3-28-2014

Hybridization between previously isolated ancestors may explain the persistence of exactly two ancient lineages in the genome of the oyster parasite Perkinsus marinus

Peter C. Thompson
University of Maryland, pete.c.thompson@gmail.com

Benjamin M. Rosenthal
United States Department of Agriculture

Matthew P. Hare
University of Maryland

Infection, Genetics and Evolution 24 (2014) 167–176

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Hybridization between previously isolated ancestors may explain the persistence of exactly two ancient lineages in the genome of the oyster parasite *Perkinsus marinus*

Peter C. Thompson [a,*], Benjamin M. Rosenthal [b], Matthew P. Hare [a,1]

[a] University of Maryland, Department of Biology, 1210 Biology-Psychology Bldg, College Park, MD 20742, USA
[b] Animal Parasitic Diseases Lab, Beltsville Agricultural Research Center, Agricultural Research Service, United States Department of Agriculture, Beltsville, MD 20705, USA

ARTICLE INFO

Article history:
Received 12 July 2013
Received in revised form 18 February 2014
Accepted 4 March 2014
Available online 28 March 2014

Keywords:
Dimorphism
Hybridization
Perkinsus
Coalescent
Parasite
Evolution

ABSTRACT

Theory predicts that neutral genetic variation accumulates within populations to a level determined by gains through mutation and losses by genetic drift. This balance results in a characteristic distribution of allelic variation with the maximum allelic difference determined by effective population size. Here, we report a striking departure from these expectations in the form of allelic dimorphism, observed at the majority of seven loci examined in *Perkinsus marinus*, an important oyster parasite that causes Dermo disease. DNA sequences were collected from five loci flanking microsatellite repeats and two loci coding for superoxide dismutase enzymes that may mediate the parasite's interaction with its host. Based on 474 sequences, sampled across 5000 km of the eastern United States coastline, no more than two alleles were observed at each locus (discounting singletons). Depending on the locus, the common allele ranged in overall frequency from 72% to 92%. At each locus the two alleles differed substantially (3.8% sequence difference, on average), and the among-locus variance in divergences was not sufficient to reject a simultaneous origin for all dimorphisms using approximate Bayesian methods. Dimorphic alleles were estimated to have diverged from a common ancestral allele at least 0.9 million years ago. Across these seven loci, only five other alleles were ever observed, always as singletons and differing from the dimorphic alleles by no more than two nucleotides. Free recombination could potentially have shuffled these dimorphisms into as many as 243 multilocus combinations, but the existence of only ten combinations among all samples strongly supports low recombination frequencies and is consistent with the observed absence of intragenic recombination. We consider several demographic and evolutionary hypotheses to explain these patterns. Few can be conclusively rejected with the present data, but we advance a recent hybridization of ancient divergent lineages scenario as the most parsimonious.

1. Introduction

Measures of genetic diversity are correlated with the adaptive capacity of populations (Reed and Frankham, 2001), and different patterns of variation within and among populations can illuminate the evolutionary processes shaping variation. Within interbreeding populations, observed neutral alleles represent extant lineages that trace back through the genealogy to the most recent common ancestral allele (MRCA), a process that can be modeled by coalescent theory (Rosenberg and Nordborg, 2002). According to this theory, the expected depth of a genealogy (theta) is proportional to the effective population size (Ne) and the mutation rate (μ). In a population of constant size, DNA sequences from contemporary samples "coalesce" to a common ancestor back through time following a geometric pattern, with frequent coalescences in the recent past but progressively longer waiting times (genealogical branch lengths) between coalescent nodes toward the MRCA. Although a large variance in genealogical shape and depth may occur among independent neutral loci in the same population, the neutral expectation is for half the genealogical depth to be defined by the oldest two lineages before they coalesce at the MRCA.

One of the most intriguing patterns observed repeatedly in pathogen populations is a deviation from these neutral expectations, termed allelic dimorphism, where the variation at a locus is distributed primarily between two highly diverged allelic classes

* Corresponding author. Address: Animal Parasitic Diseases Lab, USDA-ARS, Bldg 1180 BARC-East, Beltsville, MD 20705, USA. Tel.: +1 301 466 4174; fax: +1 301 504 5306.
E-mail address: pete.c.thompson@gmail.com (P.C. Thompson).
[1] Current address: Department of Natural Resources, 208 Fernow Hall, Cornell University, Ithaca, NY 14853, USA.

http://dx.doi.org/10.1016/j.meegid.2014.03.003
1567-1348/Published by Elsevier B.V.

(Roy et al., 2008; Sibley and Ajioka, 2008). Where allelic dimorphism is observed, sequence variants at a locus (haplotypes) show a pattern of genealogical relatedness in which only two classes of alleles are encountered and those classes are much more divergent from each other than seen for any allelic comparison within either class. Genealogically, this appears as a pair of long branches separating the two allelic classes (clades), with the branch lengths too long relative to the overall genetic diversity to be explained by neutral coalescent variance.

Explanations for allelic dimorphism have been varied, including clonal population structure and selection for antigen diversity within pathogens. As summarized by Roy et al. (2008), few hypotheses provide satisfactory explanations for all the typical features of allelic dimorphism: (1) ancient allelic classes with relatively low intra-class diversity, (2) no inter-class recombination, (3) never more than two allelic classes, and (4) dimorphism restricted to certain functional categories of loci. Given that standing levels of genetic diversity can strongly influence the rate of a pathogen's co-evolutionary adaptation to the host (Frank, 1996), it is important to develop and test more mechanistic hypotheses for genetic dimorphism in pathogens.

Perkinsus marinus is a protistan parasite that causes substantial mortality across most of the geographic range of the eastern oyster, *Crassostrea virginica*. This host species is the subject of a commercial fishery worth over $100 million in 2012 (U.S. Department of Commerce, 2012), and can also provide economic benefits of more than $10,000 per hectare each year through shore stabilization, providing habitat for other valuable organisms, and improving water quality (Grabowski et al., 2012). *P. marinus* prevalence and abundance varies widely by location (Burreson and Ragone Calvo, 1996; Craig et al., 1989), with increased disease associated with higher water temperature and salinity (Burreson and Ragone Calvo, 1996; Mackin, 1956; Soniat, 1985). When growth conditions for the parasite are optimal, oyster mortality may exceed 95% in the second year of infection (Albright et al., 2007; McCollough et al., 2007). Prior to the 1980s, *P. marinus* was believed to be limited to estuaries south of Chesapeake Bay, but expanded its range 600 km over 10 years and is now routinely observed in locations as far north as Maine (Ford, 1996). Laboratory tests suggest that virulence of this ecologically and economically important pathogen varies regionally, as does host resistance (Bushek and Allen, 1996).

P. marinus belongs to an early branching dinoflagellate lineage (Saldarriaga et al., 2003; Siddall et al., 1997) and is most closely related to the congeners *Perkinsus olseni* and *Perkinsus honshuensis* (Dungan and Reece, 2006; Moss et al., 2008). Recent studies using microsatellite loci have inferred that *Perkinsus* is diploid and capable of sexual reproduction, but genotypic diversification across its geographic range is constrained by both asexual reproduction and substantial inbreeding (Thompson et al., 2011; Vilas et al., 2011). However, diverse *P. marinus* multilocus microsatellite lineages were observed in some locations (Thompson et al., 2011), providing opportunities for genetically distinct parasite strains to compete within local host populations and drive coevolutionary dynamics (Frank, 1996). In other locations where this pathogen recently expanded its range, population genetic patterns were suggestive of hybridization between distinct strains (Thompson et al., 2011).

In single gene DNA sequencing studies of *P. marinus* nuclear loci, an interesting pattern of large divergence between haplotypes has been observed (ranging from 1.5% to 3.6% maximal difference, Brown et al., 2004; Reece et al., 1997; Robledo et al., 1999), implying a large effective population size for this pathogen. Some loci were characterized by multiple haplotypes (Brown et al., 2004), while others were limited to only two (Reece et al., 1997; Robledo et al., 1999). A seemingly dimorphic pattern was also found at microsatellite loci as reflected in bimodal frequency distributions of simple tandem repeat lengths (Thompson et al., 2011). Microsatellite loci are often considered to be selectively neutral, but may contribute to regulatory function in genomes or may be linked to other functional loci under selection (Selkoe and Toonan, 2006). Therefore, the extent of nucleotide divergence among these simple tandem repeat length alleles deserves further scrutiny to rigorously test for dimorphic patterns and characterize the genomic extent of this pattern.

In this study, nucleotide sequences were obtained from seven *P. marinus* nuclear loci, each amplified by polymerase chain reaction (PCR) from infected oyster tissue from wild oysters collected from Massachusetts to Texas. Flanking sequences were collected from five microsatellite loci, and coding and intron sequences were collected from two superoxide dismutase genes (SOD1 and SOD2), loci potentially involved in responses to host defense mechanisms (Schott et al., 2003). Given that genotypic fragment analysis with these microsatellite loci demonstrated mixed patterns of sexual and clonal reproduction across many *P. marinus* populations (Thompson et al., 2011), recombination was investigated both within and between loci to determine whether haplotypes have remained associated in the same breeding population for an extended period of time. Estimates of coalescence time among alleles were used to estimate their age. Finally, to distinguish among potential mechanisms shaping allelic diversity in the *P. marinus* genome, we examined whether individual loci followed neutral coalescent patterns, and tested whether the collection of dimorphic loci are consistent with a single evolutionary origin for the two allelic classes.

2. Materials and methods

2.1. Samples and loci

PCR amplification conditions were optimized using clonally cultured isolates from Mozambique Point, LA (ATCC #50763), Fort Pierce, FL (ATCC #50775), Bennet Point, MD (ATCC #PRA-240), Delaware Bay, NJ (ATCC #50509), and Narragansett Bay, RI (HCTR, graciously provided by Marta Gomez-Chiarri). Additionally, four clonal isolates of other *Perkinsus* species (*Perkinsus chesapeaki* (ATCC #50866, ATCC #50864, and ATCC #50807) and *P. olseni* (ATCC #PRA-31)) were used to attempt amplifying outgroup sequences.

Genomic DNA was extracted from large samples of oysters collected at 15 geographic locations across the known *P. marinus* range, and molecular assays were used to test for *P. marinus* infection. Six infected oysters from each geographic collection were selected for PCR amplification and sequencing of microsatellite flanking regions (Table 1). These loci were developed from 40 microsatellite-containing contigs identified in the *P. marinus* draft genome (Refseq: NZ_AAXJ00000000). After evaluating all 40 loci, seven loci were chosen for further study based on their reliable amplification and variable numbers of microsatellite repeats (Thompson, 2010). Flanking regions from one side of five of these microsatellite loci were Sanger sequenced on both strands (Table 1). Microsatellite flanking region primers are in Table 2 and the size of each analyzed flanking region is in Table 3. Only host individuals that had been judged by multilocus microsatellite analysis to contain DNA from exactly one strain of *P. marinus* were included for present purposes, according to methods described in Thompson et al. (2011). Briefly, based on genotypes at seven highly variable microsatellite loci, if three or more alleles were observed at any locus, the individual was considered to be infected by more than one *P. marinus* individual. Given the observed variability in microsatellite repeats (12–18 alleles per locus), the probability of

Table 1
Sampling details. The first two columns provide collection locality and its abbreviation (Abbrev) used in figures. Number of individuals (*N*) reports total attempts to collect a full sequence data set, including singleton clonal cultures (*). The number of successful sequencing reactions for each locus is provided in the remaining columns. Note that samples used for the coding regions were a subset of those for the microsatellite loci.

Collection Location	Abbrev	*N*	Microsatellite Flanking Regions					Coding Loci		
			Pm2232	Pm2903	Pm2988	Pm4488	Pm8517	SOD1-short	SOD2-short	SOD2-long
Dickinson, TX	DTX	6	3	4	5	5	3			
Sabine Lake, TX	TLB	6	2	4	4	5	3		1	
Mozambique Point, LA	LA	1*	1	1	1	1	1	1	1	
Dauphin Island, AL	ALA	6	5	3	3	2	1			
Apalachicola, FL	APA	6	6	5	6	6	6			
Port Charlotte, FL	PCH	6	5	4	4	5	5	4	26	
St. Lucie River, FL	SLR	6	6	4	6	6	6	2		
Fort Pierce, FL	FTP	7*	5	4	7	6	6	4	1	1
Cape Canaveral, FL	LPA	6	4	5	5	6	4	3	11	
New Smyrna Beach, FL	NSB	6	4	4	3	5	4	2	13	
Port Orange, FL	POR	6	6	6	6	4	6	3	4	
Georgetown, SC	GSC	6	1	1	3	0	0			
Skidaway Island, GA	SKW	6							5	
Bennet Point, MD	MD	1*	1	1	1	1	1	1	1	
Delaware Bay, NJ	DBA	7*	5	5	5	5	4	1	1	1
Narragansett Bay, RI	NRI	7*	7	6	7	7	6			
Tisbury, MA	TMA	6	6	6	6	6	6			
Edgartown, MA	EMA	6	5	5	4	4	5			

Table 2
Primers for amplification and sequencing. Primer names are denoted first by locus name, then by specifics regarding the location within a locus. "F" and "L" refer to forward primers, and "R" refers to reverse primers. Internal primers for nesting were generally given the designation "int." SOD1 primers additionally have numbers indicating the starting position of the primers in the gene sequence provided in Schott et al. (2003).

Locus	Reaction	Primer Name	Sequence (5′–3′)	Tm	Amplicon size
Pm2232	Primary	Pm2232seqF1	tcaacggagcttttctcgat	60	1016
		Pm2232seqR1	tcgagtacactggcagcatc	60	
	Secondary	Pm2232seqFint	ccaggacatatcgtaaagca	56	124
		Pm2232seqR2	tacgctcgcagtgaccatag	60	
Pm2903	Primary	Pm2903seqF2	cagggagccatacgacttgtc	62	1031
		Pm2903Rseq	ctacccaccctatggctgaa	60	
	Secondary	Pm2903seqF	cacgtggtgtcgcatatttc	60	215
		Pm2903seqRint	aacatggcttgggtgatagg	57	
Pm2988	Primary	Pm2988SeqF2	aggcttcaacgcttccaata	60	869
		Pm2988seqR2	gaaatgaatcccggaaaggt	60	
	Secondary	Pm2988SeqF1	tcgatctttgcttgacatcg	60	267
		Pm2988seqRint	gaacgcgtttaacccatagc	60	
Pm4488	Primary	Pm4488seqF1	gacttgatcccttcggtgaa	60	803
		Pm4488seqR1	cctcgctctcattcttggtc	60	
	Secondary	Pm4488seqF1		60	321
		Pm4488seqRint	agctttggagacgtcgttgt	60	
Pm8517	Primary	Pm8517seqF2	gtcaaacggagacgtgtgtg	60	1222
		Pm8517seqR2	cgagcctacgaccaactttc	60	
	Secondary	Pm8517seqFint	ctggaccaggtcaggtttgt	60	405
		Pm8517seqR2		60	
SOD2 short	Primary	SOD2L	ggggagaatgtgttcaatgc	60	442
		SOD2R	gcctttcgcatgaagttctg	60	
	Secondary	SOD2L		60	317
		SOD2Rint	tgccatacaaagacgaggaa	60	
SOD2-Full Length	5′ amplicon	SOD2-5′d	tgatcggagacacgtactgc	60	994
		SOD2R			
	3′ amplicon	SOD2L		60	994
		SOD2-3′c	cgtgtcaccacaattcaac	60	
	Sequencing	SOD2-5′d		60	
	5′ amplicon	SOD2-x1L	tttcgtcccttcaaggtctg	60	
		SOD2-x2R	ggatgtaaccctggtgatgc	60	
		SOD2R		60	
	3′ amplicon	SOD2L		60	
		SOD2-x5R	cgcaagtgaagagtggaat	60	
		SOD2-3′c		60	
		SOD2-x4L	ggggaacgttgatgagatga	60	
		SOD2-AtlL1	gggtcgaccacatatgttgcg	60	
		SOD2-GulfL1	agggtgatctcatatgttggt	60	
SOD1-short	Primary	SOD1-70L	cgttgcttctcgtcagtcac	60	661
		SOD1-731R	actctcccttatcgggttgc	60	
	Secondary	SOD1-290L	actctaccatcgccagcaag	60	441
		SOD1-731R	actctcccttatcgggttgc	60	

Table 3
Summary of sequence polymorphism, heterozygosity, and neutrality test results for all samples combined. Chromosome number assumes diploidy. Singletons are haplotypes observed only once. No tests of selection were significant ($p > 0.05$).

Locus	Pm2232	Pm2903	Pm2988	Pm4488	Pm8517	SOD2 long	SOD2 short	SOD1-short
Sequence polymorphism								
# Chromosomes sequenced	142	142	142	146	134	4	158	44
Total length	91	181	232	222	298	1507	257	273
# Differing haplotypes	3	3	2	2	3	2	4	2
Singletons	1	1	0	0	1	0	2	0
Segregating sites (S)	2	4	5	11	10	82	21	6
Nucleotide Diversity (pi)	0.004	0.007	0.002	0.012	0.011	0.036	0.056	0.006
Max % difference	2.2	2.7	2.2	5.9	2.7	5.5	7.4	3.3
Heterozygosity								
Observed	0.014	0.077	0.000	0.000	0.092		0.048	0.045
Expected	0.317	0.411	0.138	0.182	0.383		0.294	0.165
Inbreeding coefficient (*F*)	0.954	0.813	1.000	1.000	0.759		0.836	0.725
Tests of selection								
Tajima's D	−0.327	1.229	−0.822	0.692	1.912		1.324	0.127
Fu and Li's D*	−0.987	−0.340	1.008	1.404	−0.117		0.811	1.185
Fu and Li's F*	−0.918	0.206	0.471	1.368	0.721		1.217	1.001

an individual being multiply infected, yet having fewer than three alleles at all seven microsatellite loci was small. Microsatellite simple sequence repeat variation has been reported elsewhere (Thompson, 2010; Thompson et al., 2011).

Two partial segments were also sequenced from the superoxide dismutase 1 and 2 genes (SOD1 and SOD2-short, respectively). These loci provide the ability to compare patterns at functional loci possibly under selection relating to host-pathogen interactions (Schott et al., 2003) versus the putatively neutral microsatellite loci. SOD1 included all of intron 2, all of exon 3, and 25 bases from intron 3. SOD2-short contained sequence only from intron 3. Samples used for sequencing both partial SOD loci were an overlapping subset from the same collections used for sequencing microsatellite flanking regions (details in Table 1). In addition, the full SOD2 coding sequence (SOD2-long) was amplified in two overlapping (diploid) segments from ATCC strains 50775 and 50509.

2.2. PCR

Except for those sequences derived from the clonal *P. marinus* isolates, loci were amplified directly from infected oyster genomic DNA extracts from wild oysters collected from the geographic locations listed in Table 1. For all microsatellites and partial SOD amplicons, primary amplifications were followed by nested or hemi-nested amplifications using diluted amplification product as template. Each 20 µl PCR reaction used a standard amplification protocol with 30 amplification cycles in both the primary and secondary amplification as follows. Each PCR reaction contained 1× PCR buffer (Invitrogen, Carlsbad, CA), 2.5 mM magnesium chloride, 0.25 mM mixed dNTP's, 0.1 mg/ml bovine serum albumin, 0.5 units *Taq* DNA polymerase (Invitrogen), and 40 nM of each primer (Table 2). Reactions were subjected to a standard cycling procedure with an initial denaturation at 94 °C for 2 min followed by 30 cycles of 94 °C for 30 s, 60 °C for 30 s, and extension at 72 °C for 1 min. The cycling was followed by a final extension step at 72 °C for 5 min, and reactions were stored at 4 °C until analyzed. For each 96 well amplification plate, reactions containing no template DNA or uninfected oyster genomic DNA were used to control for cross-contamination and non-specific amplifications, respectively. Two µl of undiluted oyster genomic DNA was used as template in each primary amplification. Genomic DNA template concentration ranged from 40 to 250 ng/reaction. In all cases, 2 µl of the primary amplification were used as template for the secondary reaction.

Amplifications of congeneric *Perkinsus* species were attempted using temperature gradient PCR for all microsatellite flanking sequences and the SOD2 locus, varying annealing temperatures from 40 °C to 60 °C and increasing the annealing time to 1 min per cycle.

2.3. Sequencing

PCR products were Sanger sequenced directly on both strands using Applied Biosystems BigDye v.3.0 chemistry (Life Technologies, Carlsbad, CA) with standard procedures for removing excess primers and sequencing reaction purification (Thompson, 2010). For sequencing we used primers designed for secondary amplification, even in clonal isolates where no secondary amplification was needed. Sequencing of the entire coding sequence of *SOD2* was accomplished using two primers on each strand (Table 2) to guarantee sufficient coverage, and included two allele specific primers designed from SOD2-short sequencing results and targeting the two dimorphic allelic classes (Table 2).

2.4. Alignment

DNA sequence chromatograms were edited and aligned in Sequencher v 4.7 (Gene Codes Corporation, Ann Arbor, MI). Because heterozygous sites were called from direct diploid sequence, we used only sequence portions where both DNA strands produced reliable base calls. Chromatograms with overlapping peaks of approximately equal height at a particular nucleotide position were scored as heterozygous for that nucleotide, even if this pattern was only seen on one strand (disagreement between strands regarding heterozygosity only occurred once). Multiple sequence alignments were then constructed in Sequencher using diploid sequences, manually adjusting the alignment to minimize the number of gaps when necessary.

2.5. Analysis

All DNA sequences were partitioned into two haplotypes, assuming diploidy (Thompson et al., 2011). The number of sequences provided in Table 3 reflects the number of chromosomes sampled (two chromosomes for each contig). When multiple heterozygous sites were observed within diploid sequences, the phase of these polymorphisms was inferred and haplotypes determined using PHASE v 2.1.

Summary statistics were calculated for each locus from all samples pooled using DNAsp v 5.0 (Librado and Rozas, 2009).

Nucleotide variation was estimated as the number of segregating sites, S, and the average number of pairwise nucleotide differences, pi. Tests of neutrality included Tajima's D, Fu and Li's D* and Fu and Li's F*. The number of synonymous substitutions per synonymous site (Ks) and the equivalent for nonsynonymous substitutions (Ka) were calculated according to Nei and Gojobori (1986). This method is conservative with respect to tests of positive selection when transitions outnumber transversions (Nei and Kumar, 2000). A one-tailed Fisher's exact test was used to test whether Ka > Ks, a pattern expected from strong positive selection. Strobeck's *S* statistic was used to test whether the observed number of haplotypes was less than expected, given the nucleotide diversity at each locus and assuming panmixia (Strobeck, 1987). Intragenic recombination was estimated using Hudson's R as implemented in DNAsp.

Observed data were compared against statistical distributions generated under relevant null hypotheses in order to test (1) the plausibility that neutral alleles in a constant sized population lacking subdivision could have produced the observed degree of dimorphism, and (2) compatibility of locus-specific estimates of time to MRCA with a single divergence time between two allelic classes at all seven loci. To test dimorphism for each locus, we compared the inferred length of the longest unrooted gene tree branch with the distribution of longest branch lengths under a neutral coalescent model for a single panmictic, constant sized population. If the longest branch of the observed unrooted gene tree was longer than 95% of the longest branches in neutral genealogies, we rejected neutrality in favor of dimorphic allelic classes separated by the longest branch. For these analyses, coalescent simulations were conducted in mlcoalsim v 1.42 (Ramos-Onsins and Mitchell-Olds, 2007). For each locus, 100 independent coalescent simulations generated sequences of equal length and with the same sample size and number of segregating sites as a particular locus. For instance, at locus Pm2903 simulated datasets generated under the neutral coalescent each contained 142 DNA sequences of length 181 nucleotides with 4 segregating sites. For each of the 100 simulated datasets an unrooted neighbor-joining tree was constructed in MEGA 4 based on the number of nucleotide differences, and subsequently, the number of differences along the longest branch in the tree was recorded. The distribution of the number of differences among 100 simulated trees for each locus was used for the statistical test described above.

For loci with only two haplotypes observed (matching the strict definition of dimorphic, that is, having two forms), we used a similar procedure to test for deviation from neutrality using coalescent simulations in mlcoalsim. The probability of observing exactly two haplotypes, given the length of each locus, the number of observed segregating sites, and the number of chromosomes sampled was determined based on the Kw statistic (Strobeck, 1987) from 10,000 coalescent simulations for each locus with only two forms.

In order to examine the possibility that allelic classes at all loci began diverging at the same time, the average length and variance of the longest branch of each gene tree was estimated using mrBayes v 3.2 (Ronquist et al., 2012). For each of the seven observed loci, analyses were conducted using two runs with four chains for 100,000 generations, sampling every 100th tree after a burn-in of 25,000 generations. Chain length and sampling of trees was determined to be adequate as the standard deviation of split frequencies for each mrBayes run was less than 0.01 for all loci, and the Potential Scale Reduction Factor of the longest branch was very close to 1 for all loci (range 0.999–1.006). Branch lengths and their associated 95% highest probability distribution (HPD) interval were generated from the resulting 750 trees.

We tested for simultaneous divergence among loci using MTML-msBayes v 20120510 (Huang et al., 2011). For this analysis, each locus was treated as if representing a species pair, providing seven potential divergence events. In order to conduct this analysis, the haplotypes from either side of a midpoint rooted neighbor joining tree were designated as "species A" and the remaining haplotypes were designated "species B." The observed data from each of seven "species pairs," corresponding to each locus, were used to generate 10,000,000 simulated datasets, restricting the effective population size to between 0.00001 and 1. The probability that one, two, or more divergence events best explained the divergence among all seven loci was estimated from the posterior distribution of the 10,000 simulations (tolerance = 0.001) according to the default parameters in MTML-msBayes.

To estimate time to the most recent common ancestor based on the combined information from all loci, nucleotide mutation models were compared and tested for each locus using FindModel on the web (http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html). The best-fit model for each locus was evaluated by AIC and likelihood. Corrected genetic distances for the length of each gene tree were calculated in MEGA 4 (Tamura et al., 2007). Divergence time was calculated based on the average corrected genetic distance (d) among all loci. Time to the most recent common ancestor (t) was calculated as $t = d/2\mu$ assuming a range of mutation rates (μ) from 2×10^{-8} to 1×10^{-9} substitutions per site per year in accordance with mutation rates reported for other single-celled organisms (Kondrashov and Kondrashov, 2010; Machado and Ayala, 2001).

All previous phylogeographic inference for *P. marinus* has been based on RFLPs or microsatellites where homoplasy can mimic recombination. To compare previous inferences with the sequence-based results here, genotypes were pooled within three geographic regions according to hypothesized subdivision in Reece et al. (2001), corresponding to the Gulf of Mexico, southeastern United States (Atlantic coast south of Maryland), and northeastern United States (this region corresponds to the recent range expansion since the 1980s). Using only those individuals with complete multilocus genotypes (MLGs) from the microsatellite flanking regions, genotypic diversity was quantified using Nei's unbiased estimate of genotypic diversity (Nei, 1987). Regional differences in MLG diversity were tested using bootstrap sampling with replacement as implemented in GenoDive (Meirmans and vanTienderen, 2004) and regional genetic differentiation was tested using AMOVA in Arlequin v 3.01 (Excoffier et al., 2005) with 1000 randomizations of the data. Recombination was tested pairwise between loci using linkage disequilibrium based on likelihood as implemented in GenePop on the Web (Raymond and Rousset, 1995).

3. Results

3.1. Locus specific patterns

All loci amplified successfully in each of the five *P. marinus* clonal cultures. All clonal isolates were homozygous except for those deriving from Maryland and New Jersey, both of which were heterozygous for the same four loci (Pm2903, Pm8517, SOD1-short, and SOD2-short). In spite of multiple attempts to generate outgroup sequence, no other *Perkinsus* species amplified for any locus.

Sequence data were collected for a total of 474 amplicons from all samples, with the final analyzed sequence covering a total of 3928 bases per individual across the seven *P. marinus* loci (Table 3). For the microsatellite flanking loci, 70.4% of sequencing attempts were successful for both strands. Similar to microsatellite genotyping results (Thompson et al., 2011), the distribution of missing sequences varied randomly among samples and locations for each locus, resulting in 54 samples (52%) with complete multilocus genotypes. Only 13 heterozygous genotypes (2.7%) were

encountered among all sequences. For these heterozygous sequences, all PHASE inferences were highly significant ($p < 0.001$), suggesting reliable haplotype determination despite deviations of the data from the Hardy–Weinberg equilibrium assumption.

Allelic diversity was minimal within each locus, but marked divergence occurred between observed alleles (Fig. 1, Table 3). An average of 130 (stdev = 38.5) chromosomes were sequenced at each locus, but no more than 4 alleles were found at any locus. One common (frequency = 72 to 94%) and one alternate haplotype was observed at every locus, with singletons also observed at some loci (Fig. 1). The two major alleles differed in DNA sequence by an uncorrected average of 3.8% (stdev = 2%). The largest allelic difference (7.4%, Table 3) was observed at the 257 bp SOD2-short locus. Locus Pm2232 had the lowest sequence difference (2.2%, Table 3). Insertions or deletions (indels) were observed at two loci, with a three nucleotide indel in Pm4488, and six different indels over a total of 14 nucleotide positions within SOD2 noncoding regions. No locus deviated from neutral expectations by Tajima's D, Fu and Li's D*, or Fu and Li's Fs metrics (Table 3).

For all but two loci, the observed number of haplotypes was much lower than coalescent expectations (two-tailed test of Strobeck's *S*, $p < 0.001$ for loci Pm2988, Pm4488, Pm8517, and SOD2-short, $p < 0.01$ at SOD1), given the observed number of segregating sites and number of chromosomes sampled at each locus.

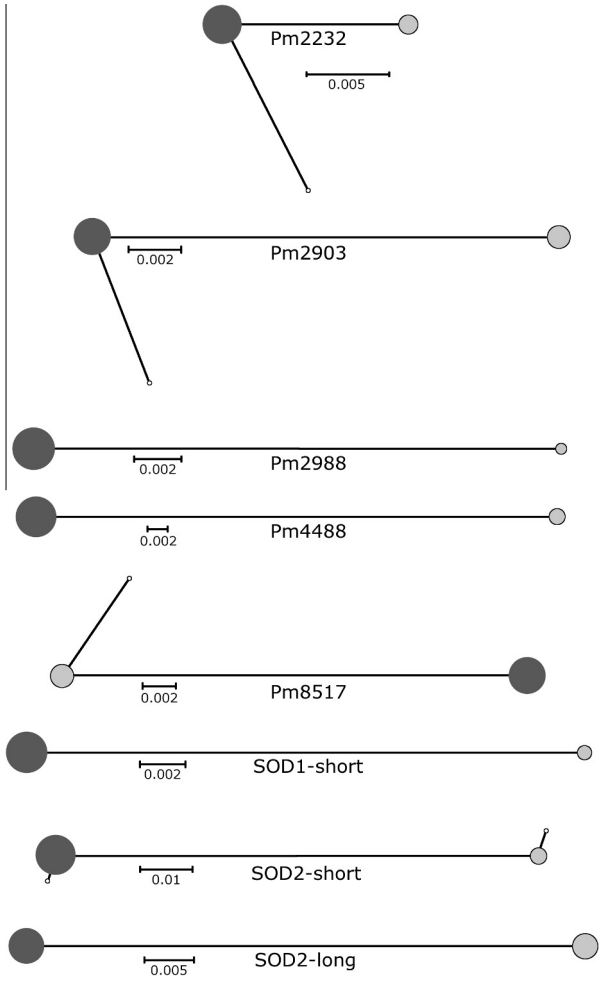

Fig. 1. Unrooted neighbor-joining trees for seven loci based on corrected genetic distances (scale bars). Common haplotypes are represented in black, alternate haplotypes are denoted in gray, and singletons are represented in white. The size of each circle is proportional to the frequency of that haplotype in the entire dataset. Every locus except microsatellite flanking region Pm2232 is characterized by a long internal branch separating two classes of alleles.

Within the SOD1 and SOD2 gene sequences, variation was present in both introns and exons. At SOD1, 2 of 6 single nucleotide polymorphisms (SNPs) were in coding portions of the sequence and both were synonymous. The full SOD2 gene sequence had a total alignment length of 1507 bp. Eighty-two nucleotides (5.5%) were polymorphic, including an overall transition to transversion ratio of 2.0. Exons included 843 bp and contained 32 polymorphisms (3.8%), 15 of which were nonsynonymous and changed a total of 14 amino acids in the inferred polypeptide. Introns contained a total of 648 bp with 50 polymorphic sites (7.7%). One allele from each locus was identical to published SOD1 and SOD2 sequence (GenBank accessions AY13779 and AY137780, respectively).

Treating the two SOD2 long alleles as if they are from different taxa and applying the method of Nei and Gojobori (1986) using the ciliate nuclear genetic code, Ka (0.0248) was significantly smaller than Ks in coding regions (0.0842, $p = 0.0008$) and Ks overall (0.0842), failing to support positive selection by the simplest test of H_1: Ka > Ks.

3.2. Allelic dimorphism

The low haplotype diversity and high allelic divergence imply dimorphism, but to specifically test dimorphism against neutral expectations, coalescent simulations of sequences evolving under a strict neutral model were conducted for statistical comparison with observed patterns. Only two haplotypes were observed at three of the seven loci. Based on coalescent simulations, the probability of observing only two haplotypes was less than 0.001 at Pm2988 and Pm4488, and less than 0.01 for SOD1-short, given the observed number of segregating sites. Among the remaining loci, the observed longest branch length was significantly longer than the distribution of longest branches recovered from neutral simulations for three loci, Pm2903, Pm8517, and SOD2-short ($p = 0.05$, 0.01, and 0.01 respectively), indicating allelic dimorphism. Only Pm2232 matched neutral expectations ($p = 0.4$). Taken together, six of seven loci met this statistical definition of dimorphism.

3.3. Time of divergence

Approximate divergence time between dimorphic allele classes was estimated as the time to MRCA for haplotypes at each short locus. Corrected genetic distances for total gene tree lengths ranged from 0.011 at locus Pm2232 to 0.087 at SOD2-short (Table 4). The mean of all gene tree lengths was 0.036 (stdev = 0.026). Assuming a range of mutation rates and independent evolution of allele classes, the mean gene tree length yielded a range of divergence times between dimorphic alleles from 900,000 years to 18 million years ago.

In order to calculate the probability that dimorphic alleles began diverging simultaneously at all loci, we used mrBayes to esti-

Table 4
Corrected genetic distances for gene tree lengths. Distances were calculated using models determined using AIC criteria in FindModel. K2P = Kimura two parameter model and HKY indicates Hasegawa, Kishino, and Yano six parameter model.

Locus	Model	Distance
Pm2232	K2P	0.011
Pm2903	K2P	0.018
Pm2988	HKY	0.022
Pm4488	HKY	0.052
Pm8517	HKY	0.028
SOD1-short	K2P	0.036
SOD2-short	HKY	0.087

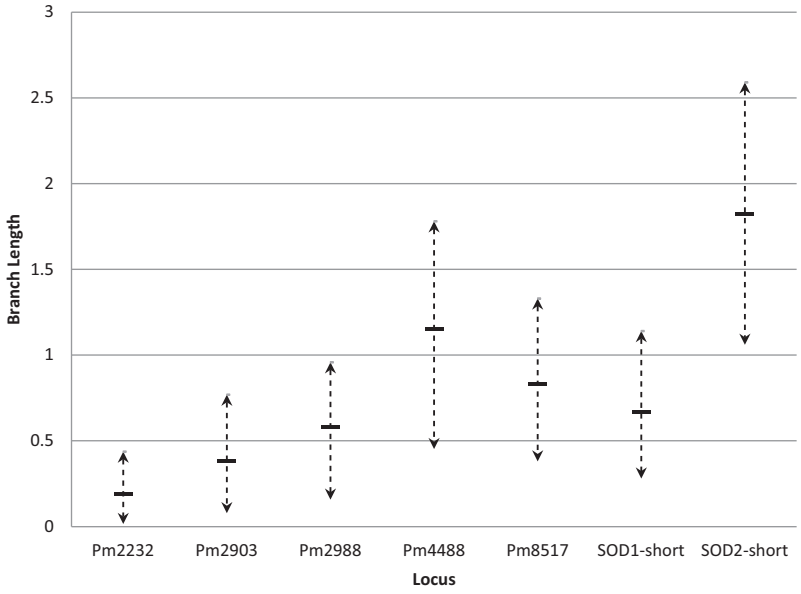

Fig. 2. Comparison of central branch lengths among loci. For each locus, the mean and 95% highest posterior distribution (HPD) interval of the length of the longest branch of the gene tree from each locus is shown. Five out of seven 95% HPDs overlap considerably, consistent with variance among loci that began diverging at the same time. While Pm2232 and SOD2-short were at the extremes of the distribution of HPDs, both loci share 95% HPD overlap with at least three other loci.

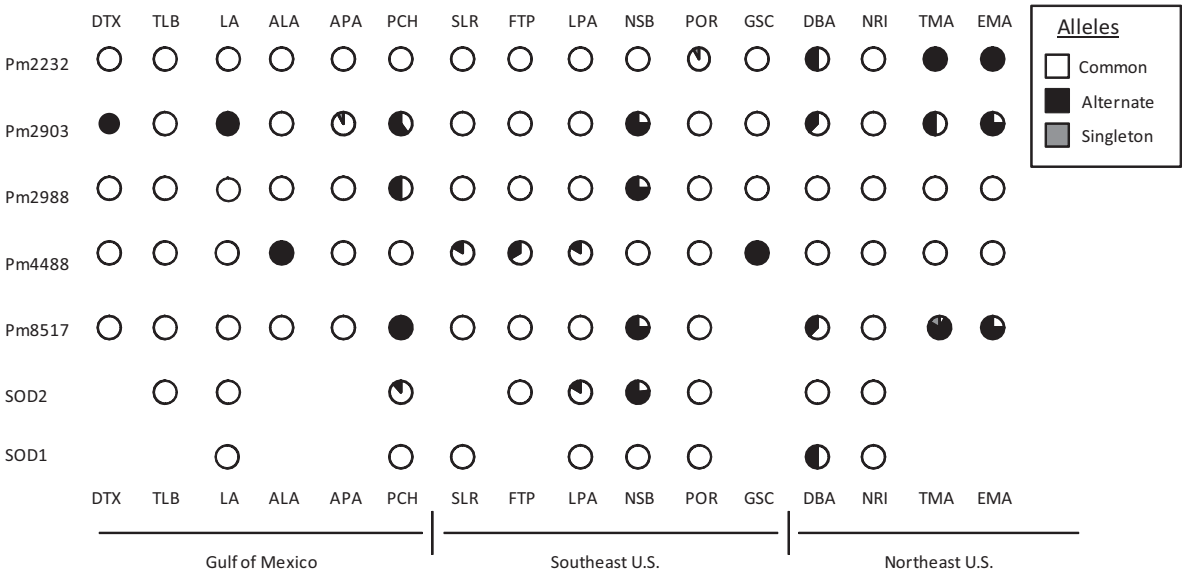

Fig. 3. Geographic distribution of alleles. Sampling locations are provided in a linear array along the coastline from the Gulf of Mexico to Massachusetts. Allele frequencies are represented by pie charts for every locus and location.

mate the length of the longest branch in each of the seven gene trees and MTML-msBayes to conduct approximate Bayesian computation of the most likely number of divergence events given the seven locus data set. The mean length of the longest branch varied from 0.195 (locus Pm2232) to 1.82 (locus SOD2). The 95% highest probability density (HPD) interval for length of longest branch was broadly overlapping across most loci (Fig. 2), as would be expected if all dimorphic loci had allele classes that began diverging at the same time. Approximate Bayesian computation indicated a 65.6% posterior probability that all seven loci began diverging simultaneously. There was 17.2% posterior probability that loci began diverging at two separate times. Each scenario involving more than two divergences had less than 6% support and cumulatively had less than 20% overall probability. Additionally, the 95% HPD of omega (the ratio of the variance in mean divergence times among loci to the mean divergence time) encompasses 0, indicating that there is not sufficient evidence to refute simultaneous divergence among all loci examined here (Hickerson et al., 2006).

3.4. Geographic distribution of alleles and genotypes

Both common and alternate haplotypes were widely distributed for most loci (Fig. 3), suggesting little population substructure. However, the alternate allele for two loci, locus Pm2232 and SOD1, was not observed south of Chesapeake Bay. Likewise, the alternate allele for locus Pm2988 was restricted to Florida and for locus Pm4488 was not observed north of Chesapeake Bay.

Narrower geographic restriction of the alternate allele was seen with locus Pm2988, observed only in Port Charlotte and New Smyrna Beach, FL (Fig. 3).

A total of ten multilocus genotypes were observed. The most frequent genotype (56%) was homozygous for the common allele at all loci and was observed in nine locations from Texas to Rhode Island. Heterozygosity for dimorphic alleles was only observed in individuals from Maryland, New Jersey, and Massachusetts (both Tisbury and Edgartown). South of Chesapeake Bay, heterozygosity occurred only in two Florida individuals carrying singleton variants. The lack of observed heterozygotes is reflected in high inbreeding coefficients for every locus (Table 3, average $F = 0.897$, SE = 0.055).

In order to assess the geographic distribution of genotypic diversity, samples were partitioned according to regional genetic subdivisions reported previously (Reece et al., 2001). Samples from the southeastern region had significantly lower genotypic diversity ($\hat{G} = 0.486$) than in the northeast ($\hat{G} = 0.837$) based on bootstrap sampling with replacement (one-tailed test, $p = 0.005$). Genotypic diversity in the Gulf of Mexico was intermediate ($\hat{G} = 0.614$) among regions, with comparison to the northeast showing marginal significance ($p = 0.054$). Significant genetic differences were observed among regions using AMOVA (PhiCT = 0.229, $p < 0.001$). Pairwise tests between regions indicated that Gulf of Mexico and southeastern Atlantic samples were not significantly different (PhiCT = .026, $p = 0.185$), but each was significantly different from the northeastern region (PhiCT = 0.298 and 0.321, respectively, $p < 0.001$ for both comparisons).

3.5. Recombination

Recombination within and between loci was examined in order to ascertain whether reproductive modes contribute to the maintenance of allelic dimorphism in *P. marinus*. At the intragenic level there was no indication of recombination because intermediate haplotypes were not observed in haplotype trees (Fig. 1), and Hudson's $R = 0.001$ for all loci with all samples pooled. Long-term asexuality would also result in extensive among-locus linkage disequilibrium (LD), so pairwise tests of LD were conducted on microsatellite loci for samples with all five loci genotyped. Comparing pairs of loci, significant LD was found in 6 of 10 tests following sequential Bonferroni correction ($p < 0.01$). Population structure also can cause LD so we repeated tests at the regional scale. Using regional partitions as before, significant LD was found in 2 of 6 tests in the Gulf of Mexico and Southeastern US, and 3 of 3 tests in Northeastern US samples after sequential Bonferroni correction. For regional tests of LD, some pairwise comparisons were not possible as a particular locus or loci were fixed within the region.

4. Discussion

To characterize the genomic extent of allelic dimorphism in *P. marinus*, nucleotide sequences were obtained from seven nuclear loci based on PCR amplicons from infected tissue of wild oysters collected between Massachusetts and Texas. Allelic dimorphism strong enough to reject Wright–Fisher neutrality was found at six out of seven nuclear loci in *P. marinus*, in agreement with similarly strong dimorphic patterns reported for three other loci (Brown et al., 2004; Reece et al., 1997; Robledo et al., 1999). Nucleotide differences between allele classes averaged 3.8% across all loci with no apparent intragenic recombination, indicating that alleles have been evolving independently for an extended period of time, even though both allele classes were found within contemporary samples from specific geographic locations. Treating the allelic clades as independent lineages and focusing on the variance of gene tree lengths among loci failed to reject the hypothesis of a single divergence time for all loci. This allelic divergence time was potentially as long as 18 million years but no less than 900,000 years if mutation rates in *P. marinus* are similar to other single-celled organisms (Kondrashov and Kondrashov, 2010; Machado and Ayala, 2001). These conclusions rest on some assumptions that we critically evaluate below before ultimately discussing mechanisms and implications.

4.1. Assumptions

One assumption critical to our analyses is that *Perkinsus* species are diploid, which has not been determined empirically at this time. Based on population genetic data for one locus, Robledo et al. (1999) hypothesized that *P. marinus* may be diploid during infection and culture. There is increasing evidence from other genetic studies that heterozygous individuals are frequently encountered in clonally cultured isolates (Reece et al., 2001; Thompson, 2010; Vilas et al., 2011) and wild oyster infections (Thompson et al., 2011), arguing for diploidy as the primary form in shellfish hosts. Here we showed that the clonal *P. marinus* isolate ATCC #PRA-240 had both common and alternate alleles at four of the seven loci tested. This cultured isolate was subjected to two rounds of limiting dilution in order to gain assurance that it derived from a single cell prior to whole genome sequencing (Fernández-Robledo et al., 2008).

Another assumption is that recombination rarely occurs within loci. This assumption could have influenced our accurate inference of phase for the mutations in the 13 heterozygous genotypes observed. There is high confidence in our inference (all PHASE p-values < 0.001) because homozygous genotypes (461 total) far outnumber heterozygotes at every locus. Also, the lack of recombination inferred within loci is reinforced by high linkage disequilibrium found among loci.

A third assumption is that null alleles have not biased our conclusions. Nearly one third of all amplification attempts failed, and if this was caused by sequence variation at PCR priming sites then it could lower observed haplotype diversity. However, two results suggest that a likelier general cause for missing data was scarcity of template: (1) Null amplifications were distributed relatively evenly among both loci and locations, indicating that if null alleles were present, they would have to occur at every locus and be present throughout the entire sampling range. (2) There was very little sequence variation uncovered within any of the "common" allelic classes, in spite of robust sampling effort (a minimum of 93 chromosomes sampled at any locus). Therefore, minor-frequency sequence variants at the flanking priming sites are unlikely.

Thus, the best evidence suggests that *P. marinus* is diploid, only rarely recombines, and that null alleles are not common in the data reported here. Those assumptions serve as a basis for considering how this parasite could maintain two highly diverged alleles at every locus without also having accumulated additional variability within any locus.

4.2. Origins of dimorphism

How can we possibly explain the striking lack of genetic variants given the occurrence of exactly two, remarkably distinct allelic classes at each of these loci? Of the 474 DNA sequences collected in this study, 469 could be categorized as one of two types: a high frequency "common" allele or an "alternate" allele at low to moderate frequencies. Given two alleles at each of five loci there are $3^5 = 243$ possible diploid genotypes of which only 10 were observed, consistent with primarily clonal reproduction,

reinforced by frequent inbreeding when sexual reproduction occurs.

Sexual populations typified by low levels of neutral genetic diversity should be characterized by shallow gene genealogies, owing to low mutation rates and/or small genetically effective population size. In *P. marinus,* the divergence between allelic classes demonstrates that there is no lack of mutation, and implicates an extremely small effective population size as the primary cause of low haplotype diversity. This is reinforced by the remarkable lack of diversity within any allele class. If the two allelic classes began diverging about one million years ago, greater diversity should be seen within each allelic class (if effective population sizes have been even moderately large). The lack of diversity presumably results from primarily clonal propagation and at least one recent genetic bottleneck. The data do not implicate a specific mechanism for any genetic bottleneck, but success of a hybrid between distant lineages could have contributed to such a phenomenon by sweeping through the parasite range.

Allelic dimorphism may be a common trait for certain loci in parasitic protists. The dimorphic patterns described here are similar to those observed at antigen-producing loci in *Plasmodium* and *Trypanosomes* (Machado and Ayala, 2001; Putaporntip et al., 2006; Rich et al., 2000). Explanations for allelic dimorphism have been varied and may be specific to each particular taxon or class of locus (Roy et al., 2008; Sibley and Ajioka, 2008). Here, allelic dimorphism was shown to be ubiquitous within the genome of *P. marinus*.

Hypothesized causes of allelic dimorphism include balancing selection (Polley and Conway, 2001), ancient asexuality (Birky, 1996; Mark Welch and Meselson, 2000), genome duplication (Hartl et al., 2002; Miller et al., 1993), or introgression between two divergent lineages (Polley et al., 2005). Neither balancing selection nor ancient asexuality seems to fit the dimorphism reported here. Balancing selection, though a force that can in theory elevate genetic diversity, would be expected to maintain more than two allelic classes. Also, balancing selection is not expected to be manifest across entire genomes, making this an unsatisfactory explanation for the allelic dimorphism observed here. Strict asexuality would be expected to promote genomic heterozygosity, as mutations would accumulate independently on homologous chromosomes. However, in cultured isolates analyzed here, heterozygosity was never seen for more than two out of five microsatellite loci, making this mechanism untenable.

Genome duplication may explain the patterns observed here as a single duplication would explain the simultaneous divergence observed at every locus tested. Under the genome duplication hypothesis, cases of apparent heterozygosity result from amplifying both duplicate genes whereas apparent homozygosity would occur at loci too diverged to be simultaneously amplified, or when one copy had been subsequently deleted. However, we observed too little sequence divergence between allelic classes to expect frequent PCR priming problems. Thus, seemingly nonrandom patterns of secondary deletion would need to be invoked to explain some clonal isolates with multiple 'heterozygous' loci while other isolates remain uniformly 'homozygous'. Furthermore, the geographic pattern of occurrence for major and minor alleles, were it to be explained by copy number variation (CNV), would require that nearly all populations south of Chesapeake Bay have one but not both copies, a scenario that seems implausible (Fig. 3). Nonetheless, given the millions of years potentially involved and a possible mosaic history of local clonal expansions generating geographic variation in CNV patterns, this hypothesis is viable.

Finally, introgression between two divergent parasite lineages seems plausible based on two contemporary observations: (1) genotypic patterns north of Chesapeake Bay suggest that sexual recombination between divergent strains caused elevated heterozygosity, in both the DNA sequence data reported here and in microsatellite loci (Thompson et al., 2011) and (2) the rapid rate of range expansion north of Chesapeake Bay, over 600 km of coastline in less than 10 years, suggests that the whole continent could have experienced a similarly rapid expansion in the past, broadly distributing introgressed lineages from the original point of hybridization. Furthermore, hybridization between two long-diverged lineages could explain a consistent divergence time for dimorphic alleles at loci across the genome. Assuming that genetic drift is controlling the fate of alleles at most loci analyzed here, the fact that both dimorphic alleles were observed at every locus strongly suggests that the hybridization event must be relatively recent. Otherwise, given the low effective population size implied by negligible intra-class nucleotide diversities, one or the other allele should have disappeared from the species through random genetic drift. Taken altogether, recent hybridization seems the simplest explanation for the dimorphic patterns reported here, as genome duplication requires complex explanations of geographically variable loss and gain of gene copies within individuals, and other evolutionary processes (such as asexuality or balancing selection) poorly fit these data.

The hybridization hypothesis holds that independently evolving parasite populations were physically separated yet retained their ability to mate, and did so, upon secondary contact. There have been two recent reports of *P. marinus* infection in different oyster hosts (*Crassostrea corteziensis* and *Crassostrea gigas*) on the Pacific coast of Mexico (Cáceres-Martínez et al., 2008; Enriquez-Espinoza et al., 2010) and in *Crassostrea rhizophorae* along the coast of South America (da Silva et al., 2013). Two forms of the internal transcribed spacer locus were observed in *P. marinus* isolates from the Pacific, suggesting dimorphism may also be present in these populations (Escobedo-Fregoso et al., 2013). Nevertheless, these newly identified infections could represent a potential source of the highly diverged alleles found in this study.

If hybridization between two ancient lineages caused the allelic patterns in *P. marinus*, ecological and epidemiological characteristics may have been altered. Current populations of *Spartina* hybrids are expanding their geographic range (Ayres et al., 2004). Hybridization apparently played a role in the establishment of anoles in Florida (Kolbe et al., 2004). Hybrid fungal diseases of plants have shown increased virulence (Barrett et al., 2007; Brasier et al., 1999), and changes in life-history, including timing of infection in rust infections (Barrett et al., 2007). Consequently, the putative hybridization of *P. marinus* lineages may have similarly changed the ecological and epidemiological properties of this parasite.

Acknowledgments

We would like to thank Scott Roy for assistance in designing a formal test of neutrality based on the coalescent. We are also thankful to W. Pecher and G. Vasta for sharing samples collected in the northeastern portion of the range as well as multiple state agencies for collecting wild samples from the Gulf of Mexico and South Carolina. We would also like to thank the two anonymous reviewers whose comments greatly improved this manuscript. This publication was supported by the Behavior, Ecology, Evolution, and Systematics Program at the University of Maryland College Park and salary support to PCT from the U.S. Department of Agriculture.

References

Albright, B.W., Abbe, G.R., McCollough, C.B., Barker, L.S., Dungan, C.F., 2007. Growth and mortality of dermo-disease-free juvenile oysters (*Crassostrea virginica*) at

three salinity regimes in an enzootic area of Chesapeake Bay. J. Shellfish Res. 26, 451–463.
Ayres, D., Smith, D., Zaremba, K., Klohr, S., Strong, D., 2004. Spread of exotic cordgrasses and hybrids (*Spartina* sp.) in the tidal marshes of San Francisco Bay, California, USA. Biol. Invasions 6, 221–231.
Barrett, L.G., Thrall, P.H., Burdon, J.J., 2007. Evolutionary diversification through hybridization in a wild host-pathogen interaction. Evolution 61, 1613–1621.
Birky Jr., C.W., 1996. Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. Genetics 144, 427–437.
Brasier, C.M., Cooke, D.E.L., Duncan, J.M., 1999. Origin of a new Phytophthora pathogen through interspecific hybridization. Proc. Natl. Acad. Sci. USA 96, 5878–5883.
Brown, G.D., Hudson, K.L., Reece, K.S., 2004. Multiple polymorphic sites at the ITS and ATAN loci in cultured isolates of *Perkinsus marinus*. J. Eukaryot. Microbiol. 51, 312–320.
Burreson, E.M., Ragone Calvo, L.M., 1996. Epizootiology of *Perkinsus marinus* disease of oysters in Chesapeake Bay, with emphasis on data since 1985. J. Shellfish Res. 15, 17–34.
Bushek, D., Allen, S., 1996. Host-parasite interactions among broadly distributed populations of the eastern oyster *Crassostrea virginica* and the protozoan *Perkinsus marinus*. Mar. Ecol. Prog. Ser. 139, 127–141.
Cáceres-Martínez, J., Vásquez-Yeomans, R., Padilla-Lardizábal, G., del Río Portilla, M.A., 2008. *Perkinsus marinus* in pleasure oyster *Crassostrea corteziensis* from Nayarit, Pacific coast of México. J. Invertebr. Pathol. 99, 66–73.
Craig, A., Powell, E.N., Fay, R.R., Brooks, J.M., 1989. Distribution of *Perkinsus marinus* in gulf coast oyster populations. Estuaries 12, 82–91.
Da Silva, P.M., Vianna, R.T., Guertler, C., Ferreira, L.P., Santana, L.N., Fernández-Boo, S., Ramilo, A., Cao, A., Villalba, A., 2013. First report of the protozoan parasite *Perkinsus marinus* in South America, infecting mangrove oysters *Crassostrea rhizophorae* from the Paraíba River (NE, Brazil). J. Invertebr. Pathol. 113, 96–103.
Dungan, C.F., Reece, K.S., 2006. In vitro propagation of two *Perkinsus spp*. parasites from Japanese Manila clams *Venerupis philippinarum* and description of *Perkinsus honshuensis n. sp*. J. Eukaryot. Microbiol. 53, 316–326.
Enriquez-Espinoza, T., Grijalva-Chon, J., Castro-Longoria, R., Ramos-Paredes, J., 2010. *Perkinsus marinus* in *Crassostrea gigas* in the Gulf of California. Dis. Aquat. Organ. 89, 269–273.
Escobedo-Fregoso, C., Arzul, I., Carrasco, N., Gutiérrez-Rivera, J.N., Llera-Herrera, R., Vázquez-Juárez, R., 2013. Polymorphism at the ITS and NTS loci of *Perkinsus marinus* isolated from cultivated oyster *Crassostrea corteziensis* in Nayarit, Mexico and phylogentic relationship to *P. marinus* along the Atlantic coast. Transbound Emerg. Dis.. http://dx.doi.org/10.1111/tbed.12090.
Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinform. Online 1, 47–50.
Fernández-Robledo, J.A., Lin, Z., Vasta, G.R., 2008. Transfection of the protozoan parasite *Perkinsus marinus*. Mol. Biochem. Parasitol. 157, 44–53.
Ford, S.E., 1996. Range extension by the oyster parasite *Perkinsus marinus* into the northeastern United States: response to climate change? J. Shellfish Res. 15, 45–56.
Frank, S.A., 1996. Models of parasite virulence. Q. Rev. Biol. 71, 37–78.
Grabowski, J.H., Brumbaugh, R.D., Conrad, R.F., Keeler, A.G., Opaluch, J.J., Peterson, C.H., Piehler, M.F., Powers, S.P., Smyth, A.R., 2012. Economic valuation of ecosystem services provided by oyster reefs. Bioscience 62, 900–909.
Hartl, D.L., Volkman, S.K., Nielsen, K.M., Barry, A.E., Day, K.P., Wirth, D.F., Winzeler, E.A., 2002. The paradoxical population genetics of *Plasmodium falciparum*. Trends Parasitol. 18, 266–272.
Hickerson, M.J., Stahl, E.A., Lessios, H.A., 2006. Test for simultaneous divergence using approximate Bayesian computation. Evolution 60, 2435–2453.
Huang, W., Takebayashi, N., Qi, Y., Hickerson, M.J., 2011. MTML-msBayes: approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. BMC Bioinformatics 12, 1. http://dx.doi.org/10.1186/1471-2105-12-1.
Kolbe, J.J., Glor, R.E., Rodriguez Schettino, L., Lara, A.C., Larson, A., Losos, J.B., 2004. Genetic variation increases during biological invasion by a Cuban lizard. Nature 431, 177–181.
Kondrashov, F.A., Kondrashov, A.S., 2010. Measurements of spontaneous rates of mutations in the recent past and the near future. Philos. Trans. R. Soc. Lond. B, Biol. Sci. 365, 1169–1176.
Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.
Machado, C.A., Ayala, F.J., 2001. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. Proc. Natl. Acad. Sci. USA. 98, 7396–7401.
Mackin, J., 1956. *Dermocystidium marinum* and salinity. Proc. Natl. Shellfish Assoc. 46, 116–128.
Mark Welch, M., Meselson, M., 2000. Evidence for evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science 288, 1211–1215.
McCollough, C.B., Albright, B.W., Abbe, G.R., Barker, L.S., Dungan, C.F., 2007. Acquisition and progression of *Perkinsus marinus* infections by specific-pathogen-free juvenile oysters (*Crassostrea virginica* Gmelin) in a mesohaline Chesapeake Bay tributary. J. Shellfish Res. 26, 465–477.
Meirmans, P.G., vanTienderen, P.H., 2004. GenoType and GenoDive: two programs for the analysis of genetic diversity of asexual organisms. Mol. Ecol. Notes 4, 792–794.
Miller, L.H., Roberts, T., Shahabuddin, M., McCutchan, T.F., 1993. Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1). Mol. Biochem. Parasitol. 59, 1–14.
Moss, J.A., Xiao, J., Dungan, C.F., Reece, K.S., 2008. Description of *Perkinsus beihaiensis* n. sp., a new *Perkinsus* sp. parasite in oysters of southern China. J. Eukaryot. Microbiol. 55, 117–130.
Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press.
Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.
Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press.
Polley, S.D., Conway, D.J., 2001. Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. Genetics 158, 1505–1512.
Polley, S.D., Weedall, G.D., Thomas, A.W., Golightly, L.M., Conway, D.J., 2005. Orthologous gene sequences of merozoite surface protein 1 (MSP1) from *Plasmodium reichenowi* and *P. gallinaceum* confirm an ancient divergence of *P. falciparum* alleles. Mol. Biochem. Parasitol. 142, 25–31.
Putaporntip, C., Jongwutiwes, S., Iwasaki, T., Kanbara, H., Hughes, A.L., 2006. Ancient common ancestry of the merozoite surface protein 1 of *Plasmodium vivax* as inferred from its homologue in *Plasmodium knowlesi*. Mol. Biochem. Parasitol. 146, 105–108.
Ramos-Onsins, S.E., Mitchell-Olds, T., 2007. Mlcoalsim: multilocus coalescent simulations. Evol. Bioinform. Online 3, 41.
Raymond, M., Rousset, F., 1995. Genepop (version 3.4): population genetics software for exact tests and ecumenicism. J. Hered. 86, 248–249.
Reece, K.S., Siddall, M.E., Burreson, E.M., Graves, J.E., 1997. Phylogenetic analysis of *Perkinsus* based on actin gene sequences. J. Parasitol. 83, 417–423.
Reece, K., Bushek, D., Hudson, K., Graves, J., 2001. Geographic distribution of *Perkinsus marinus* genetic strains along the Atlantic and Gulf coasts of the USA. Mar. Biol. 139, 1047–1055.
Reed, D.H., Frankham, R., 2001. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution 55, 1095–1103.
Rich, S.M., Ferreira, M.U., Ayala, F.J., 2000. The origin of antigenic diversity in *Plasmodium falciparum*. Parasitol. Today 16, 390–396.
Robledo, J.A.F., Wright, A.C., Marsh, A.G., Vasta, G.R., 1999. Nucleotide sequence variability in the nontranscribed spacer of the rRNA locus in the oyster parasite *Perkinsus marinus*. J. Parasitol. 85, 650–656.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.
Rosenberg, N.A., Nordborg, M., 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat. Rev. Genet. 3, 380–390.
Roy, S.W., Ferreira, M.U., Hartl, D.L., 2008. Evolution of allelic dimorphism in malarial surface antigens. Heredity 100, 103–110.
Saldarriaga, J.F., McEwan, M.L., Fast, N.M., Taylor, F.J.R., Keeling, P.J., 2003. Multiple protein phylogenies show that *Oxyrrhis marina* and *Perkinsus marinus* are early branches of the dinoflagellate lineage. Soc. General Microbiol. 53, 355–365.
Schott, E.J., Robledo, J.-A.F., Wright, A.C., Silva, A.M., Vasta, G.R., 2003. Gene organization and homology modeling of two iron superoxide dismutases of the early branching protist *Perkinsus marinus*. Gene 309, 1–9.
Selkoe, K.A., Toonen, R.J., 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol. Lett. 9, 615–629.
Sibley, L.D., Ajioka, J.W., 2008. Population structure of *Toxoplasma gondii*: clonal expansion driven by infrequent recombination and selective sweeps. Annu. Rev. Microbiol. 62, 329–351. http://dx.doi.org/10.1146/annurev.micro.62.081307.162925.
Siddall, M.E., Reece, K.S., Graves, J.E., Burreson, E.M., 1997. "Total evidence" refutes the inclusion of *Perkinsus* species in the phylum *Apicomplexa*. Parasitology 115, 165–176.
Soniat, T., 1985. Changes in levels of infection of oysters by *Perkinsus marinus*, with special reference to the interaction of temperature and salinity upon parasitism. NE Gulf Sci. 7, 171–174.
Strobeck, C., 1987. Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. Genetics 117, 149–153.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
Thompson, P.C., 2010. Sex, Dispersal, and Deep Divergence: The Population Genetics of the Protistan Parasite *Perkinsus marinus* (Ph.D Dissertation). University of Maryland, College Park, MD.
Thompson, P.C., Rosenthal, B.M., Hare, M.P., 2011. An evolutionary legacy of sex and clonal reproduction in the protistan oyster parasite *Perkinsus marinus*. Infect. Genet. Evol. 11, 598–609.
U.S. Department of Commerce, National Oceanic and Atmospheric Administration, 2012. Fisheries of the United States, 2012. Available from http://www.st.nmfs.noaa.gov/pls/webpls/MF_ANNUAL_LANDINGS.RESULTS on 11/13/2013.
Vilas, R., Cao, A., Pardo, B.G., Fernández, S., Villalba, A., Martínez, P., 2011. Very low microsatellite polymorphism and large heterozygote deficits suggest founder effects and cryptic structure in the parasite *Perkinsus olseni*. Infect. Genet. Evol. 11, 904–911.