Evidence for Positive Selection at the Pantophysin (Pan I) Locus in Walleye Pollock, *Theragra chalcogramma*

M. F. Canino[1] *and P. Bentzen*[2]

School of Aquatic and Fishery Sciences, University of Washington, Seattle

Nucleotide polymorphism at the pantophysin (*Pan* I) locus in walleye pollock, *Theragra chalcogramma*, was examined using DNA sequence data. Two distinct allelic lineages were detected in pollock, resulting from three amino acid replacement mutations in the first intravesicular domain of the protein. The common *Pan* I allelic group, comprising 94% of the samples, was less polymorphic ($\pi = 0.005$) than the uncommon group ($\pi = 0.008$), and nucleotide diversity in both was higher than for two allelic lineages in the related Atlantic cod, *Gadus morhua*. Phylogenetic analyses of *Pan* I sequences from these two species did not clearly resolve orthology among allelic groups, in part because of recombination that has occurred between the two pollock lineages. Conventional tests of neutrality comparing polymorphisms within and between homologous regions of the *Pan* I locus in walleye pollock and Atlantic cod did not detect the effects of selection. This result is likely attributed to low levels of synonymous divergence among allelic lineages and a lack of mutation-drift equilibrium inferred from nucleotide mismatch frequency distributions. However, the ratio of nonsynonymous to synonymous substitutions per site (d_N/d_S) exceeded unity in two intravesicular domains of the protein and the influence of positive selection at multiple codon sites was strongly inferred through the use of maximum-likelihood analyses. In addition, the frequency spectrum of linked neutral variation showed indirect effects of adaptive hitchhiking in pollock resulting from a selective sweep of the common allelic lineage. Recombination between the two allelic classes may have prevented complete loss of the older, more polymorphic lineage. The results suggest that recurrent sweeps driven by positive selection is the principle mode of evolution at the *Pan* I locus in gadid fishes.

Introduction

The role of natural selection in shaping patterns of molecular evolution has been a topic of considerable interest over the past 3 decades. The proportion and types of genetic changes favored by selection generate patterns of polymorphism within and among species that can be distinguished from neutral modes of evolution. Under an infinite-sites neutral model, the level of DNA polymorphism within a species is proportional to the amount of divergence at the locus among closely related species (Nei 1987) and deviations from this pattern form the basis for various tests for natural selection, such as the HKA test (Hudson, Kreitman, and Aguadé 1987), Tajima's *D* statistic (Tajima 1989), and the M-K test (McDonald and Kreitman 1991). Comparisons of rate differences for synonymous (d_S) and nonsynonymous (d_N) substitutions per nucleotide site in DNA sequences provide a direct method to measure selective pressure on the protein. If nonsynonymous substitutions are deleterious, d_N will be less than the neutral rate of substitution (d_S). However, d_N can exceed d_S when natural selection favors nonsynonymous change at the amino acid level. Moreover, the criterion for d_N/d_S to be greater than 1 as proof of positive selection is overly stringent when averaged over all possible replacement mutations, as many codons may be invariant because of strong functional constraints at the protein level. Evidence for positive selection at many loci has been accumulating rapidly (Endo, Ikeo, and Gojobori 1996; Akashi 1999; see reviews by Yang and Bielawski [2000] and [Ford 2002]) because of to the increased availability of DNA sequence data and the development of more realistic statistical approaches for detecting selection (Goldman and Yang 1994; Nielsen and Yang 1998; Yang et al. 2000; Yang and Swanson 2002; Su, Nguyen, and Nei 2002).

Pogson (2001) presented evidence for the effects of selection at the nucleotide level for the pantophysin (*Pan* I) locus in Atlantic cod (*Gadus morhua*). This integral membrane protein has been localized to small cytoplasmic vesicles, but its exact functions in microvesicle trafficking and exocytotic pathways are poorly understood (Windoffer et al. 1999; Brooks et al. 2000). Two major allelic lineages in cod, *Pan* I^A and *Pan* I^B, were characterized by six fixed radical amino acid replacement substitutions within a small intravesicular (IV1) domain of the protein, suggesting a long-lived polymorphism at the locus. Both lineages exhibited strong linkage disequilibrium and low nucleotide diversity, with nearly all of the nucleotide variation occurring between, rather than within, *Pan* I^A and I^B groups. The low levels of linked variation within the two lineages, combined with their geographic distributions, suggested that strong directional selection had caused selective sweeps (Charlesworth, Morgan, and Charlesworth 1993), where recently derived and favored alleles within both classes were spreading at the expense of older alleles. Population studies of Atlantic cod (Pogson, Mesa, and Boutilier 1995; Fevolden and Pogson 1997; Pogson et al. 2001; Jónsdóttir et al. 1999; Jónsdóttir, Daníelsdóttir, and Nævdal 2001; Pogson and Fevolden 2003) have shown strong levels of genetic differentiation at the *Pan* I locus that does not appear to conform to an isolation-by-distance pattern, although Karlsson and Morke (2003) recently suggested that the geographic distribution of *Pan* I alleles in cod may be related to water temperature.

In this study, we examined the evidence for selection at the *Pan* I locus in walleye pollock, *Theragra chalcogramma*, an abundant arctoboreal gadid species inhabiting

[1] Present address: Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, Wash.

[2] Present address: Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada.

Key words: positive selection, pantophysin, walleye pollock, maximum likelihood.

E-mail: mike.canino@noaa.gov.

Mol. Biol. Evol. 21(7):1391–1400. 2004
doi:10.1093/molbev/msh141
Advance Access publication April 14, 2004

Table 1
Sample Collection Information

Location	Year	Latitude (°N)	Longitude	Number of Sequences
Unimak Pass, Alaska	1997	54.5	165.6° W	21
Shelikof Strait, Alaska	1997	57.9	154.7° W	20
Prince William Sound, Alaska	1998	60.1	148.3° W	17
Funka Bay, Japan	1998	45.3	143.2° E	19
Puget Sound, Washington	1998	48.1	122.7° W	20
Kronotsky Bay, Russia	1999	52.0	161.0° E	20

coastal shelves and slopes of the north Pacific Ocean and Bering Sea. Variation among *Pan* I sequences in walleye pollock, as in Atlantic cod, showed a characteristic footprint of selection at the nucleotide level. Comparisons of replacement and silent substitution rate ratios using maximum-likelihood models of codon evolution (Goldman and Yang 1994; Nielsen and Yang 1998; Yang et al. 2000) indicated that positive selection in two intravesicular domains has played a dominant role in the evolution of pantophysin in walleye pollock.

Materials and Methods

Samples

Fin clip tissues from walleye pollock were collected at seven locations across the species range from 1997 to 1999 (table 1) and preserved in 95% ETOH until extraction. All samples were from adult fish, except for 1998 Puget Sound samples of age-1 pollock collected by beach seine.

DNA Extraction and Sequencing

Genomic DNA was isolated from preserved tissues after lysis in proteinase K (10 mg ml^{-1} at 60°C for 1h) using Qiagen DNeasy extraction protocols (Valencia, Calif.) and resuspended in 50 to 100 μl of low TE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0). Polymerase chain reaction (PCR) amplifications were performed in 10 μl volumes containing 10 mM Tris-HCL (pH 8.3), 20 mM KCL, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, approximately 100 ng template DNA, 0.8 units of Gene Choice *Taq* DNA polymerase (Kemp Biotechnologies, Inc., Frederick, Md), and 0.5 μM oligonucleotide primers (F, 5′-TCTACAAATGCGT-GAAAGTGG-3′; R, 5′-CCAGACGCTACAGGGAT-CAT-3′) designed to amplify a 985-bp region of the *Pan* I locus. The thermal profile for amplification consisted of a denaturation step of 94°C for 2 min followed by 30 cycles of 94°C (30s)+57°C (30s)+72°C (1min) in a MJ PTC-200 DNA Engine thermocycler (MJ Research, Inc., Waltham, Mass.). PCR amplicons from 117 walleye pollock were cloned into chemically competent *E. coli* using the TOPO-TA kit according to the manufacturer's protocol (Invitrogen Corp., Carlsbad, Calif.).

One random clone from each PCR (i.e., one allele per fish) was chosen for sequencing. Clones were incubated overnight in 3 ml of LB broth (1% tryptone, 0.5% yeast extract, 1.0% NaCl, pH = 7.0) containing 50 ng ml^{-1} ampicillin at 37°C, with vigorous shaking at 220 rpm. Plasmid DNA was extracted and purified using Qiagen (Valencia, Calif.) miniprep spin columns. Approximately 200 ng of plasmid DNA was incorporated in cycle sequencing reactions using DYEnamic ET dye terminators and both strands were sequenced using M13 forward and reverse primers on a MegaBACE 1000 automated sequencer according to the manufacturer's protocols (Amersham Pharmacia Biotech, Inc., Piscataway, NJ). To evaluate the potential effects of *Taq* replication errors during PCR, alleles from five individuals showing singleton nucleotide mutations were amplified, cloned, and sequenced a second time. Preliminary editing and alignment of DNA sequences was accomplished using Sequencher® software before final alignment with the ClustalW algorithm (Thompson, Higgins, and Gibson 1994) using default alignment parameters implemented in BioEdit (Hall 1999). Sequences were deposited in GenBank under accession numbers AY291126 to AY291203.

Data Analyses

Several standard measures of genetic variation were calculated using the DnaSP analysis program (Rozas and Rozas 1999): S is the number of segregating sites, π is nucleotide diversity (Nei 1987), and k is the average number of nucleotide differences (Tajima 1983). Polymorphism at the *Pan* I locus was visualized using a sliding window of nucleotide diversity (Kreitman and Hudson 1991). Linkage disequilibrium among polymorphic sites was computed and tested for significance using Fischer's exact test implemented in the DnaSP program after Bonferroni adjustment for multiple tests (Rice 1989).

The null hypothesis of neutral evolution of pollock *Pan* I alleles was tested using Tajima's *D* statistic (Tajima 1989), derived from the total number of segregating sites and the average number of pairwise differences between sequences. Similar tests based upon the number of singleton mutations versus the total number of mutations (D* [Fu and Li 1993]) or based upon the average number of pairwise differences between sequences (F*) were conducted with the DnaSP program. Interspecific tests of neutral evolution between pollock and Atlantic cod (M-K test [McDonald and Kreitman 1991], HKA test [Hudson, Kreitman and Aguadé 1987]) were conducted using aligned regions of cod *Pan* I^A and I^B alleles (Pogson 2001) retrieved from GenBank (accession numbers AF288943 to AF288977).

Rates of nonsynonymous substitutions per nonsynonymous site (d_N) and synonymous substitution per synonymous site (d_S) were estimated for different coding regions of the *Pan* I locus in both pollock and Atlantic cod sequences according to Nei and Gojobori (1986). Phylogenetic relationships among *Pan* I alleles were examined using aligned sequences from Atlantic cod (*G. morhua*), Greenland cod (*G. ogac*), Arctic cod (*Boreogadus saida*), Polar cod (*Arctogadus glacialis*), and walleye pollock, rooting the tree with *A. glacialis*. Six pollock sequences that showed evidence of recombination (four gamete test [Hudson and Kaplan 1985]) were omitted from the analyses. A phylogeny was estimated for 109 *Pan* I alleles using the neighbor-joining method (Saitou and Nei 1987) with Kimura's two-parameter distance (Kimura 1980) implemented in the PHYLIP program (Felsenstein 1993) and tested using 1,000 bootstrap replicates of the data.

The effects of selection on *Pan* I alleles were investigated through several maximum-likelihood models for estimating d_N/d_S ratios (ω) within allele phylogenies (Goldman and Yang 1994; Nielsen and Yang 1998; Yang et al. 2000). First, trees based upon coding DNA (164 amino acids) for pollock, *G. morhua*, and *G. ogac* alleles were constructed using neighbor-joining, parsimony and maximum-likelihood methods in the PHYLIP program (Felsenstein 1993). The resultant tree topologies were used to construct two subsets of data within and among *Pan* I allelic lineages for analyses with likelihood methods. The first group consisted of 10 pollock sequences used for an intraspecific comparison. The second group of 15 sequences included the same 10 individuals but added two Atlantic cod *Pan* I^A and I^B alleles and one *G. ogac* sequence as an outgroup.

Maximum-likelihood tree topologies for both data sets were then estimated using the DNAML program in PHYLIP and input for analyses using various maximum-likelihood models implemented in the PAML program (Yang 1997). All models were run using the F3×4 option in PAML where expected codon frequencies were based upon nucleotide frequencies occurring at the three codon positions. The one-ratio model (M0) provided an average d_N/d_S ratio (ω) for all lineages (branches) in the tree, whereas the free-ratio model estimated ω separately for each branch. Additional models were tested that allowed for heterogeneity in ω ratios among codons. A neutral model (M1) contained two types of sites, one subject to strong selection against replacement mutations ($\omega = 0$) and an alternate site where substitutions were considered to be selectively neutral ($\omega = 1$). The positive selection model (M2) extended the neutral model to include a third class of sites where ω exceeds 1 and an empirical Bayesian approach was used to assign codons to ω categories and calculate posterior probabilities of the assignments. More complex models of detecting positive selection that allow for different heterogeneous distributions of ω ratios among sites (Yang et al. 2000) were also tested. The beta model (M7) assumes that ω follows a beta distribution over the interval (0,1) and does not allow for sites where ω is greater than 1. An extra class of sites, where ω is estimated from the data and thus can exceed 1, is incorporated in the beta&ω model (M8). Empirical tests of these models (Yang et al. 2000) indicate that comparisons between M7 and M8 can provide a robust test for selection.

Fits of the likelihood models to the data were evaluated using a likelihood ratio test (LRT), in which twice the difference in log-likelihood values between models ($2\Delta l$) is expected to follow a χ^2 distribution, with the number of degrees of freedom equal to the difference in the number of parameters estimated between them. In some cases, the ω ratio was undefined for a given branch of the phylogeny when estimates of d_N were positive and those for d_S were zero. A two-ratio model was then specified to estimate two ω values for the tree, one ratio for the branch of interest (ω_b) and another mean ratio (ω_0) for all remaining branches. The model was run with ω_b constrained to equal 1 (i.e., neutral) with the same mean ω_0 for all other branches. An LRT with 1 df was then used to infer whether ω_b was likely to be significantly different than 1.

Results

Nucleotide Polymorphism

DNA sequences were obtained from 117 individuals in six samples and yielded 78 unique alleles. Aligned sequences are provided in table S1 of Supplementary Material online (http://www.molbiolevol.org). Most sequences were 985 bp in length and consisted of 164 codons in three exons plus two introns of 117 and 376 bp, respectively. A total of 120 polymorphisms, including 88 singleton and 32 parsimony informative sites plus three indels were found among *Pan* I alleles. There was no evidence for *Taq* replication errors in five alleles that were amplified, cloned, and sequenced twice. Within coding regions, 61 mutations were observed at 57 segregating sites, resulting in 14 silent and 47 replacement substitutions, with the majority (75%) of nonsynonymous differences occurring as singletons (table 2).

Three amino acid replacement mutations within the IV1 region distinguished two distinct *Pan* I allelic lineages within pollock: a common allele type found in 94% of the sequences and a rare type found in only eight individuals. In keeping with nomenclature introduced by Pogson (2001) for *Pan* I alleles in Atlantic cod, the common and rare allele types in walleye pollock were designated as *Pan* I^C and *Pan* I^D, respectively. Excluding one sequence identified as a recombinant allele, three replacement mutations in the IV1 domain appear to be fixed between any two *Pan* I lineages (fig. 1). A minimum of two amino acid replacements had occurred between Atlantic cod and pollock *Pan* I allelic groups, including apparent fixation for different amino acids at codon position 48 in the IV1 domain of walleye pollock and Atlantic cod alleles. Higher levels of nucleotide diversity (π) and average number of nucleotide differences among alleles (k) were observed in the pollock *Pan* I^D lineage compared with the pollock *Pan* I^C lineage or with homologous regions of 25 *Pan* I^A and nine *Pan* I^B alleles in Atlantic cod (table 3). The common allele group in pollock, *Pan* I^C, was about equally divergent from *Pan* I^D alleles or from the I^A or I^B lineages in Atlantic cod, averaging 14 to 17 nucleotide differences among pairwise comparisons (table 4).

Three peaks of nucleotide diversity were evident in the pollock *Pan* I sequences (fig. 2). The first occurred in the second exon containing mutations in the IV1 domain fixed between *Pan* I^C and *Pan* I^D lineages. *Pan* I^D alleles were considerably more polymorphic than *Pan* I^C alleles in the second intron ($\pi = 0.0122$ versus 0.0036, respectively) where a second peak of diversity was observed. The third peak of polymorphism occurred in the small IV2 domain (114 bp) as a result of seven synonymous and 14 replacement mutations, including a total of five amino acid replacements in two adjacent codons (fig. 1). Four of the five amino acid polymorphisms occurred in both *Pan* I lineages; the asparagine/leucine combination of residues observed at codon positions 158 and 159 in *Pan* I^C alleles was not detected in the small number of *Pan* I^D sequences. Linkage disequilibrium was detected in 135 out of 6,670 pairwise comparisons among polymorphic sites, less than expected by chance alone (334), and 35 of these were significant after corrections for multiple tests. There was no

Table 2
Amino Acid Replacement Mutations in Walleye Pollock Pantophysin Domains

Codon Position	Amino Acid Change	Location	Singleton?	Classification
8	A→K	M1	N	R
11	A→K	M1	Y	R
19	I→V	IV1	Y	C
24	K→N	IV1	N	R
30	E→V	IV1	Y	R
31	I→V	IV1	Y	C
35	F→L	IV1	Y	C
46	S→P	IV1	N*	R
47	Y→C	IV1	Y	R
48	T→A	IV1	Y	R
48	T→K	IV1	N*[a]	R
51	T→N	IV1	N*[a]	R
55	G→D	IV1	Y	R
56	T→N	IV1	Y	R
56	T→S	IV1	Y	C
57	T→A	IV1	Y	R
68	S→P	IV1	Y	R
69	A→T	IV1	Y	R
70	E→G	IV1	Y	R
71	F→S	M2	Y	R
72	F→S	M2	Y	R
77	V→A	M2	N	R
81	L→P	M2	Y	R
82	Y→H	M2	Y	R
91	L→P	M2	Y	R
95	H→R	M3	Y	C
103	G→S	M3	Y	C
106	V→A	M3	Y	C
109	F→S	M3	Y	R
116	F→L	M3	Y	C
116	F→S	M3	Y	R
118	W→R	M3	Y	R
118	W→R	M3	Y	R
129	L→P	IV2	Y	C
140	S→G	IV2	Y	C
147	V→A	IV2	N	C
150	S→F	IV2	Y	R
151	D→G	IV2	Y	R
152	G→R	IV2	Y	R
156	A→V	IV2	Y	C
156	A→T	IV2	Y	R
158	S→I	IV2	N	R
158	S→N	IV2	N	C
159	G→V	IV2	N	R
159	G→L	IV2	N	R
159	G→M	IV2	N	R
162	M→V	IV2	Y	C

NOTE.—M1, M2, and M3 are the first, second, and third transmembrane domains, respectively; IV1 and IV2 are the first and second intravesicular domains, respectively. Asterisks show replacement mutations fixed between *Pan* I^C and I^D alleles. Classification refers to conservative (C) and radical (R) replacements based upon Grantham (1974).

[a] Excluding one recombinant *Pan* I^D allele.

evidence for linkage disequilibrium between replacement mutations in the IV2 domain with those in the IV1 region used to define the *Pan* I^C and I^D allelic groups.

Tests of Neutrality

Tajima's *D* statistic and Fu and Li's D* and F* tests (Fu and Li 1993) for selective neutrality were statistically significant for *Pan* I^C alleles ($P < 0.05$ in all comparisons), but no significant deviations from neutral expectations were observed in *Pan* I^D alleles. Under a model of neutral evolution, the ratio of replacement to synonymous mutations (R/S ratio) for fixed differences between species is expected to be equal to the R/S ratio for polymorphic sites within species (McDonald and Kreitman 1991), and this expectation forms the basis of the HKA test (Hudson, Kreitman, and Aguadé 1987). We applied this rationale in tests of selective neutrality between the extant pantophysin lineages occurring in Atlantic cod and pollock. If directional selection has contributed to the fixation of replacement mutations between two lineages, the R/S ratio of these fixed differences (divergence) should exceed R/S based upon polymorphism within lineages. Conversely, retention of ancestral polymorphisms by positive selection acting upon nonsynonymous sites could result in a larger R/S ratio of polymorphic sites as opposed to those that are fixed. HKA tests at the *Pan* I locus were not significant between any two allelic lineages. Pollock I^C and I^D alleles had half of the number of fixed amino acid replacements (three) relative to the two cod *Pan* I lineages, no fixed synonymous differences compared with each other, and a R/S ratio of polymorphic sites of 3.07. M-K tests (McDonald and Kreitman 1991) of the neutral expectation that regions of the genome evolving at high rates will also have high levels of intraspecific polymorphism within species yielded a single significant test result between cod I^A and I^B lineages (Fischer's exact test, $P = 0.030$) that was consistent with the effects of selection at the *Pan* I locus described in Atlantic cod (Pogson 2001) but was not detected in comparisons between pollock allelic groups. These conventional tests of neutral evolution are sensitive to assumptions regarding underlying demographic population parameters. For example, the HKA test assumes constant population sizes with migration-drift equilibrium conditions for both species used in the comparison. This assumption was tested by examining the nucleotide mismatch frequency distribution (Rogers and Harpending 1992; Rogers 1995) of neutral sites in pollock *Pan* I alleles (fig. 3). The strong unimodal distribution of low frequency mutations is consistent with a pattern resulting either from population expansion or the effects of a recent selective sweep and suggest that nonequilibrium conditions likely exist at the *Pan* I locus, thus violating a fundamental assumption of the HKA test.

Nonsynonymous nucleotide substitution was highest in the two intravesicular domains of the *Pan* I locus (fig. 4). Mean d_N/d_S ratios exceeded unity in both domains and significant differences between mean d_S and d_N estimates were found in both the IV1 and IV2 domains (z test, $P < 0.001$ in both comparisons). There were no significant differences between estimates of mean d_S and d_N in other coding domains or when averaged over all coding regions combined. Nonsynonymous substitution in both *Pan* I^C and I^D alleles were significantly higher ($P < 0.001$) compared with homologous regions of Atlantic cod *Pan* I^A and I^B alleles (fig. 5). Silent substitutions per site were significantly higher for the pollock *Pan* I^C lineage, but not for *Pan* I^D, in comparisons with cod *Pan* I^A and I^B allelic groups.

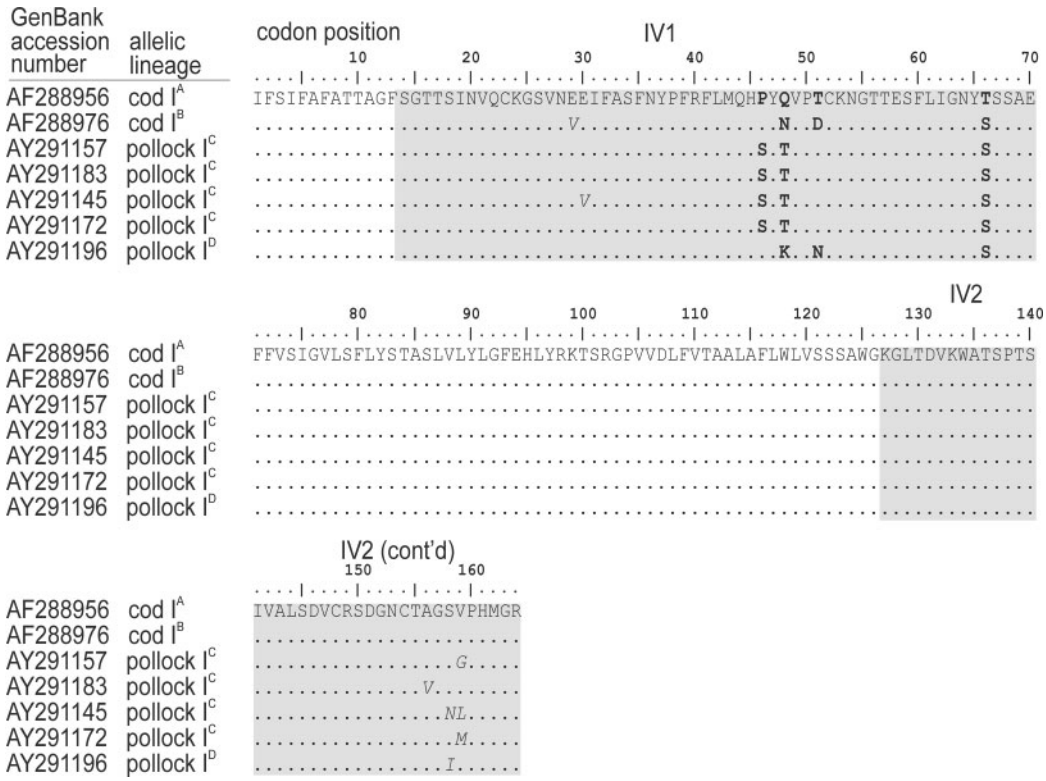

FIG. 1.—Representative amino acid sequences of *Pan* I allelic groups in walleye pollock and Atlantic cod. Fixed (bold) and polymorphic (italicized) residues within the first (IV1) and second (IV2) domains (shaded) are indicated.

Phylogeny of *Pan* I Alleles

A neighbor-joining phylogeny derived from *Pan* I sequences resolved the major allelic lineages reported for cod (Pogson 2001) and those observed within pollock, although support for nodes of the trees joining the groups was equivocal. Analyses using complete *Pan* I sequences or those from coding regions alone suggested a weak paraphyletic relationship among *Pan* I lineages, where pollock I^C and cod I^A alleles were more closely related to each other, as were cod I^B and pollock I^D alleles (fig. 6). Results from maximum-parsimony methods or analyses of third position sites in coding DNA (not shown) did not resolve this ambiguity. However, a phylogeny constructed from noncoding (intron) DNA sequences provided stronger support for reciprocal monophyly of *Pan* I lineages within each species (fig. 6*C*).

Results from maximum-likelihood models provided strong evidence for positive selection at the *Pan* I locus. Maximum-likelihood tree topologies used as input for the models in PAML were highly concordant with those derived from neighbor-joining and parsimony methods, indicating that the allelic phylogenies were sufficiently robust to produce consistent results from the models. The single-ratio model estimated an average d_N/d_S ratio (ω) exceeding 1 in the pollock data set and slightly less than unity for the combined data (table 5), suggesting an overall increase in nonsynonymous substitution rates in pollock *Pan* I lineages. Positive selection models (M2 and M8) estimated a similar proportion of sites subject to positive selection ($p_S = 0.02$) in the combined data set but a much higher fraction (0.14) for the pollock data alone. One codon in the IV1 domain (position 48) and two within the IV2 domain (positions 158 and 159) were consistently identified as sites with high likelihood for positive selection in both analyses (table 5).

Table 3
Pantophysin Nucleotide Polymorphism in Walleye Pollock and Atlantic Cod

Allele Type	n	S	k	π	θ
Pollock *Pan* I^C	71	106	4.99	0.0051	0.0239
Pollock *Pan* I^D	7	20	7.43	0.0078	0.0087
Cod *Pan* I^A	25	11	1.78	0.0018	0.0030
Cod *Pan* I^B	9	6	1.33	0.0013	0.0022

NOTE.—S is the number of polymorphic sites, k is the mean number of nucleotide differences, π is the nucleotide diversity, and θ is estimated theta per site.

Table 4
Pairwise Nucleotide Differences Between Pantophysin Lineages

Allele type	Pollock *Pan* I^C	Pollock *Pan* I^D	Cod *Pan* I^A	Cod *Pan* I^B
Pollock *Pan* I^C	—	13.69	12.98	16.62
Pollock *Pan* I^D	3[a]	—	12.29	13.84
Cod *Pan* I^A	7	5	—	1.780
Cod *Pan* I^B	11	7	11	—

NOTE.—Above diagonal: average number of nucleotide differences (k). Below diagonal: number of fixed nucleotide differences.

[a] After deletion of one recombinant *Pan* I^D allele.

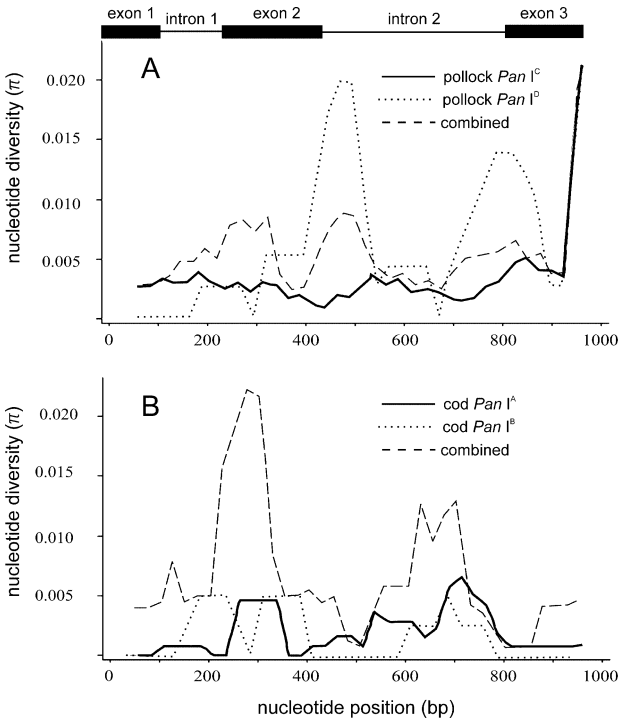

FIG. 2.—Sliding window view of nucleotide polymorphism across the *Pan I* locus. The window size is 100 bp and the step size is 25 bp. (*A*) Walleye pollock *Pan I* alleles. (*B*) Atlantic cod *Pan I* alleles. Exon 1 consists of the first transmembrane domain (M1, 13 codons) and a 27-codon portion of the first intravesicular domain (IV1). Exon 2 contains the remaining 30 codons of the IV1 domain, the second transmembrane domain (M2, 23 codons), a nine-codon cytoplasmic loop (CYTO) domain, and four codons of the third transmembrane domain (M3). Exon 3 consists of the remaining 20 codons of the M3 domain and the second intravesicular domain (IV2, 38 codons).

Selection models provided a significantly better fit to the data; comparisons of M2 versus single-ratio and neutral models yielded likelihood ratio test values of 27.00 and 19.54 (df = 2, $P < 0.001$) for the combined and pollock-only data sets, respectively. LRTs using the pollock data for the same model comparisons were also highly significant ($2\Delta l = 21.68$ and 20.88; df = 2, $P < 0.001$). Tests between the beta (M7) and beta&ω (M8) models likewise strongly supported positive selection ($2\Delta l = 19.88$ for combined data and 17.36 for pollock

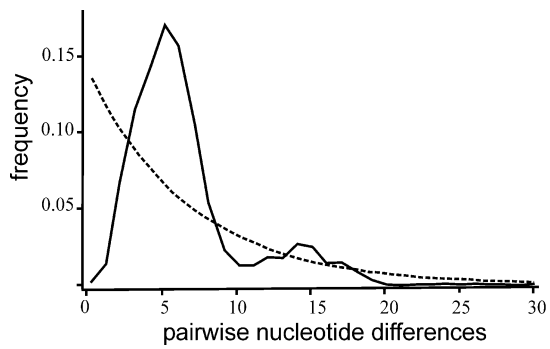

FIG. 3.—Nucleotide mismatch distribution of neutral sites in walleye pollock *Pan I* alleles (solid line) versus expected values under neutrality in a population of constant size (dashed line).

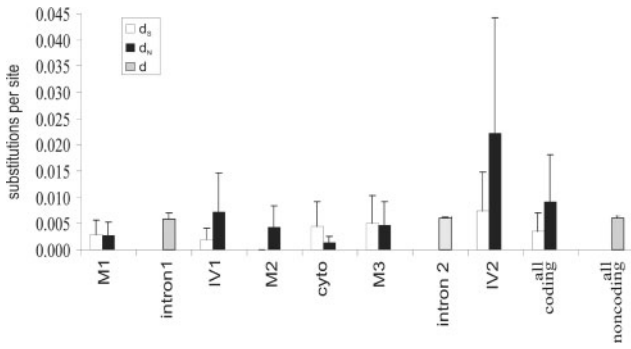

FIG. 4.—Synonymous (d_S), nonsynonymous (d_N), and neutral (d) substitution rates (±SE) in *Pan I* coding domains and introns. Domain abbreviations as in figure 2.

data; $P < 0.001$ for both). The free-ratio model did not perform significantly better than other models in either data set, but did provide estimates of ω ratios for branches in the phylogeny when estimates of d_S were greater than 0 (fig. 7). Nine branches in the tree have ω values below 1; the remaining branches had positive estimates of d_N but not for d_S. Two-ratio models, constraining one branch to neutrality (i.e., $\omega_b = 1$) were tested against an unconstrained two-ratio model for each of the branches where d_S had been estimated as zero using the free-ratio model. LRTs for all comparisons showed that fixing $\omega_b = 1$ resulted in a significantly poorer fit than an unconstrained model ($0.01 < P < 0.025$ in all tests) and did not support a hypothesis of neutral evolution along branches where d_N/d_S could not be estimated directly.

Discussion

Comparisons of DNA sequence differences within and between closely related species often provides insight into the temporal scales of molecular evolutionary processes. The persistence of nonsynonymous intraspecific and transspecific *Pan* I polymorphisms within and between walleye pollock and Atlantic cod suggests that they have been maintained by some form(s) of historical or contemporary selection. Neutral polymorphisms are transient phenomena and are not expected to persist long with respect to evolutionary time. Under balancing selection, polymorphisms may be maintained for much longer

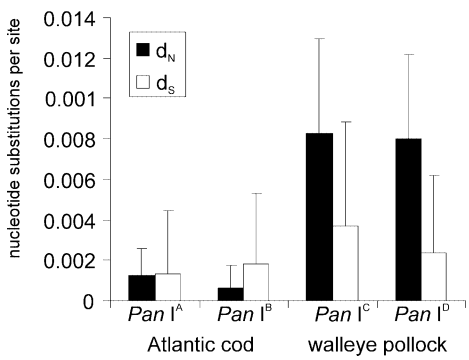

FIG. 5.—Synonymous (d_S) and nonsynonymous (d_N) substitution rates (±SE) in *Pan* I coding domains of walleye pollock and Atlantic cod.

periods than the expected neutral coalescence time of $2N_e$ generations (Takahata 1990), allowing for the accumulation of variation at linked neutral sites. Directional selection has the opposite effect, shortening coalescence times and reducing neutral linked variation through sweeps of derived advantageous mutations (Charlesworth, Morgan, and Charlesworth 1993; Otto 2000).

The observed patterns of *Pan* I diversity are not necessarily concordant with known examples of persistent balancing selection that results in many allelic variants, such as seen in *Mhc* class I and II alleles in fishes (Garrigan and Hedrick 2001). Rather, the divergence among allelic groups in pollock and Atlantic cod appears to have resulted from selective sweeps of advantageous *Pan* I lineages. Results from d_N/d_S ratio comparisons and maximum-likelihood analyses strongly suggest that positive selection acting at codon sites in the two intravesicular domains is largely responsible for the observed divergence. Multiple fixed replacement mutations in the IV1 region (i.e., positions 48 and 51) suggest that they may have important functional roles in the protein and polymorphic replacement mutations in the IV2 domain, identified as sites for selection through maximum-likelihood analyses, appear to be unique to pollock (fig. 1). Pogson and Mesa (2004) recently examined *Pan* I sequence variation in 18 species of gadid fishes and also found evidence for positive selection in both the IV1 and IV2 domains, although not at codon positions in the IV2 region (158 and 159) identified in this study. Indirect evidence for the effects of diversifying selection at the pantophysin locus comes from the distribution of segregating neutral sites in pollock *Pan* I alleles. Nucleotide diversity is high at the beginning and end of the second intron in the *Pan* I^D allelic group, where polymorphisms tightly linked to the second and third exons (containing sites under selection in the IV1 and IV2 domains, respectively) have accrued (fig. 2). This variation is not evident in the *Pan* I^C lineage, which now occurs at high frequency in the populations, or in Atlantic cod allelic groups characterized by low intra-allelic diversity and strong linkage disequilibrium (Pogson 2001).

A reduction in genetic diversity in the *Pan* I^C allelic group could result from demographic processes that increase rates of genetic drift (e.g., bottlenecks, founder events). However, the spread of advantageous mutations through selective sweeps also reduces variation at linked neutral sites (Maynard Smith and Haigh 1974; Fay and Wu 2000; Parsch, Meiklejohn, and Hartl 2001), producing a negative skew in their frequency spectrum (Andolfatto 2001; Galtier, Depaulis, and Barton 2000). The pollock *Pan* I^C lineage exhibits an excess of low frequency variants (fig. 3), a pattern that could result from a combination of directional selection and/or changes in population size but is unlikely to have arisen from background selection that does not predict changes in the frequency distribution of mutations (Hudson and Kaplan 1995).

The unique signature of positive selection is a high frequency of alleles derived from a successful variant (Przeworski 2002). Fay and Wu (2000) developed a statistic, H, to detect the hitchhiking effects of positive selection upon linked neutral sites by comparing high versus intermediate frequencies of derived neutral alleles. Several recent studies have reported significant H values for the *achaete, Acp26A*, and *dsat2* loci in *D. melanogaster* (Fay and Wu 2000; Takahashi et al. 2001) and the *janus-ocnus* locus in *D. simulans* (Parsch, Meiklejohn,

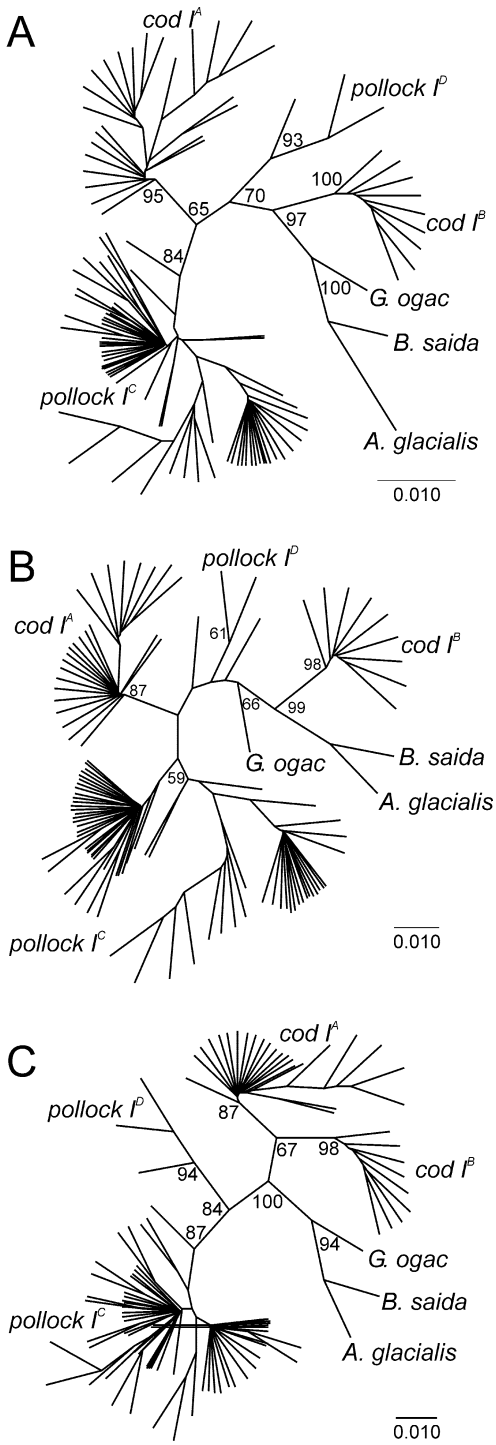

FIG. 6.—Neighbor-joining phylogenies of Kimura two-parameter distances from different regions of walleye pollock and Atlantic cod *Pan I* sequences using Polar cod, *Arctogadus glacialis*, as an outgroup root. (*A*) Full *Pan* I sequence (985 bp). (*B*) Coding DNA only (492 bp). (*C*) Two noncoding introns (493 bp). Percent bootstrap support is shown for nodes joining major allelic lineages.

Table 5
Maximum Likelihood Model Results for Combined Data and Walleye Pollock Pantophysin DNA Sequences

Model	Number of Parameters	*l*	ts/tv	Estimates of Parameters	Codon Sites for Positive Selection
Combined data					
Free-ratio	55	−935.59	3.34		
(M0) Single-ratio	29	−944.38	3.35	$\omega = 0.92$	
(M1) Neutral	29	−940.65	2.98	$\rho_0 = 0.55$, $\rho_1 = 0.45$	
(M2) Selection	31	−930.88	3.36	$\rho_0 = 0.43$, $\rho_1 = 0.54$, $\rho_S = 0.02$, $\omega_S = 20.95$	48**, 158, 159**
(M7) beta	30	−940.77	2.93	$p = 0.058$, $q = 0.098$	
(M8) beta&ω	32	−930.83	3.34	$\rho_0 = 0.98$, $\rho_S = 0.02$, $\omega_S = 19.59$, $\rho p = 1.23$, $q = 1.32$	48**, 158*, 159**
Pollock only					
Free-ratio	35	−827.73	4.88		
(M0) Single-ratio	19	−834.45	4.91	$\omega = 1.65$	
(M1) Neutral	19	−834.05	4.25	$\rho_0 = 0.42$, $\rho_1 = 0.58$	
(M2) Selection	21	−823.61	4.93	$\rho_0 = 0.11$, $\rho_1 = 0.87$, $\rho_S = 0.02$, $\omega_S = 47.05$	48, 158**, 159**
(M7) beta	20	−834.98	4.62	$p = 0.526$, $q = 0.001$	
(M8) beta&ω	22	−826.30	4.84	$\rho_0 = 0.86$, $\rho_S = 0.14$, $\omega_S = 12.47$, $p = 0.001$, $q = 1.54$	8, 24, 46, 48, 51, 57, 71, 140, 147, 150, 152, 158, 159 (all **)

NOTE.—Combined data consists of sequences from walleye pollock, Atlantic cod, and Greenland cod. Log-likelihood values (*l*) and estimated transition/transversion ratios (ts/tv) are given for each model. The value ω is the average d_N/d_S rate ratio estimated for the single-ratio model (M0), ω_S is the average d_N/d_S ratio for sites under positive selection in the models M2 and M8, p and q are the shape parameters for the beta distribution of ω in M7 and M8. ρ_0, ρ_1, and ρ_S are the proportions of codons subject to purifying, neutral, and positive selection, respectively. Probability $\omega_S < 1$; *** <0.001, ** <0.01, * <0.05.

and Hartl 2001), attesting to its power in detecting selection among closely related species. H is the difference between two estimators of the neutral parameter θ and requires identification of the ancestral allele, which can sometimes be inferred from comparisons with closely related species. Although phylogenetic analyses did not resolve the homology of *Pan* I alleles in Atlantic cod and pollock, there is evidence (see below) that the pollock I^D lineage is ancestral to the common *Pan* I^C allelic group.

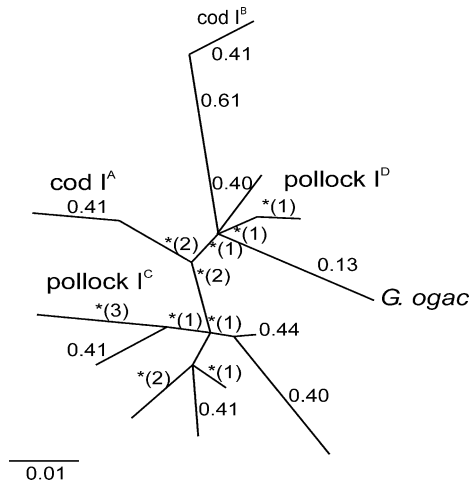

FIG. 7.—Maximum likelihood phylogeny of *Pan* I alleles in walleye pollock, Atlantic cod, and Greenland cod, *G. ogac*. Branch lengths represent estimated number of codon substitutions per site. Numbers are estimated d_N/d_S ratios (ω) for a given branch indicates branches having estimates of $d_N > 0$ and $d_S = 0$. Associated numbers in parentheses are the number of nonsynonymous codon changes along a given branch.

The H value estimated from the 376-bp second intron using *Pan* I^D alleles as the putative ancestors was large (10.31), suggesting that positive selection was responsible for the sweep event that has driven *Pan* I^C alleles to high frequencies. Taken together, the large H value and Tajima's *D* statistic results are inconsistent with either background selection or demographic processes such as population expansion but do support an adaptive hitchhiking model under positive selection (Otto 2000).

The H test has limited power to detect old sweeps as high-frequency alleles go to fixation and recombination breaks down linkage disequilibrium between the favored allele and neutral hitchhiking sites (Przeworski 2002). This suggests that selective sweeps in both Atlantic cod and pollock are evolutionarily recent and have not yet resulted in fixation for one allelic lineage in either species. Loss of variation in regions of low recombination appears to be mainly determined by hitchhiking unless background selection is very strong (Kim and Stephan 2000). This seems to be the case in Atlantic cod, where much silent variation has been purged through selective sweeps (Pogson 2001). In pollock, the hitchhiking is incomplete, as recombination has prevented the complete loss of ancestral allelic variation. Although sweeps in both species may be regarded as recent, the overall higher levels of nucleotide variation and evidence for intragenic recombination in pollock *Pan* I allelic groups suggest that the observed polymorphisms may have persisted for a considerably longer period of time in pollock than in cod.

The existence of two pantophysin lineages in both cod and pollock raises questions regarding speciation and phylogeographic origins of both species. Phylogenetic

analyses consistently placed the cod I^A lineage closer to pollock *Pan* I^C alleles and placed cod *Pan* I^B alleles with the other gadid species (fig. 5). These results are consistent with those from a recent phylogenetic survey of *Pan* I variation in 18 species of gadid fishes (Pogson and Mesa 2004) that provided some evidence for classifying *G. morhua* and *T. chalcogramma* as sister taxa descended from an ancestor giving rise to Pacific cod, *Gadus macrocephalus*. Only the pollock *Pan* I^C lineage was included in their analysis, and it grouped with the cod *Pan* I^A allele because of a shared replacement mutation at codon position 51 in the IV1 domain (fig. 1). This codon was identified as a site with a high likelihood for positive selection in both Atlantic cod (Pogson and Mesa 2004) and walleye pollock (table 5); thus, convergent evolution may have produced this phylogenetic result. Although we did not examine *Pan* I DNA sequences from *Gadus macrocephalus* in this study, we note that all three amino acid residues at codon positions defining the rare pollock Pan I^D allelic lineage (46, 48, and 51) are identical with those found in *G. macrocephalus* or *G. ogac*, but none are shared with *G. morhua* (fig. 1). Although these similarities could be homoplasious rather than homologous, the greater amount of neutral variation exhibited by the rare pollock Pan I^D lineage may represent a trans-species polymorphism predating speciation of pollock and Atlantic cod from the *G. macrocephalus* lineage. This interpretation is in accord with a hypothesis recently presented by Pogson and Mesa (2004) for biogeographic origins of *T. chalcogramma* and *G. morhua* in the Pacific from a lineage leading to *G. macrocephalus*, followed by subsequent invasion of the north Atlantic by *G. morhua*. Our results do not appear to support a hypothesis for separate biogeographic origins of *G. macrocephalus* and *T. chalcogramma* resulting from successive invasions by an Atlantic ancestor of *G. morhua*, as proposed by Carr et al. (1999) from analyses of mtDNA sequence data.

In summary, evidence for positive selection at the *Pan* I locus in walleye pollock has been inferred directly from variation in d_N/d_S ratios and indirectly through the frequency spectra of linked neutral variation. Results from maximum-likelihood models of codon evolution indicate that several replacement mutations within the intravesicular domains of the *Pan* I locus are responsible for the divergence among allelic groups, a result consistent with the general observation that positive selection alters only a few sites at different times but with potentially large effects (e.g., Takahashi et al. 2001). The detection of this locus in a small screening of anonymous nuclear cDNA clones (Pogson, Mesa, and Boutilier 1995) suggests that positive selection may operate on more loci than previously suspected, an assumption that has some support from human studies, where positive selection appears to be responsible for a large fraction of protein evolution (Fay, Wycoff, and Wu 2001). Intraspecific (Pogson 2001; this study) and interspecific comparisons (Pogson and Mesa 2004) of *Pan* I sequences indicate that sweeps resulting from positive selection appear to be a common mode of evolution at this locus in gadid species. Future research efforts should be directed toward understanding the physiological mechanism(s) responsible for positive selection and the universality of this phenomenon in other fishes.

Supplementary Material

Table S1 of aligned *Pan* I DNA sequences is available online at the MBE Web site (http://www.molbiolevol.org).

Acknowledgments

The authors are grateful to Takahashi Yanagimoto, Jim Seeb, and Lisa Seeb for contributing samples used in this study. We thank G. H. Pogson for providing *Pan* I sequences for *Arctogadus glacialis* and *Boreogadus saida* used in the phylogenetic analyses and for helpful comments during the early stage of manuscript preparation. Two anonymous referees gave us valuable advice for improving the manuscript. This research was supported by Washington State Sea Grant (Project # 61–9242) and the Alaska Fisheries Science Center and is contribution FOCI-0487 to NOAA's Fisheries-Oceanography Coordinated Investigations.

Literature Cited

Akashi, H. 1999. Within- and between-species DNA sequence variation and the 'footprint' of natural selection. Gene **238**:39–541.

Andolfatto, P. 2001. Adaptive hitchiking effects on genome variability. Curr. Opin. Genet. Dev. **11**:635–641.

Brooks, C. C., P. E. Scherer, K. Cleveland, J. L. Whittmore, H. F. Lodish, and B. Cheatham. 2000. Pantophysin is a phosphoprotein component of adipoctye transport vesicles and associates with GLUT4-containing vesicles. J. Biol. Chem. **275**:2029–2036.

Carr, S. M., D. S. Kivlichan, P. Pepin, and D. C. Crutcher. 1999. Molecular systematics of gadid fishes: implications for the biogeographic origins of Pacific species. Can. J. Zool. **77**: 19–26.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics **134**:1289–1303.

Endo, T., T. Ikeo, and T. Gojobori. 1996. Large scale search for genes on which positive selection may operate. Mol. Biol. Evol. **13**:685–690.

Fay, J., and C.-I. Wu. 2000. Hitchiking under positive Darwinian selection. Genetics **155**:1405–1413.

Fay, J., G. J. Wyckoff, and C.-I. Wu. 2001. Positive and negative selection on the human genome. Genetics **158**:1227–1234.

Felsenstein, J. 1993. PHYLIP (phylogeny inference package). Version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle.

Fevolden, S. E., and G. H. Pogson. 1997. Genetic divergence at the synaptophysin (*Syp* I) locus among Norwegian coastal and north-east Arctic populations of Atlantic cod. J. Fish. Biol. **51**:895–908.

Ford, M. 2002. Applications of selective neutrality tests to molecular ecology. Mol. Ecol. **11**:1245–1262.

Fu, Y.-X., and W.-H. Li. 1993. Statistical tests of neutrality of mutations. Genetics **133**:693–709.

Galtier, N., F. Depaulis, and N. H. Barton. 2000. Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. Genetics **155**:981–987.

Garrigan, D., and P. W. Hedrick. 2001. Class I *Mhc* polymorphism and evolution in endangered California chinook and other Pacific salmon. Immunogenetics **53**:483–489.

Goldman, N. and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**:725–736.

Grantham, R. 1974. Amino acid difference formula to help explain protein evolution. Science **18**:862–864.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. **41**:95–98.

Hudson, R. R., and N. L. Kaplan. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111**:147–264.

———. 1995. Deleterious background selection with recombination. Genetics **141**:1605–1617.

Hudson, R. R., M. Kreitman, and M. Aguadé. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **116**:153–159.

Jónsdóttir, Ó. D. B., A. K. Daníelsdóttir, and G. Nævdal. 2001. Genetic differentiation among Atlantic cod (*Gadus morhua* L.) in Icelandic waters: temporal stability. ICES J. Mar. Sci. **58**:114–122.

Jónsdóttir, Ó. D. B., A. K. Imsland, A. K. Daníelsdóttir, V. Thorsteinsson, and G. Nævdal. 1999. Genetic differentiation among Atlantic cod in south and south-east Icelandic waters: synaptophysin (*Syp* I) and haemoglobin (*HbI*) variation. J. Fish Biol. **54**:1259–1274.

Karlsson, S., and J. Mork. 2003. Selection-induced variation at the pantophysin (*Pan* I) in a Norwegian fjord population of cod (*Gadus morhua* L.). Mol. Ecol. **12**:3265–3274.

Kim, Y., and W. Stephan. 2000. Joint effects of genetic hitchhiking and background selection on neutral variation. Genetics **155**:1415–1427.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

Kreitman, M., and R. R. Hudson. 1991. Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127**:565–582.

Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favorable gene. Genet. Res. **23**:23–35.

McDonald, J., and M. Kreitman. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351**:652–654.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

Nielsen, R., and Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148**:929–936.

Otto, S. P. 2000. Detecting the form of selection from DNA sequence data. Trends Genet. **16**(12):526–529.

Parsch, J., C. D. Meiklejohn, and D. L. Hartl. 2001. Patterns of DNA sequence variation suggest the recent action of positive selection in the *janus-ocnus* region of *Drosophila simulans*. Genetics **159**:647–657.

Pogson, G. H. 2001. Nucleotide polymorphism and natural selection at the pantophysin (*PAN* I) locus in the Atlantic cod, *Gadus morhua* (L.). Genetics **157**:317–330.

Pogson, G. H., and S.-E. Fevolden. 2003. Natural selection and the genetic differentiation of coastal and Arctic populations of the Atlantic cod in northern Norway: a test involving nucleotide sequence variation at the pantophysin (*Pan* I) locus. Mol. Ecol. **12**:63–74.

Pogson, G. H., and K. A. Mesa. 2004. Positive Darwinian selection at the pantophysin (*Pan* I) locus in marine gadid fishes. Mol. Biol. Ecol. **21**:65–75.

Pogson, G. H., K. A. Mesa, and R. G. Boutilier. 1995. Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci. Genetics **139**:375–385.

Pogson, G. H., C. T. Taggart, K. A. Mesa, and R. G. Boutilier. 2001. Isolation by distance in the Atlantic cod, *Gadus morhua*, at large and small geographic scales. Evolution **55**:131–146.

Przeworski, M. 2002. The signature of positive selection at randomly chosen loci. Genetics **160**:1179–1189.

Rice, W. R. 1989. Analyzing tables of statistical data. Evolution **43**:223–225.

Rogers, A. R. 1995. Genetic evidence for a Pleistocene population explosion. Evolution **49**:608–615.

Rogers, A. R., and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. **9**:552–569.

Rozas, J., and R. Rozas. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15**:174–175.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for constructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

Su, C., V. K. Nguyen, and M. Nei. 2002. Adaptive evolution of variable region genes encoding an unusual type of immunoglobin in camelids. Mol. Biol. Evol. **19**:205–215.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics **105**:437–460.

———. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**:585–595.

Takahashi, A., S.-C. Tsaur, J. A. Coyne, and C.-I. Wu. 2001. The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. Proc. Nat. Acad. Sci. USA **98**:3920–3925.

Takahata, N. 1990. A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc. Natl. Acad. Sci. USA **87**:2419–2423.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4637–4680.

Windoffer, R., M. Borchert-Stuhltrager, N. Haass, S. Thomas, M. Hergt, C. J. Bulitta, and R. E. Leube. 1999. Tissue expression of the vesicle protein pantophysin. Cell Tissue Res. **296**:499–510.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS **13**:555–556.

Yang, Z., and J. P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. **15**:496–503.

Yang, Z., R. Nielsen, N. Goldman, and A.-M. Krabbe Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155**:431–449.

Yang, Z. and W. J. Swanson. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol. Biol. Evol. **19**:49–57.

David Rand, Associate Editor

Accepted March 17, 2004