples, requires replication of a substantial fraction of representative samples (>10%). This investment also improves estimates of locus-specific mismatch error rates for accurate ranking of loci and high-error culling. A new tool for automated binset optimization was recently described along with novel descriptive statistics for evaluating the ‘information content’ of bins without the benefit of duplicate genotypes (Arrigo *et al.*, 2009); comparing this approach to the optimization methods applied here will be a valuable next step for advancing the utility of automated binsetting.

Regardless of methods used for binset creation, we recommend using liberal locus acceptance criteria initially, producing genotypes for many loci with a high variance in locus-specific mismatch error rates. Subsequent removal of high-error loci should be empirical, proceeding until detectable artifactual substructure is eliminated. Then a nested series of data sets that vary with respect to average mismatch error can be evaluated with respect to population substructure. Trends in the degree and pattern of substructure across data sets can be used to justify data set choice and evaluate robustness of biological inferences. For biologically plausible patterns of structure, we assumed that the data set showing maximum divergence provides both accuracy and optimum power for detecting subtle patterns. Study objectives and the degree of independent knowledge on substructure patterns will determine how conservative the data set choice should be. Even for population-level analyses that account for genotyping error in parameter estimates (Foll *et al.*, 2010), the comparative approach described here can help empirically evaluate unpredictable consequences of genotyping error.

Relating population structure to larval dispersal

It is widely recognized that both habitat preference and life history can be important factors in determining larval dispersal capacity in sedentary benthic species. For example, estuarine species usually show lower gene-flow potential than comparable demersal marine species (Bilton *et al.*, 2002). Also, species with direct development often show more population structure than species with planktonic larvae (reviewed in Pelc *et al.*, 2009).

Higher population differentiation in *C. virginica* than *O. equestris* suggests the latter species has higher dispersal potential or lower gene flow constraints along eastern Florida. However, co-distributed species don’t necessarily share the same historical processes leading to today’s patterns. Furthermore, different habitat use by co-distributed species can lead to differences in population connectivity. For these two oyster species there are several plausible causes of their different degrees of population substructure.

First, lower tolerance of marine salinities may constrain gene flow for *C. virginica* relative to *O. equestris*. Effective larval dispersal may be largely confined within lagoons for *C. virginica* and be relatively ‘open’ along the continental shelf for *O. equestris*. There are hydrodynamic scenarios that could limit larval dispersal along the continental shelf connecting populations north and south of Cape Canaveral (Hare *et al.*, 2005), but results here indicate that *O. equestris* experiences no major barrier.

The second potential factor affecting dispersal propensity is larval life history. Although *O. equestris* has internal fertilization and broods young, veliger larvae are released after several days (Menzel, 1955). We are not aware of any precise data on planktonic larval duration for *O. equestris*, but 2 weeks is typical of the genus (Castanos *et al.*, 2005) and this is similar to planktonic duration of *C. virginica* in southern waters. Still, by brooding its larvae for several days during early development when physiological tolerances are most constrained (Wright *et al.*, 1983), *O. equestris* may increase larval survivorship and long distance dispersal potential.

The third important factor in this comparison is historical. The phylogeographic break and clinal allele frequencies in *C. virginica* at Cape Canaveral were speculated to be a result of secondary contact (Reeb and Avise, 1990; Hare and Avise, 1996). Results from this study using new samples and novel markers indicate that this regional break has remained strong for nearly two decades. Here, as in the AFLP analysis of oysters by Murray and Hare (2006), there is only a small minority of *C. virginica* loci showing strong differentiation, many of them with geographically coincident clines. It is possible that the two oyster species compared here experience similar contemporary barriers to gene flow near Cape Canaveral, but effective population sizes were too large to promote much differentiation by genetic drift and it is only in the case of *C. virginica* that pre-contact north/south differences remain in sharp contrast across the secondary contact zone.

Despite the regional contrast between these two oyster species, it is interesting that they showed a geographically concordant pattern of subtle population structure. In both species the BWP population is distinguished from the other two southern populations by having a slightly different admixture pattern. There is no detectable difference in DNA quality between BWP and other accessions, nor other technical artifacts identified that could explain this concordance. There appears to be ample opportunity for strong genetic drift with BWP because it is from a narrow section of intracoastal waterway cut between Lake Worth and Palm Beach Inlet to the south and Jupiter Inlet to the north.

DATA ARCHIVING

Data have been deposited at Dryad: doi:10.5061/dryad.4f64ck01.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work benefitted from field collection efforts by Nate Bendik, logistical field support by John Scarpa and laboratory support by Harmony Borchardt-Wier, for which we are thankful. Comments on earlier drafts from Hare Lab members and three anonymous reviewers helped improve this manuscript. This work was funded by NSF Bio-OCE 0648528 to MPH.

Althoff DM, Gitzendanner MA, Segraves KA (2007). The utility of amplified fragment length polymorphisms in phylogenetics: a comparison of homology within and between genomes. *Syst Biol* **56**: 477–484.

Arrigo N, Tuszynski JW, Ehrich D, Gerdes T, Alvarez N (2009). Evaluating the impact of scoring parameters on the structure of intra-specific genetic variation using RawGeno, an R package for automating AFLP scoring. *BMC Bioinformatics* **10**: 33.

Bilton DT, Paula J, Bishop JDD (2002). Dispersal, genetic differentiation and speciation in estuarine organisms. *Estuar Coast Shelf Sci* **55**: 937–952.

Bonin A, Bellemain E, Eidesen PB, Pompanon F, Brochmann C, Taberlet P (2004). How to track and assess genotyping errors in population genetics studies. *Mol Ecol* **13**: 3261–3273.

Bonin A, Ehrich D, Manel S (2007). Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Mol Ecol* **16**: 3737–3758.

Caballero A, Quesada H, Rolan-Alvarez E (2008). Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. *Genetics* **179**: 539–554.

Castanos C, Pascual MS, Agulleiro I, Zampatti E, Elvira M (2005). Brooding pattern and larval production in wild stocks of the puelche oyster, *Ostrea puelchana* D’orbigny. *J Shellfish Res* **24**: 191–196.

Coart E, Van Glabeke S, Petit RJ, Van Bockstaele E, Roldan-Ruiz I (2005). Range wide versus local patterns of genetic diversity in hornbeam (*Carpinus betulus* L. *Conserv Genet* **6**: 259–273.

Crawford LA, Desjardins S, Keyghobadi N (2011). Fine-scale genetic structure of an endangered population of the Mormon metalmark butterfly (*Apodemia mormo*) revealed using AFLPs. *Conserv Genet* **12**: 991–1001.

DeHaan LR, Belina RAK, Ehlke NJ (2002). Peakmatcher: software for semi-automated fluorescence-based AFLP. *Crop Sci* **42**: 1361–1364.

Duchesne P, Bernatchez L (2002). AFLPOP: a computer program for simulated and real population allocation, based on AFLP data. *Mol Ecol Notes* **2**: 380–383.

Foll M, Fischer MC, Heckel G, Excoffier L (2010). Estimating population structure from AFLP amplification intensity. *Mol Ecol* **19**: 4638–4647.

Hare MP, Avise JC (1996). Molecular genetic analysis of a stepped multilocus cline in the American oyster (*Crassostrea virginica*). *Evolution* **50**: 2305–2315.

Hare MP, Guenther C, Fagan WF (2005). Nonrandom larval dispersal can steepen marine clines. *Evolution* **59**: 2509–2517.

Hedrick PW (2005). *Genetics of Populations* 3rd edn Jones and Bartlett: Boston..

Herrmann D, Poncet BN, Manel S, Rioux D, Gielly L, Taberlet P *et al.* (2010). Selection criteria for scoring amplified fragment length polymorphisms (AFLPs) positively affect the reliability of population genetic parameter estimates. *Genome* **53**: 302–310.

Hoese HD (1960). Biotic changes in a bay associated with the end of a drought. *Limnol Oceanogr* **5**: 326–336.

Holland BR, Clarke AC, Meudt HM (2008). Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. *Syst Biol* **57**: 347–366.

Karl SA, Avise JC (1992). Balancing selection at Allozyme Loci in oysters—implications from nuclear Rflps. *Science* **256**: 100–102.

Kirkendale L, Lee T, Baker P, Foighil DO (2004). Oysters of the Conch Republic (Florida Keys): a molecular phylogenetic study of *Parahyotissa mcgintyi*, *Teskeyostrea weberi* and *Ostreola equestris*. *Malacologia* **46**: 309–326.

McDonald JH (2009). *Handbook of Biological Statistics* 2nd edn Sparky House Publishing: Baltimore, Maryland.

Menzel RW (1955). Some phases of the biology of *Ostrea equestris* say and a comparison with *Crassostrea virginica* (Gmelin). *Pubi Inst Mar Sci Univ Tex* **4**: 69–153.

Meudt HM, Clarke AC (2007). Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends Plant Sci* **12**: 106–117.

Morin PA, Luikart G, Wayne RK (2004). SNPs in ecology, evolution and conservation. *Trends Ecol Evol* **19**: 208–216.

Murray MC, Hare MP (2006). A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. *Mol Ecol* **15**: 4229–4242.

Peakall R, Smouse PE (2006). GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. *Mol Ecol Notes* **6**: 288–295.

Pelc RA, Warner RR, Gaines SD (2009). Geographical patterns of genetic structure in marine species with contrasting life histories. *J Biogeogr* **36**: 1881–1890.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Reeb CA, Avise JC (1990). A genetic discontinuity in a continuously distributed species - mitochondrial-DNA in the American oyster, *Crassostrea virginica*. *Genetics* **124**: 397–406.

Rosenberg NA (2004). Distruct: a program for the graphical display of population structure. *Mol Ecol Notes* **4**: 137–138.

Shumway SS (1996). Natural environmental factors. In: Kennedy VS, Newell RIE, Eble AF (eds) *The Eastern Oyster, Crassostrea Virginica*. Maryland Sea Grant, University of Maryland System: College Park, MD. pp 467–513.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002). Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Mol Ecol* **11**: 139–151.

Vos P, Hogers R, Bleeker M, Reijans M, Vandelee T, Hornes M *et al.* (1995). Aflp - a new technique for DNA-fingerprinting. *Nucleic Acids Res* **23**: 4407–4414.

Whitlock R, Hipperson H, Mannarelli M, Butlin RK, Burke T (2008). An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Mol Ecol Notes* **8**: 725–735.

Wright DA, Kennedy VS, Roosenburg WH, Castagna M, Mihursky JA (1983). Temperature tolerance of embryos and larvae of five bivalve species under simulated power plant entrainment conditions: a synthesis. *Mar Biol* **77**: 271–278.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)