

Positive selection on mitochondrial DNA in the eastern oyster, *Crassostrea virginica*

Colin G. Rose¹ and Matthew P. Hare²

University of Maryland, Department of Biology, College Park, Maryland
 1 – current address: Department of Biology, University of Southern California, Los Angeles, CA
 2 – current address: Department of Natural Resources, Cornell University, Ithaca, NY

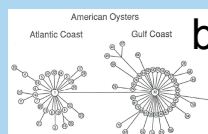
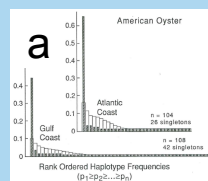


ABSTRACT

Recent meta-analyses of metazoan mitochondrial (mt)DNA data suggest that invertebrate species have effective population sizes (N_e) sufficiently large that natural selection, and not mutation-drift equilibrium, determines mtDNA diversity (Bazin et al. 2006). This finding has important implications for the use of mtDNA to make conservation-related demographic inferences but strong multilocus data sets allowing tests within species are few. Large invertebrate N_e contradicts a well known pattern in many marine invertebrate species with high fecundity and high early mortality, such as oysters, where N_e is estimated to be many orders of magnitude below census size (N) and in some cases as low as $N_e = 10^2 - 10^3$. To resolve this contradiction at the level of a single species, we compared mitochondrial and nuclear loci in the eastern oyster, *Crassostrea virginica*. Based on a battery of tests for selection using DNA sequence data, neutrality was rejected for mitochondrial DNA but mostly neutral patterns and high polymorphism were found at seven unlinked nuclear loci, implicating natural selection as the force responsible for low, non-neutral mtDNA variation in *C. virginica*. We conclude that either background selection or periodic selective sweeps shape mitochondrial variation in eastern oysters despite sweepstakes reproduction or other processes creating a small N_e/N ratio.

BACKGROUND

A prominent hypothesis explaining low N_e/N ratios is “sweepstakes” reproduction generating huge variance in family size, and some of the strongest



empirical data supporting sweepstakes reproduction are from non-neutral patterns of mtDNA variation in oysters.

Figure 1: American (= eastern) oyster mtDNA haplotype frequency spectra (a) and networks (b). Open histogram bars show neutral expectations. From Beckenbach (1994).

Beckenbach (1994) observed that skewed mtDNA haplotype frequency distributions and “hub and spokes” haplotype networks were characteristic of multiple independent oyster populations. He proposed that very small N_e coupled with very high mutation rates could produce these patterns.

Eldon & Wakeley (2006) applied a multiple-mergers model, an alternative to coalescent theory when family size is strongly skewed and the assumption of binary gene lineage coalescence is likely violated. Using Pacific oyster mtDNA data similar to that shown above, they inferred that a single individual replaces as much as 8% of the population with its offspring.

Patterns of mitochondrial DNA variation are repeatedly being used to infer oyster demography, providing support for extreme variance in family size (sweepstakes reproduction) and very small N_e . This demographic mechanism should produce similar patterns of variation at all neutral loci, both mitochondrial and nuclear.

RESULTS

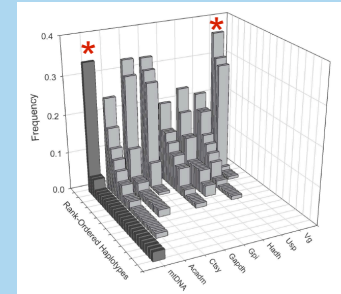


Fig. 2: Haplotype frequency distributions for combined mtDNA data and seven nDNA loci. Asterisks mark significant Tajima's D .

Table 1: Nucleotide polymorphism and divergence with tests of selection using *C. rhizophorae* as outgroup

	Locus	bp	m	S	P_s	D_{xy}	FL-D	FL-F
mtDNA	Atp6	127	44	2	0.016	0.203	-1.83	-1.89
	Cox1	658	18	3	0.009	0.192	-2.51*	-2.69*
	Cox2	216	47	5	0.029	0.131	-0.11	-0.69
	Cytb	348	46	7	0.02	0.235	-0.1	-0.76
Nuclear DNA	Acadm	498	27	23	0.047	0.025	-2.02	-2.35
	Ctsy	555	22	21	0.038	0.041	-0.61	-0.71
	Gapdh	624	26	6	0.011	0.033	-2.04	-2.09
	Gpi	297	24	18	0.109	0.035	-0.94	-0.93
	Hadh	585	15	12	0.021	0.05	-0.14	-0.01
	Usp	438	28	12	0.027	0.057	-0.64	-0.98
	Vg	547	26	12	0.03	0.035	-2.46*	-2.80*

- A single Chesapeake Bay population of oysters was sampled and compared with the closest congener, *C. rhizophorae*
- Proportion of polymorphic silent sites, P_s , was lower in mtDNA than nDNA loci despite much smaller numbers of chromosomes sampled (m , Table 1) for nDNA (mtDNA $P_s = 0.045$, nDNA $P_s = 0.085$, $p < 0.05$).
- In contrast, mean divergence between species was significantly higher for mtDNA than nDNA loci at silent sites (mtDNA $D_{xy} = 0.574$, nDNA $D_{xy} = 0.089$, $p < 0.01$), but not at replacement sites (mtDNA $D_{xy} = 0.033$, nDNA $D_{xy} = 0.016$, $p > 0.05$).
- Only the concatenated mtDNA data had a significant McDonald-Kreitman test ($p = 0.027$, neutrality index = 0.173). dN/dS was uniformly low, suggesting purifying selection.

- Pairwise Hudson, Kreitman, Aguade (HKA) tests rejected neutrality at the $p < 0.01$ level for only five locus contrasts out of 55, all of them between mt-Cox1 and nDNA loci.
- In contrast to nDNA loci (Fig. 2), mtDNA had an extreme excess of both high and low frequency haplotypes (the latter mostly singletons) and a deficit at intermediate frequency.
- Combined mtDNA data (excluding Cox1) rejected neutrality based on Tajima's D (-2.58, $p < 0.00001$) and Fu's F_s (-27.1, $p < 0.00001$). Only one nDNA locus was significant for each of these tests.
- Compound DH and DHEW tests (Table 2) are the most powerful frequency-based tests for positive selection and are relatively insensitive to background selection and demography. These tests show strong support for selection on mtDNA and moderate support for 2 nDNA loci.

Table 2: Tests of Selection based on frequency spectrum

Tajima's D	Fu's F_s	EW	FW-H	DH (P)	DHEW (P)
-1.14	-1.67*	0.956	0.115	0.166	0.159
-1.714	-2.60**	0.704*	0.378	0.227	0.190
-1.82*	-5.34***	0.766*	-1.93*	0.006	0.002
-2.18*	-5.23***	0.836**	-6.40*	< 0.001	< 0.001
-1.52	-13.29	0.125*	0.699	0.595	0.491
-0.52	-0.32	0.149*	-0.555	0.172	0.090
-1.18	-3.20*	0.500*	-1.40	0.033	0.016
-0.54	-4.8	0.101	-0.400	0.189	0.188
0.29	-2.98	0.120	0.876	0.792	0.774
-0.65	-6.59	0.212	-1.54	0.018	0.011
-2.08*	-2.35	0.497*	-0.369	0.086	0.045

- Effective Population Size: Coalescent estimate of N_e based on all nDNA loci was $N_e = 1.8 \times 10^5$

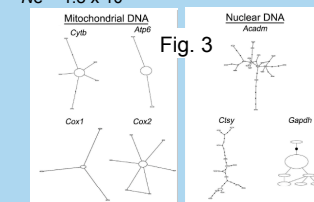


Fig. 3: As expected, the mtDNA haplotype skew is associated with simple “hub and spokes” minimum spanning networks that contrast in pattern with nDNA loci.

CONCLUSIONS

- Because our sample of nuclear genes does not show deviations from neutrality as extreme or consistently as mtDNA, we conclude that selection is shaping patterns of variation at mtDNA and lowering within population variation below mutation-drift expectations.
- There is support for both purifying selection and positive selection on the mtDNA. The strongest evidence for positive selection comes from the McDonald-Kreitman test because it is insensitive to demographic history.
- The N_e estimated with nDNA loci is several orders of magnitude lower than census N but does not support an enigmatically extreme N_e/N ratio as did mtDNA. One conservation implication of large N_e is that hatchery-based supplementation (e.g., Chesapeake Bay oyster restoration) has a greater risk of increasing inbreeding within wild supplemented populations.

ACKNOWLEDGMENTS

Thanks to K. Paynter and H. Hertler for samples, J. Wakeley for multiple merger software and M. Kuhner for advice. Thanks to Cornell Center for Comparative and Population Genomics for travel funds.