Gene 515 (2013) 28–33

Contents lists available at SciVerse ScienceDirect

Gene

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

Accumulation of slightly deleterious mutations in the mitochondrial genome: A hallmark of animal domestication

Austin L. Hughes *

Department of Biological Sciences, University of South Carolina, Coker Life Sciences Bldg., 700 Sumter St., Columbia SC 29208, USA

ARTICLE INFO

Article history:
Accepted 29 November 2012
Available online 10 December 2012

Keywords:
Domestication
Mitochondrial DNA
Purifying selection
Slightly deleterious mutations

ABSTRACT

The hypothesis that domestication leads to a relaxation of purifying selection on mitochondrial (mt) genomes was tested by comparative analysis of mt genes from dog, pig, chicken, and silkworm. The three vertebrate species showed mt genome phylogenies in which domestic and wild isolates were intermingled, whereas the domestic silkworm (*Bombyx mori*) formed a distinct cluster nested within its closest wild relative (*Bombyx mandarina*). In spite of these differences in phylogenetic pattern, significantly greater proportions of nonsynonymous SNPs than of synonymous SNPs were unique to the domestic populations of all four species. Likewise, in all four species, significantly greater proportions of RNA-encoding SNPs than of synonymous SNPs were unique to the domestic populations. Thus, domestic populations were characterized by an excess of unique polymorphisms in two categories generally subject to purifying selection: nonsynonymous sites and RNA-encoding sites. Many of these unique polymorphisms thus seem likely to be slightly deleterious; the latter hypothesis was supported by the generally lower gene diversities of polymorphisms unique to domestic populations in comparison to those of polymorphisms shared by domestic and wild populations.

1. Introduction

The availability of genomic data for domestic animals and plants has stimulated interest in discovering the genetic changes underlying the distinctive characteristics of domestic species (Amaral et al., 2011; Andersson, 2011; Andersson and Georges, 2004; MacEachern et al., 2009; Rubin et al., 2010). Some such genetic changes presumably involve the loci responsible for traits consciously or inadvertently subjected to artificial selection by humans in the process of domestication (Andersson and Georges, 2004). In addition, certain neutral or even deleterious alleles may be fixed because they are linked to alleles selectively favored in the process of domestication (Fisher, 1930). Other genetic changes may involve the relaxation of purifying selection after domestication, as a result of the less stringent selective pressures in the domestic environment (Björnerfeldt et al., 2006; Cruz et al., 2008; MacEachern et al., 2009). Because domestication may involve population bottlenecks, the effectiveness of purifying selection in eliminating slightly deleterious mutations may also be reduced in domesticated populations (Cruz et al., 2008).

Comparisons of the ratio of nonsynonymous to synonymous substitutions have been used to test for a relaxation of purifying selection on both mitochondrial and nuclear genes of the domestic dog (Björnerfeldt et al., 2006; Cruz et al., 2008) and on nuclear genes of domestic cattle (MacEachern et al., 2009). Mitochondrial (mt) genomes are expected to particularly prone to the accumulation of slightly deleterious mutations because of their high mutation rate and lack of recombination (Hasegawa et al., 1998; Hughes and Hughes, 2007; Rand and Kann, 1996).

Although some studies have claimed evidence of positive selection on mt protein-coding genes (Bazin et al., 2006; Li et al., 2010; Mishmar et al., 2006) and RNA-encoding genes (Ruiz-Pesini and Wallace, 2006), the methods used in these analyses are of questionable validity (Hughes, 2007; Hughes and Friedman, 2008; Hughes et al., 2008). Moreover, these studies have at best merely reported statistical patterns allegedly consistent with positive selection. Not even a single study has shown experimentally that any of the allegedly selectively favored substitutions is actually associated with an advantageous phenotype.

By contrast, there is strong evidence of purifying selection on mt protein-coding genes. First, mt protein coding genes almost always show higher numbers of synonymous substitutions per synonymous site (d_S) than of nonsynonymous substitutions per nonsynonymous site (d_N), a pattern indicating that purifying selection has acted to remove deleterious nonsynonymous mutations (Hughes and Hughes, 2007; Kumar, 1996). Moreover, there are numerous known diseases linked substitutions in mt protein-coding and RNA-encoding genes of humans and other mammals (Baranowska et al., 2009; Schaeffer et al., 2001; Wallace, 1992). The existence of such mt-linked disorders supports the hypothesis of strong purifying selection by proving evidence of mutations that have a deleterious phenotypic effect (Kimura and Ohta, 1973).

Abbreviations: ML, maximum likelihood; mt, mitochondrial; SNP, single nucleotide polymorphism; π_S, nucleotide diversity at synonymous sites; π_N, nucleotide diversity at nonsynonymous sites.

* Tel.: +1 803 777 9186; fax: +1 803 777 4002.
E-mail address: austin@biol.sc.edu.

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.11.064

Phylogenies of mt genomes of domestic animal species and their closest wild relatives have revealed surprising complexity, with wild and domestic genomes not forming separate clades (Björnerfeldt et al., 2006; Giuffra et al., 2000; Larson et al., 2005; Liu et al., 2006). Such topologies may arise from independent domestication events, from domestication of populations involving several distinct matrilines, from post-domestication introgression, or from some combination of these processes. Nonetheless, on the hypothesis that domestication relaxes purifying selection present in the wild, one would expect to see the accumulation of slightly deleterious mutations in domesticated populations whatever their origin. Conversely one would expect to see intensified purifying selection in free-living populations, whether those represent the wild ancestral species or feral populations of domestic origin. Here I test this prediction using published sequences of complete mitochondrial genomes of the domestic dog, pig, chicken, and silkworm.

2. Methods

2.1. Sequence data

Complete or nearly complete mt genome sequences were obtained for the following taxa: 254 from the domestic dog (*Canis lupus familiaris*) and 19 from the wolf (*Canis lupus*); 59 from the domestic pig and 27 from wild boar (*Sus scrofa*); 41 from the domestic chicken (*Gallus gallus domesticus*) and 17 from the red junglefowl (*Gallus gallus*); and 33 from the domestic silkworm (*Bombyx mori*) and 15 from its closest known wild relative (*Bombyx mandarina*). Sequences were aligned by the CLUSTALX program (Thompson et al., 1997). When pairwise comparisons were made among a set of aligned sequences, any site at which the alignment postulated a gap in any of the sequences was excluded from the analysis so that a comparable set of sequences was available for each pairwise comparison. Phylogenetic trees were rooted with appropriate outgroup taxa: (1) for the dog, four sequences from the coyote (*Canis latrans*); (2) for the pig, a sequence from the common warthog (*Phacochoerus africanus*); for the chicken, a sequence from the green junglefowl (*Gallus varius*); and (4) for the silkworm, one sequence from the oriental fruit moth (*Grapholita molesta*) and one sequence from the gypsy moth (Lymantria dispar). For Genbank identifiers (gi numbers) of all sequences, see Supplementary Figs. S1–S4.

2.2. Statistical methods

Phylogenetic trees were constructed on the basis of the entire mt genome DNA sequence by the maximum likelihood (ML) method in the MEGA program, version 5.05 (Tamura et al., 2011). The Model test function in MEGA was used to choose models for ML analyses by the Bayes Information Criterion (BIC). The reliability of branching patterns in ML trees was tested by bootstrapping (1000 samples). The following DNA sequence evolution models were used: (1) for the dog, GTR + I; (2) for the pig, TN93 + G + I; (3) for the chicken, HKY + G + I; and for the silkworm, GTR + G + I.

The number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous substitutions per nonsynonymous site (d_N) were estimated by Li's (1993) method, which takes into account transitional bias (known to be marked in mitochondrial genomes). In coding regions, the mean for all pairwise comparisons of d_S provided an estimate of nucleotide diversity at synonymous sites (π_S); and the mean for all pairwise comparisons of d_N provided an estimate of nucleotide diversity at nonsynonymous sites (π_N) (Nei and Kumar, 2000). The number of nucleotide substitutions per site (d) in non-protein-coding regions was estimated by maximum composite likelihood method in MEGA (Tamura et al., 2011); the nucleotide diversity (π) in non-protein-coding regions was estimated by the mean of d in all pairwise comparisons. The computation of π_S, π_N, and π was equivalent to equation 10.6, p. 256, of Nei (1987). Standard errors of π_S, π_N, and π were estimated by the bootstrap method, which takes into account the non-independence of pairwise comparisons (Nei and Kumar, 2000); and z-tests were used to test equality of nucleotide diversities in different genomic regions.

Gene diversity (Nei, 1987, p. 177) was estimated separately at individual single nucleotide polymorphism (SNP) sites, using the PolyAna program (Hughes, 2005; Hughes et al., 2003; Knapp et al., 2011); where x_i is the frequency of the ith allele (nucleotide) at a given locus (site), the gene diversity is $1 - \Sigma x_i^2$. Polymorphic (SNP) sites were classified as synonymous or nonsynonymous, based on the coding effect of the nucleotide change. There were certain SNP sites that could not be classified unambiguously as synonymous or nonsynonymous (ambiguous sites), either because both synonymous and nonsynonymous variants occurred at the same site or because, given polymorphic sites within a single codon, the coding effect of a given substitution depended on the pathway taken by evolution. Ambiguous sites constituted 15 of 1439 polymorphic sites in coding regions of the four domesticated taxa (1.0%); and 41 of 2055 (2.0%) of polymorphic sites in coding regions of the four wild taxa. Ambiguous sites were excluded from the analyses of gene diversity at individual polymorphic sites.

3. Results

3.1. Phylogenetic analyses

In rooted ML phylogenies of dog, pig, and chicken mt genomes, domestic and wild isolates were intermingled (Fig. 1 and Supplementary Figs. S1–S3). Often highly significant bootstrap support was obtained for clusters of sequences including one or more genomes of wild origin and one or more genomes of domestic origin (Fig. 1 and Supplementary Figs. S1–S3). For example, Fig. 1 illustrates the ML tree of chicken mt genomes, showing topology only. Although in this tree, many of the deeper relationships did not receive strong bootstrap support, there were five clusters with 96% bootstrap support or better that included sequences of both wild and domestic origin (Fig. 1). By contrast, in the case of the silkworm, all *B. mori* sequences clustered together in a monophyletic group that received 100% boostrap support (Fig. 2 and Supplementary Fig. S4). The *B. mori* cluster fell within *B. mandarina*, suggesting that the latter taxon is paraphyletic (Fig. 2 and Supplementary Fig. S4).

3.2. Nucleotide diversity

Synonymous (π_S) and nonsynonymous (π_N) nucleotide diversities were estimated for the concatenated coding sequences of the 9 mt protein-coding genes; π_S and π_N were estimated separately for domestic and wild isolates in each of the four taxa (Table 1). In each case, π_S was significantly greater than π_N for both domestic and wild isolates (Table 1). However, the π_N:π_S ratio was in each case higher in the domestic population than in the corresponding wild population (Table 1), consistent with relaxation of purifying selection in domestic populations. However, the differences between the π_N:π_S ratios of domestic and wild populations were generally small (Table 1).

Nucleotide diversity (π) in 12S rRNAs, 16S rRNAs, and tRNAs was significantly lower than the corresponding value of π_S in both domestic and wild populations of dog, pig, and chicken (Table 2). The same was true of wild but not domestic silkworm (Table 2). The absence of this pattern in the silkworm can probably be explained by a much lower overall level of polymorphism in *B. mori* than in the other domestic population examined. For example, π_S in *B. mori* was only about one third that in the domestic chicken, only about one fifth that in the domestic dog, and only about one tenth that in the domestic pig (Table 1).

The pattern of reduced nucleotide diversity RNA-encoding genes in comparison to synonymous sites in protein-coding genes supported the hypothesis that the RNA-encoding genes are subject to stronger purifying selection than are synonymous sites in protein-coding genes. By