Nonneutral Evolution and Differential Mutation Rate of Gender-Associated Mitochondrial DNA Lineages in the Marine Mussel Mytilus

Humberto Quesada, Mary Warren and David O. F. Skibinski

School of Biological Sciences, University of Wales, Swansea SA2 8PP, United Kingdom

Manuscript received November 11, 1997
Accepted for publication March 11, 1998

ABSTRACT

Mussels have two types of mitochondrial DNA (mtDNA). The M type is transmitted paternally, and the F type is transmitted maternally. To test hypotheses of the molecular evolution of both mtDNA genomes, 50 nucleotide sequences were obtained for 396 bp of the *COIII* gene of European populations of *Mytilus edulis* and the Atlantic and Mediterranean forms of *M. galloprovincialis.* Analysis based on the proportion of synonymous and nonsynonymous substitutions indicate that mtDNA is evolving in a non-neutral and complex fashion. Previous studies on American mussels demonstrated that the F genome experiences a higher purifying selection and that the M genome evolves faster. Here we show that these patterns also hold in European populations. However, in contrast to American populations, where an excess of replacement substitution between F and M lineages has been reported, a significant excess of replacement polymorphism within mtDNA lineages is observed in European populations of *M. galloprovincialis.* European populations also show an excess of replacement polymorphism within the F but not within the M genome with respect to American *M. trossulus,* as well as a consistent pattern of excess of rare variants in both F and M genomes. These results are consistent with a nearly neutral model of molecular evolution and a recent relaxation of selective constraints on European mtDNA. Levels of diversity are significantly higher for the M than F genome, and the M genome also accumulates synonymous and nonsynonymous substitutions at a higher rate, in contrast with earlier reports where no difference for the synonymous rate was observed. It is suggested that a subtle balance between relaxed selection and a higher mutation rate explains the faster evolutionary rate of the M lineage.

RECENT reports have challenged the axiom that mitochondrial DNA (mtDNA) evolves under a strictly neutral model of molecular evolution (reviewed in Ballard and Kreitman 1995; Rand and Kann 1996). Most of these reports on mtDNA also show a variety of complex patterns of variation, suggesting that mtDNA evolution can be governed by many of the same selective forces found to operate on the nuclear genome, but also by factors that do not affect nuclear variation (Ballard and Kreitman 1995). Tests of neutrality, beyond simply giving evidence for or against selection, can provide insights into the other diverse processes that shape and maintain genetic variation in nature.

Mussels of the genus Mytilus have an unusual mode of mtDNA inheritance that provides a novel model for investigating mechanisms of mtDNA evolution. Mussels have two mtDNA genomes called F and M (Fisher and Skibinski 1990). Females transmit the F genome to both daughters and sons, whereas males transmit the M genome to sons, thus the M genome is inherited paternally and the F genome is inherited maternally (Skibinski *et al.* 1994a,b; Zouros *et al.* 1994a,b). This mode of inheritance accounts for the high levels of heteroplasmy for the M and F genomes in males (Fisher and Skibinski 1990; Hoeh *et al.* 1991), and the high sequence divergence (>20%) between these two lineages (Skibinski *et al.* 1994b), whose separation predates current Mytilus taxa (Rawson and Hilbish 1995; Stewart *et al.* 1995; Quesada *et al.* 1996). Thus, Mytilus mtDNA provides a potentially powerful model for detecting the influence of natural selection because analyses can be applied at several different levels of divergence, between and within the F and M genomes, and between and within the different forms or species of Mytilus.

mtDNA is also noteworthy for evolving faster in Mytilus than in other metazoans, a feature that has been attributed to a relaxed selective constraint associated with biparental inheritance (Hoeh *et al.* 1996). Several studies have also suggested that the M genome is more polymorphic and evolves faster overall than the F genome (Skibinski *et al.* 1994b; Rawson and Hilbish 1995; Stewart *et al.* 1995; Quesada *et al.* 1996), though without statistical support. However, for American populations of *Mytilus edulis* and *M. trossulus,* Stewart *et al.* (1995, 1996) demonstrated a significantly higher replacement rate for the M than the F genome in the *COIII* gene and in a mtDNA segment with no assigned function. Stewart *et al.* (1996) also noted a higher

Corresponding author: David O. F. Skibinski, School of Biological Sciences, University of Wales, Swansea SA2 8PP, UK.
E-mail: d.o.f.skibinski@swansea.ac.uk

replacement rate for the M than the F genome in nonconserved regions, as well as a significant excess of replacement substitutions between M and F lineages when applying the McDonald-Kreitman test. They interpreted these observations as evidence for non-neutral evolution of Mytilus mtDNA associated with relaxed selection on the M genome. Because males carry both F and M genomes in their somatic tissue, but females usually only have the F genome, the M genome should experience more relaxed selection overall. While the departure from neutrality for mtDNA in American Mytilus taxa is opposite in direction to that reported in other non-Mytilus taxa, where excesses of replacement polymorphisms within lineages have been usually reported (e.g. Ballard and Kreitman 1995; Rand and Kann 1996), American mussels are also noteworthy for having strong nuclear / mtDNA incompatibilities leading to the breakdown of biparental inheritance in hybrids and to intrinsic barriers blocking mtDNA introgression between different species (Zouros *et al.* 1992; Rawson *et al.* 1996; Saavedra *et al.* 1996). The restricted mtDNA introgression between American Mytilus taxa is in clear contrast to the extensive levels of mtDNA introgression observed between European taxa (Quesada *et al.* 1995b, 1998), raising the possibility of diverse evolutionary forces governing mtDNA polymorphisms.

In this study, we assess this possibility by examining patterns of mtDNA evolution in European populations of *M. edulis* and in the recently reported Atlantic and Mediterranean forms of *M. galloprovincialis* (Quesada 1993). For comparative purposes, we examined the *COIII* gene for a segment that covers the same region that was analyzed previously in American mussels (Stewart *et al.* 1995, 1996). We ask the following questions: Are patterns of mtDNA nucleotide variation compatible with the expectations of a strictly neutral model when a combination of neutrality tests are applied? If not, is the high excess of replacement substitutions between F and M lineages in American mussels also seen in European mussels? Is there variation between European taxa in the degree of departure from neutral expectations? Are an evolutionary rate and diversity higher for the M than for the F genome general features of Mytilus mtDNA, and, if so, what is the role of the differences in mutation rate *vs.* relaxed constraint?

MATERIALS AND METHODS

Sampling: Mussels from the three Mytilus taxa were collected from five European localities that are known (Gosling 1992; Quesada 1993) to contain pure *M. edulis* (Redcar and Mk. Point, northeastern and southwestern Britain, respectively), Atlantic *M. galloprovincialis* (Vigo, northwestern Spain), and Mediterranean *M. galloprovincialis* (Peñiscola and Chioggia, eastern Spain and northeastern Italy, respectively). Mussels were sexed by microscopic examination of gonads. A total of 50 individuals (24 males and 26 females) were chosen at random from samples from these populations for sequence analysis.

DNA preparation and sequencing: Template for PCR amplification was prepared from single mussels using a CTAB-based method of mtDNA extraction (Fisher and Skibinski 1990). Extractions from females were used for PCR amplification of a 476-bp fragment of the F mtDNA COIII gene using the following Mytilus specific primers: forward FOR1 5′CCAAAC CCGTCATCTACTAG-3′ and reverse REV1 (Zouros *et al.* 1994b) 5′-ATGCTCTTCTTGAATATAAGCGTACC-3′, which correspond to nucleotide positions 805–824 and 1326–1301, respectively, of segment 5 described in Hoffmann *et al.* (1992). DNA from males was used for PCR amplification of the M genome using the following set of M specific primers (Skibinski *et al.* 1994b): forward FOR2 5′-AAACCCTTCGTCCAC AAGG-3′ and reverse REV2 5′-AGCCTTTTTGTCATCATT CTGT-3′, homologous to nucleotide positions 806–824 and 143–164 of segments 5 and 6, respectively. This 1.5-kb fragment was purified from 1% agarose gels and cleaned with Spin-X columns (Costar, Cambridge, MA). The eluted DNA was subsequently used as a source for the amplification of the same fragment as amplified from females using the primers FOR2 and REV1 given above. PCR reactions and cycling conditions were as described in Quesada *et al.* (1996), except that a total of 25 cycles were used. Double-stranded F and M PCR products were cleaned with the Quiaquick kit (Quiagen, Chatsworth, CA) before sequencing. Both PCR strands were sequenced using the Thermo Sequenase cycle sequencing kit (Amersham, Arlington Heights, IL) on an ALF automated sequencer (Pharmacia, Piscataway, NJ). About 200–300 ng of PCR product and 10 pM of primer were used per sequencing reaction.

Sequence analysis: Sequences were aligned using the ClustalW (Higgins *et al.* 1996) computer program. Nucleotide sequences were translated into amino acid sequences using the Drosophila mtDNA genetic code, as described in Hoffmann *et al.* (1992). Nucleotide diversity was calculated for each taxon using the Fu (1994) estimator of θ. This new estimator has a variance that is substantially smaller than that of any existing estimator by making full use of phylogenetic information (Fu 1994). The proportion of synonymous (K_S) and nonsynonymous (K_A) substitutions per site were calculated for all pairwise combinations of sequences according to the method of Comeron (1995). This method minimizes stochastic errors by separating the twofold degenerate sites into sites where only transitional and transversional substitutions are synonymous. The Poisson-corrected proportion of amino acid substitutions was also calculated for all pairwise comparisons of sequences.

The sequences were tested for departures from neutral expectations using several tests as implemented in the DnaSP package (Rozas and Rozas 1997). The McDonald and Kreitman test (McDonald and Kreitman 1991) was used to test that the ratio of replacement to silent substitutions should be the same within and between lineages. For their analysis, Stewart *et al.* (1996) pooled sequences of different taxa within the F and within the M lineages. Here we chose not to pool our data in this way because such a procedure has the potential to obscure any differences that might exist among taxa. Probability values for contingency tables were calculated by the Monte Carlo method (Roff and Bentzen 1989) using 10,000 permutations per analysis. To test for departures from neutrality in relation to the frequency distribution of F and M polymorphisms within each Mytilus taxon, we used the tests of Tajima (1989) and Fu and Li (1993). This was done for nondegenerate and fourfold degenerate sites to determine whether patterns of variation among lineages are as predicted by neutral theory (Gillespie 1991).

Phylogenetic relationships among sequences were esti-

mated using the neighbor-joining method (MEGA version 1.02; Kumar *et al.* 1993). The analyses were repeated using the Fitch-Margoliash method (PHYLIP 3.1; Felsenstein 1993). Pairwise distances between sequences were calculated using Kimura's two-parameter correction for multiple hits. Gaps in the sequences were removed in pairwise comparisons. The level of support for the resulting phylogenetic trees was determined using 1000 bootstrap replications for the neighbor-joining method and 100 replications for the Fitch-Margoliash method. Similar topologies were obtained using the two methods, thus for brevity, the neighbor-joining tree alone is presented.

RESULTS

Phylogenetic relationships: A total of 50 sequences (24 M and 26 F haplotypes) were scored for the same segment of 396 bp of the *COIII* gene. Polymorphic sites are given in the appendix. There are 172 polymorphic sites over the whole data set. For comparative purposes, 308 bp of these sequences were aligned against the homologous 308 bp of the seven published *COIII* sequences of the highly diverged American *M. trossulus* M and F haplotypes (Stewart *et al.* 1995). The full 396 bp of the *COIII* sequence was used in all those analyses not involving American *M. trossulus.*

The neighbor-joining tree (Figure 1) indicates a primary division of American and European sequences into highly diverged male (M lineage) and female (F lineage) types. This result is consistent with earlier reports in Mytilus (Rawson and Hilbish 1995; Stewart *et al.* 1995). The main branch leading to the M lineage is longer, as are secondary branches, compared with corresponding branches in the F lineage. The Poisson-corrected number of amino acid substitutions per site between European sequences and the highly diverged American *M. trossulus* sequences were significantly different (*z*-test, $P < 0.001$) between the F (0.018 ± 0.015) and M (0.100 ± 0.015) lineages. These results correspond with those published previously for American taxa (Rawson and Hilbish 1995; Stewart *et al.* 1995, 1996), showing a higher evolutionary rate for the M lineage. There is clear and substantial geographic structure evident in the variation between Atlantic and Mediterranean *M. galloprovincialis* (Quesada *et al.* 1995a,c). In the present study, Mediterranean *M. galloprovincialis* exhibit at least two highly diverged groups of F and M haplotypes that are not linked to particular geographical sites. F and M haplotype assemblages are not always concordant with the taxonomic identification of the individuals in which they are found. Hybridization and mtDNA introgression among European Mytilus taxa can explain these patterns of variation, as discussed elsewhere (Quesada *et al.* 1995b, 1996).

Diversity and divergence for replacement and synonymous changes: Estimates of nucleotide diversity for both F and M genomes, as measured by θ, are given in Table 1. Mediterranean *M. galloprovincialis* has the highest diversity value for both F and M genomes. Genetic diversity is higher for the M genome than for the F genome within all taxa, although the difference is never significant (*z*-test). However, a two-tailed Wilcoxon sign rank test applied for the five populations across the three taxa does give a significant ($P < 0.05$) result (Table 2). When synonymous and nonsynonymous substitutions are analyzed separately, nucleotide diversity is significantly higher ($P < 0.05$) for synonymous than for nonsynonymous substitutions (Table 2). Moreover, the M genome shows a genetic diversity significantly higher than the F genome for synonymous substitutions, but not for nonsynonymous substitutions. A three-way analysis of variance can be carried out on the data set of 20 values for nonsynonymous and synonymous substitutions given in Table 2. The two two-way interactions (population × type of substitution and genome × population) are nonsignificant when assessed against the three-way interaction. The interaction between genome and type of substitution is, however, significant when assessed against these other interactions pooled as error ($F_{1,12} = 7.836$, $P < 0.025$). These observations suggest differences in the evolutionary factors operating on each genome. However, stochastic factors could also account for a higher diversity of the M genome because local populations from the same taxa cannot be considered evolutionarily independent. However, synonymous substitutions are intermingled with nonsynonymous substitutions within both the F and M mtDNA molecules. Thus, the two types of substitutions should not show different patterns of variation between genomes as a result of drift, the opposite of what is observed. In addition, the above result is consistent with earlier claims suggesting higher diversity for the M than the F genome in populations not closely related to those analyzed here, such as American *M. edulis* and American *M. trossulus* (Stewart *et al.* 1995, 1996). These earlier reports did not provide statistical support for higher diversity for the M genome, nor did they consider the role of drift as a potential explanation for the observed differences between genomes. Consideration of all data sources suggests strongly that a higher diversity for the M genome is a general feature of Mytilus mtDNA—contradictory observations have never been reported (Skibinski *et al.* 1994b; Rawson and Hilbish 1995; Stewart *et al.* 1995, 1996; Quesada *et al.* 1996, 1998).

Table 3 gives the number of synonymous (K_S) and nonsynonymous (K_A) substitutions observed between European taxa in comparisons for the F or for the M genome. Both estimates are higher for the M lineage in five of the six comparisons, but the difference between M and F genomes is not significant, for K_A or K_S, in any one of these comparisons. The K_A/K_S ratio was unusually high, *e.g.*, compared with Stewart *et al.* 1996, in all three taxa, and F and M genomes displayed similar ratios within each of the taxa. This observation implies either that F and M mtDNA genomes are under a similar and dramatic relaxed selection, or that selective pres-

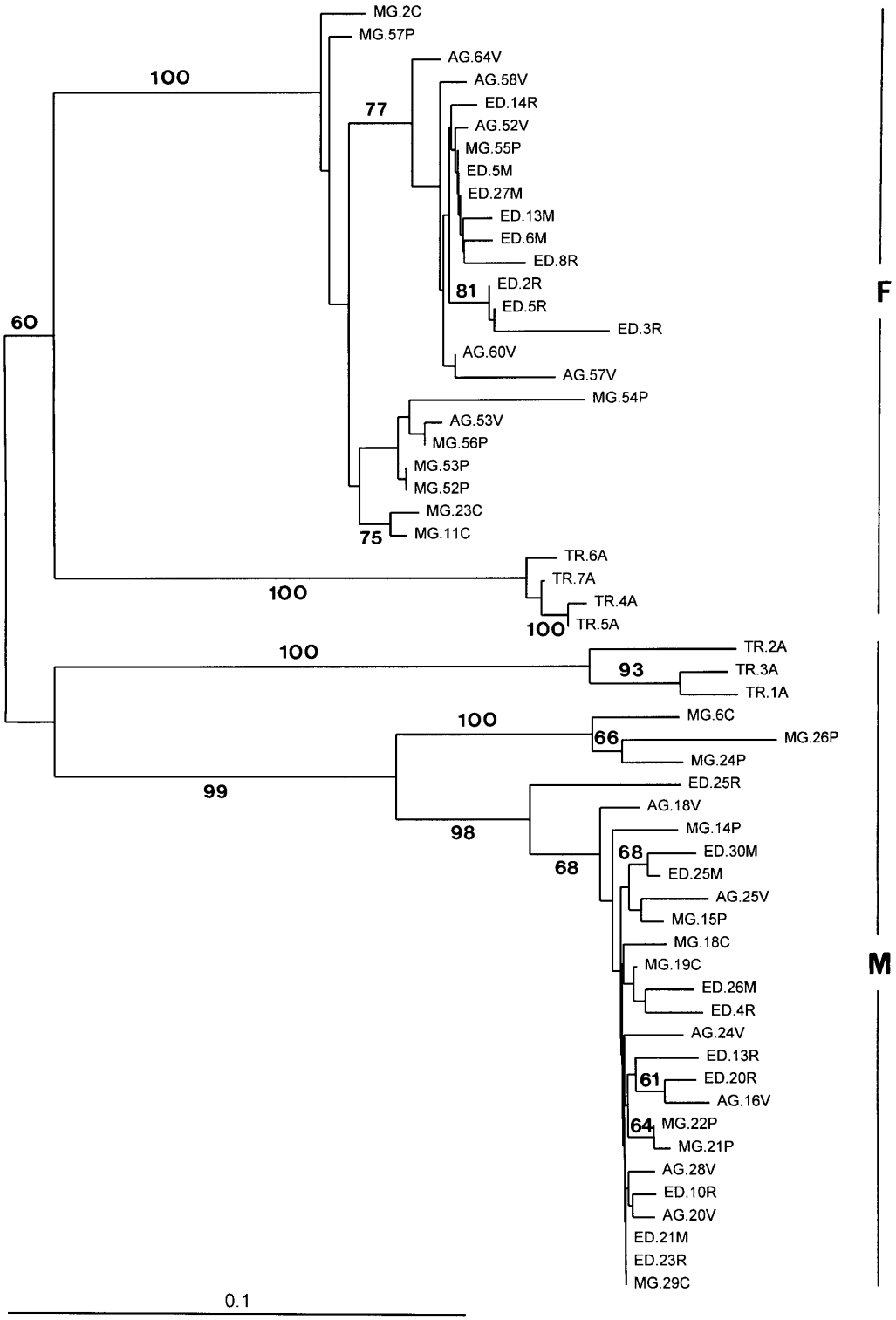

Figure 1.—Neighbor-joining tree of the 50 sequences from three European taxa, plus the seven homologous published sequences from American *M. trossulus* (Stewart *et al.* 1995). The tree is based on 308 nucleotides and is rooted between the male and female lines. Only bootstrap values >50 are presented. Sequences are indicated by an alphanumeric code indicating the taxa (ED, *M. edulis*; AG, Atlantic *M. galloprovincialis*; MG, Mediterranean *M. galloprovincialis*; TR, *M. trossulus*) followed by the individual and locality from which the sequence was obtained: M, Mk. Point; R, Redcar; V, Vigo; P, Peñiscola; C, Chioggia; A, North America.

sures of similar strength are operating on both genomes. When pooled European sequences are compared with the highly diverged American *M. trossulus* sequences, a remarkably different pattern is apparent (Table 3). First, both K_A and K_S are significantly ($P < 0.001$) greater for the M genome. Second, the K_A/K_S ratio is 3 times higher for the M lineage. Third, the K_A/K_S ratios are lower than those observed in European comparisons, 2 to 3 times lower for the M genome, but 11–19 times lower for the F genome. These contrasting results, when we compare closely related European sequences with highly diverged *M. trossulus* sequences, suggest that patterns of mtDNA variation are critically affected by the time scale considered. Thus, in the long term, the M

TABLE 1

Estimates of nucleotide variation for the F and M mtDNA genomes

Taxa	Genome	*N*	Polymorphic sites	θ[a] ± SE	Tajima's *D*[b]		Fu and Li's *D'*[c]	
					Zerofold	Fourfold	Zerofold	Fourfold
ED	F	9	26	14.204 ± 5.454	−1.633	−1.728*	−1.731	−1.943*
	M	10	46	28.873 ± 9.950	−1.383	−1.709	−1.634	−1.916
AG	F	6	29	20.058 ± 8.823	−1.453	−1.423	−1.490	−1.458
	M	6	29	24.447 ± 10.623	−1.299	−0.960	−1.228	−1.063
MG	F	9	44	29.262 ± 10.540	−1.505	−0.084	−1.576	−0.298
	M	10	69	32.512 ± 11.074	−1.174	−0.018	−1.107	0.343

[a] Estimate is from Fu (1994) and error includes stochastic variance.
[b] From Tajima (1989).
[c] From Fu and Li (1993).
* A departure from neutral expectation at the 5% level.

genome accumulates more synonymous and nonsynonymous substitutions than the F genome, although the nonsynonymous substitutions are accumulated at a higher rate. This observation is consistent with previous studies on American Mytilus taxa, suggesting a higher evolutionary rate and relaxed constraint for the M genome (Stewart *et al.* 1996).

Tests of neutrality: Table 4 summarizes the number of replacement and silent substitutions observed within and between F and M genomes for each European taxon. Because there has been some controversy about the proper method for counting substitutions in the McDonald-Kreitman test (Graur and Li 1991; Whittam and Nei 1991), differences between lineages were counted in three different ways. For method 1, a site that is fixed in one lineage and polymorphic in another is classified as a polymorphism (McDonald and Kreitman 1991). For method 2, a site fixed in one lineage and polymorphic for different substitutions in another is counted as a polymorphism and one fixed difference. This method attempts to cope with situations in which the observed polymorphism results from a single individual, most likely reflecting a multiple mutation after a fixed difference between the two highly diverged (>20%) F and M lineages has occurred. For method 3, branches on the phylogenetic tree were used to count mutations that appear on within-lineage branches and between-lineages branches (Nachman *et al.* 1994). These three methods resulted in similar patterns of association, as shown by a log linear analysis for each species ($P > 0.85$, for the three-way interactions). For the three methods considered, F/M interlineage comparisons reveal that Atlantic and Mediterranean *M. galloprovincialis* display a consistent pattern of excess replacement polymorphisms within mtDNA lineages over that predicted under a strictly neutral model of molecular evolution. The departure is always statistically significant for Atlantic *M. galloprovincialis*, and significant or marginally significant for Mediterranean *M. galloprovincialis*, depending on the method considered. By con-

TABLE 2

Fu (1994) estimates of diversity per population

Population (taxa)	Genome	*N*	θ ± SE		
			Total	Nonsynonymous	Synonymous
Mk. Point (ED)	F	4	5.802 ± 3.359	1.161 ± 0.916	4.247 ± 2.561
	M	4	9.022 ± 4.995	0.576 ± 0.566	8.619 ± 4.791
Redcar (ED)	F	5	11.508 ± 5.696	3.442 ± 1.991	5.925 ± 3.137
	M	6	26.309 ± 11.387	3.688 ± 1.988	19.859 ± 8.738
Vigo (AG)	F	6	20.058 ± 8.823	5.322 ± 2.694	10.314 ± 4.795
	M	6	24.447 ± 10.623	6.580 ± 3.229	15.229 ± 6.823
Chioggia (MG)	F	3	11.663 ± 7.164	7.584 ± 4.796	3.394 ± 2.343
	M	4	19.150 ± 10.091	5.048 ± 2.973	14.089 ± 7.549
Peñiscola (MG)	F	6	20.734 ± 9.101	5.402 ± 2.728	14.392 ± 6.487
	M	6	32.174 ± 13.753	12.097 ± 5.515	19.650 ± 8.619
Two-tailed Wilcoxon test (F *vs.* M)[a]:			$P < 0.05$	$P > 0.60$	$P < 0.05$

[a] The null hypothesis is also rejected ($P < 0.05$) for the 10 comparisons between synonymous and nonsynonymous substitutions.

TABLE 3

Jukes-Cantor corrected estimates of synonymous (K_S) and nonsynonymous (K_A) substitutions (Comeron 1995)

Compared taxa	Genome	K_A	K_S	K_A/K_S
ED to MG	F	0.011 ± 0.009	0.098 ± 0.031	0.112
	M	0.012 ± 0.010	0.130 ± 0.012	0.092
ED to AG	F	0.009 ± 0.008	0.046 ± 0.027	0.196
	M	0.008 ± 0.005	0.062 ± 0.036	0.129
AG to MG	F	0.013 ± 0.008	0.083 ± 0.031	0.157
	M	0.017 ± 0.010	0.120 ± 0.110	0.142
EUROPE to TR	F	0.007 ± 0.006	0.670 ± 0.050	0.010
	M	0.038 ± 0.006***	1.000 ± 0.063***	0.038

ED, *M. edulis*; MG, Mediterranean *M. galloprovincialis*; AG, Atlantic *M. galloprovincialis*; TR, American *M. trossulus*; EUROPE, pooled European taxa.

*** Departure ($P < 0.001$) from the null hypothesis of no difference in K_A or K_S between F and M genomes (*z*-test).

trast, *M. edulis* shows good agreement with neutral expectations. The ratio of polymorphic to fixed differences is up to two times higher for nonsynonymous than for synonymous substitutions for the two *M. galloprovincialis* taxa for all three methods.

Additional McDonald-Kreitman tests can be performed for each genome by comparing the two major branches separating American *M. trossulus* and European taxa (Figure 1), taking advantage of the prediction that the ratio of polymorphic to fixed differences should be the same in each partition of a neutral genealogy (Ballard and Kreitman 1994). In each of these further tests involving F/F or M/M intralineage comparisons, the three counting methods produced similar results, thus only those for method 1 are given (Table 5). For this analysis, a significant excess of replacement polymorphism is observed within the F genome ($P = 0.002$), and inspection of the data reveals that all these polymorphisms occur in European populations. By contrast, a nonsignificant result is obtained for the M genome ($P = 0.296$). This contrasting pattern between genomes receives additional support from a log linear analysis ($P = 0.008$ for the three-way interaction). A similar excess of replacement polymorphism within the European F genome ($P = 0.010$), but not within the European M genome ($P = 0.867$), is observed if European *M. edulis* sequences alone are compared with American *M. trossulus* sequences. This indicates that *M. edulis* contributes to the contrasting pattern between genomes in F/F and M/M intralineage comparisons, despite the conformity of *M. edulis* to neutral expectations in F/M interlineage comparisons. These results imply either a stronger selection for adaptive mutations or a more relaxed selection for the F genome in European taxa with respect to American populations. Divergence in the genetic background, following the isolation of American and European populations, could account for such differences in evolutionary factors operating on each mtDNA lineage.

Results of the Tajima (1989) and Fu and Li (1993) tests are given in Table 1. Consistent results are obtained with both methods. There is an indication of an excess of rare nucleotide polymorphisms within F and M lineages, as revealed by the consistently negative *D* estimates. However, the only significant ($P < 0.05$) result is for the F genome of *M. edulis* at fourfold degenerate sites. In all three Mytilus taxa, the F genome displayed more highly negative *D* values, consistent with higher levels of purifying selection for this genome. Several observations support the hypothesis that the significant *D* value in *M. edulis* almost certainly results from the pooling of sequences from geographical regions differing in haplotype frequencies. First, the two *M. edulis* populations included in this study (Redcar and Mk. Point) are heterogeneous ($P < 0.001$) for haplotype frequencies, as revealed by a geographical survey based on *Rsa*I and *Dde*I restriction patterns (C. Gallagher, unpublished results). Second, when the two *M. edulis* samples are analyzed separately, *D* values are substantially less negative in all cases ($D < -0.10$) and are not significant ($P > 0.10$). Third, the pooled *M. edulis* sample shows higher excesses of rare polymorphisms at fourfold than zerofold sites, but the opposite occurs in each single sample, in agreement with the pattern observed in the two other European taxa (Table 1).

DISCUSSION

Excess of replacement polymorphisms within European mtDNA lineages: Sequence data for Mytilus mtDNA show a significant excess of replacement polymorphisms in F/M interlineage comparisons for *M. galloprovincialis* and for the European F genome in intralineage comparisons between American *M. trossulus*

TABLE 4

Number of synonymous and nonsynonymous polymorphic (P) and fixed (F) substitutions between F and M mtDNA lineages of European Mytilus taxa using three different counting criteria

		Method 1			Method 2			Method 3		
Taxa		F	P	P/F	F	P	P/F	F	P	P/F
ED	Nonsynonymous	12	21	1.7	14	18	1.3	14	17	1.2
	Synonymous	47	53	1.1	51	50	1.0	57	61	1.1
	P		0.227			0.416			0.683	
AG	Nonsynymous	11	23	2.1	13	20	1.5	12	22	1.8
	Synonymous	46	42	0.9	52	37	0.7	57	39	0.7
	P		0.040			0.037			0.009	
MG	Nonsynonymous	8	45	5.6	10	42	4.2	11	42	3.8
	Synonymous	31	79	2.5	42	69	1.6	44	86	1.9
	P		0.050			0.012			0.073	

Probabilities from the Monte Carlo method (Roff and Bentzen 1989).

and European taxa. Nonsignificant results are obtained in other comparisons. These results provide evidence for both neutral and nonneutral evolution of Mytilus mtDNA, suggesting different evolutionary histories for each taxon.

In the recent work by Stewart *et al.* (1996) on American *M. trossulus* and American *M. edulis*, rejection of the null hypothesis was attributed to an excess of replacement substitutions between the F and M lineages. Our data show a departure from neutrality in the opposite direction. Several questions need to be addressed: (1) What factors are responsible for the observed excess of replacement polymorphisms within mtDNA lineages? (2) What accounts for the reversal of the pattern in American mussels? (3) What are the origins of the different patterns observed among European taxa and among genomes in relation to the extent of departure from neutral expectations?

It appears unlikely that the observed patterns could be generated by nonequilibrium conditions, in spite of good evidence supporting recent variation in population size and mtDNA introgression among European taxa (Quesada *et al.* 1995b, 1998). Such violation of equilibrium conditions might affect the validity of Tajima's and Fu and Li's tests, but not the McDonald-Kreitman test (McDonald and Kreitman 1991; Brookfield and Sharp 1994). On the other hand, mtDNA introgression should lead to significant excesses of polymorphism within European mtDNA lineages for both synonymous and nonsynonymous substitutions, not just for nonsynonymous substitutions, as observed here (Tables 4 and 5). Similarly, mtDNA introgression should generate significant excesses of polymorphisms within F and M genomes in intralineage comparisons of European sequences with American *M. trossulus*, not just within the F genome, as observed (Table 5). F and M lineages exhibit very similar ratios of replacement to silent substitutions across the three European taxa (Table 3) consistent with similar levels of introgression across genomes (Quesada *et al.* 1998). Conversely, the rejection of an equilibrium neutral model for *M. galloprovincialis* but not for *M. edulis* in F/M interlineage comparisons (Table 4) cannot be explained by introgression alone. Another possibility is that excess replacement polymorphism within European mtDNA lineages could be caused by some form of balancing selection. However, under this hypothesis, an excess of high-frequency polymorphism within European mtDNA lineages should be observed. The data are not consistent with this prediction because both Tajima's *D* and Fu and Li's *D′* show a consistent excess of low frequency polymorphisms.

A more likely explanation for excess replacement polymorphisms is that replacement variants are mildly deleterious and persist within mtDNA lineages as short-lived polymorphisms, but they do not persist long enough to become fixed between F and M lineages or between American and European populations. This explanation is supported by the dramatic decrease in F and M K_A/K_S ratios when the closely related European sequences are compared with the highly diverged Amer-

TABLE 5

Number of synonymous and nonsynonymous polymorphic (P) and fixed (F) substitutions between American *M. trossulus* and pooled European Mytilus taxa at F and M mtDNA lineages

	F genome		M genome	
	F	P	F	P
Nonsynonymous	0	21	4	29
Synonymous	27	48	18	77
P	0.002		0.296	

Probabilities from the Monte Carlo method (Roff and Bentzen 1989).

TABLE 6

Variances for the frequency distributions of the number of nucleotide substitutions per synonymous and nonsynonymous polymorphic sites

	F genome			M genome		
	Synonymous	Nonsynonymous	F^a	Synonymous	Nonsynonymous	F
ED	0.0094 (17)	0.0013 (9)	7.23**	0.0057 (38)	0.0050 (8)	1.14
AG	0.0083 (18)	0.0001 (11)	83.00***	0.0075 (20)	0.0039 (7)	1.92
MG	0.0179 (27)	0.0050 (17)	3.58**	0.0127 (49)	0.0076 (19)	1.67

[a] F-test for the null hypothesis of no differences in variance between synonymous and nonsynonymous sites.
**$P < 0.01$;
***$P < 0.001$.
The number of sites are indicated in parentheses. Sites with both synonymous and nonsynonymous substitutions were excluded.

ican *M. trossulus* sequences (Table 3), suggesting that a substantial proportion of the replacement polymorphism found within European mtDNA lineages never become fixed. This nearly neutral argument has been used to explain a similar pattern of excess replacement polymorphism in the mtDNA of *Drosophila melanogaster* (Kaneko *et al.* 1993; Rand and Kann 1996), mice (Nachman *et al.* 1994), and humans and African apes (Nachman *et al.* 1996). One prediction of the nearly neutral theory (Ohta 1992) is that because silent polymorphisms are more likely to be governed by neutral forces than replacement polymorphisms, there should be a lower variance in the frequency of replacement variants. Data for European taxa are consistent with this prediction (Table 6). Synonymous sites always show a higher variance in frequency than nonsynonymous sites, although this difference is only significant for the F genome. This result suggests higher levels of purifying selection for the F genome. This conclusion is reinforced by two further observations: (1) excess of low-frequency polymorphisms greater within the European F than the M genome revealed by Tajima's D and Fu and Li's D' estimates (Table 1) and (2) the lower K_A/K_S ratios for the F genome in American *M. trossulus*–European comparisons (Table 3). Other features of the data also fit the nearly neutral model. For both genomes, the nucleotide diversity is greater for synonymous than for nonsynonymous substitutions (Table 2), and excesses of rare substitutions are larger for zerofold than for fourfold degenerate sites in *M. galloprovincialis* (Table 1) and in individual *M. edulis* samples (see results).

Our data contrast with those of Stewart *et al.* (1996), where an excess of replacement substitutions between F and M lineages in American mussels for the same *COIII* gene region was reported. Nevertheless, both could be explained under the nearly neutral model. In a nearly neutral scenario, the magnitude of negative selection coefficients acting on mtDNA vary according to the reciprocal of the effective population size (Ohta 1972, 1992). Thus, repeated founder events or bottlenecks associated with glacial fluctuations could have led to a striking relaxation of selective constraints on slightly deleterious mutations in European mtDNA. This relaxation of constraints would have a more dramatic effect on the F lineage, given a higher level of purifying selection operating on this genome. This would explain the high and similar K_A/K_S ratios for both F and M lineages in European taxa and the significant excess of replacement polymorphisms within the European F genome with respect to American *M. trossulus.* In the long term, most slightly deleterious mutations would not persist, but the more relaxed selection for the M lineage would generate an excess of replacement substitutions between M and F lineages, as observed by Stewart *et al.* (1996). In conclusion, the deviations from neutrality opposite in direction for European and American mussels might have more to do with the time scale of selection than with a difference in the nature of evolutionary forces operating on mtDNA.

Similarly, the different patterns among European taxa with respect to the extent of deviation from neutral expectations may be related to differences in effective population size and levels of constraint. mtDNA genetic diversity is negatively correlated with latitude in European Mytilus populations, suggesting larger effective population sizes and more deme-structured populations in southern taxa (Quesada *et al.* 1995b). Thus, replacement mutations that might be neutral or nearly neutral in small Northern taxa populations (*M. edulis*) could be deleterious in Southern taxa (Mediterranean *M. galloprovincialis*).

However, despite the apparent fit of the data to the nearly neutral model, it is difficult to exclude the alternative hypothesis of a recent relaxation of selection on mtDNA that has occurred, not through a reduction in effective population size, but through a reduction in the magnitude of selection coefficients acting on mtDNA. This would result in some previously deleterious mutations becoming neutral and remaining within mtDNA lineages as polymorphisms. Support for this

explanation is provided by the increasing evidence of complex nuclear–mtDNA interactions that restrict the exchange of mtDNA between hybridizing American Mytilus taxa (Rawson *et al.* 1996; Saavedra *et al.* 1996), which contrasts clearly with the extensive levels of mtDNA introgression observed in European taxa (Quesada *et al.* 1996b, 1998). It is difficult to account for the compatibility between a single mtDNA genome and very different nuclear backgrounds in highly introgressed European populations, unless such a relaxation of selection is operating. Assumptions that need to be made in relation to variation in effective population size (N_e) and selective coefficient (s) to make this explanation consistent with a nearly neutral model seem by contrast to be more unrealistic because a general criticism of such models is that they require a very narrow range of s and N_e to fit data (Gillespie 1991, 1994). Alternative models and combinations of models, both selective and neutral, *e.g.*, Gillespie 1994, may better account for the data as a general explanation of similar excesses of mtDNA replacement polymorphism in taxa as divergent as Drosophila, humans, and mice.

Differential evolutionary rates of male and female mtDNA lineages: This study demonstrates a higher diversity for the M than the F genome, and that the M lineage evolves faster than the F lineage in European populations of *M. edulis* and *M. galloprovincialis.* These results are consistent with earlier reports in Mytilus using both RFLP (Skibinski *et al.* 1994b; Quesada *et al.* 1996, 1998) and sequence data (Rawson and Hilbish 1995; Stewart *et al.* 1995, 1996), and thus support the universality of this difference between genomes. What factors might be responsible? Our data suggest a combination of relaxed constraint and an increased mutation rate for the M genome.

In the earlier work of Stewart *et al.* (1996), a rate of replacement substitution significantly higher for the M than the F genome, but no significant difference for the synonymous rate, was reported. It was suggested that the higher rate for the M genome could thus only be the result of relaxed constraint for replacement substitutions; a role for differences in mutation rate was disregarded. In the present study, substantial differences in both the synonymous rate and, to a greater extent, the nonsynonymous rate are observed between genomes in American–European comparisons (Table 3). This result is clearly supportive of a higher mutation rate for the M genome, although it can also be explained under a nearly neutral model if the effective population size for the M genome is smaller, as suggested by Stewart *et al.* (1996). However, levels of diversity lower for the M than the F genome are expected under the assumption of a smaller effective population size for the M genome (Kimura 1983), but the opposite is observed (Table 2).

If similar levels of constraint were operating on both genomes, then differences in the synonymous but not in the nonsynonymous rate would reflect primarily differences in mutation rate. The striking and similar relaxation of constraints for both F and M genomes in European taxa, leading to no substantial differences of constraint between genomes, provides a test of this prediction. First, nucleotide diversity is significantly higher for the M than the F genome for synonymous substitutions, but not for nonsynonymous substitutions (Table 2). Second, the ratios of nonsynonymous to synonymous substitutions within the F and M lineages are very similar in European taxa (Table 3), but this ratio was always smaller for the M genome because of an increased synonymous rate. Third, no differences are observed between the F and M genomes in the number of sites showing replacement substitutions, but both *M. edulis* and Mediterranean *M. galloprovincialis* show a significant (Monte Carlo test, $P < 0.01$) increase in the number of sites with synonymous substitutions for the M genome (calculated from the data of Table 6).

Several alternative hypotheses fail to account for the observed differences in diversity and divergence between genomes. Genome-specific population bottlenecks in the recent history of the species could reduce variability of the F genome, but they should increase sequence divergence, as diversity becomes converted to divergence. However, this pattern is the opposite of that observed in the data. Hitchhiking on the F genome associated with a selective sweep could cause the observed reduction in diversity, as reported for Drosophila mtDNA (Ballard and Kreitman 1994; Rand *et al.* 1994). However, the hitchhiking effect predicts no particular change in the rate of substitution of unselected mutants (Birky and Walsh 1988; Kaplan *et al.* 1989) and, hence, cannot explain the lower divergence for the F genome. Another possibility is that hybridization and introgression between European Mytilus taxa (Gosling 1992; Quesada *et al.* 1995b) have caused higher polymorphism for the M than the F genome because of more relaxed taxonomic boundaries for the M genome. However, this should lead to lower sequence divergence among taxa for the M genome, the opposite of that observed.

In conclusion, the results presented here suggest that relaxed selection coupled with a higher mutation rate enhance each other's effect to generate a higher evolutionary rate for the M lineage. A higher level of oxidative damage in sperm mtDNA (see Skibinski *et al.* 1994b; Quesada *et al.* 1996) might cause the reported higher mutation rate for the M genome, as might an increase in the number of replications in the male germ line of mussels (Stewart *et al.* 1995; Rawson and Hilbish 1995). The increase in mutation rate for the M genome might be small, but the accumulated effect could be very dramatic on lineages that are evolving independently over a long period (Rawson and Hilbish 1995; Stewart *et al.* 1995; Quesada *et al.* 1996). In fact, the underlying differences among genomes in K_A and K_S are statistically significant only in comparisons including

the highly diverged *M. trossulsus* sequences. How general is the conclusion of a higher mutation rate for the M genome remains to be seen. It is possible that the contrary results of Stewart *et al.* (1996) could result from a difference in mutation rate stemming from different American and European thermal habits, *e.g.*, Rand 1994. However, it is also possible that divergence of genetic background after the isolation between American and European populations, or differences of statistical power between both studies, given the higher number of sequences analyzed here, could explain these contrasting results.

In summary, many factors are affecting the levels of mtDNA polymorphism in Mytilus, from rates of molecular evolution to population history. Depending on the particular interaction of these factors, it appears that levels of mtDNA polymorphism within and between species deviate from a neutral equilibrium model, and that natural selection is in part responsible for the overall patterns of nucleotide variation. The data presented here suggest that the balance between mutation and selection may be very subtle, and that different histories of adaptive and neutral evolution can lead to complex differences between populations. However, the excesses of replacement polymorphisms reported here in mussels are similar to those observed in species as divergent as mice, Drosophila, and humans, raising the possibility of a general mechanism governing mtDNA evolution.

We thank the Natural Environmental Research Council of the UK for financial support. We are grateful to A. Quesada and F. Rodríguez for help in the collection of mussels, and to K. A. Naish for introducing us to automatic sequencing. H.Q. was supported by a postdoctoral fellowship from the Ministerio de Educación y Ciencia (Spain). Sequences have been deposited in GenBank, accession numbers AF063251–AF063300.

LITERATURE CITED

Ballard, J. W. O., and M. Kreitman, 1994 Unravelling selection in the mitochondrial genome of *Drosophila*. Genetics **138:** 757–772.

Ballard, J. W. O., and M. Kreitman, 1995 Is mitochondrial DNA a strictly neutral marker? TREE **10:** 485–488.

Birky, C. W., Jr., and J. B. Walsh, 1988 Effects of linkage on rates of molecular evolution. Proc. Natl. Acad. Sci. USA **85:** 6414–6418.

Brookfield, J. F. Y., and P. M. Sharp, 1994 Neutralism and selectionism face up to DNA data. Trends Genet. **10:** 109–111.

Comeron, J. M., 1995 A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. J. Mol. Evol. **41:** 1152–1159.

Felsenstein, J., 1993 PHYLIP: phylogeny inference package. Version 3.1. University of Washington, Seattle.

Fisher, C., and D. O. F. Skibinski, 1990 Sex-biased mitochondrial DNA heteroplasmy in the marine mussel *Mytilus*. Proc. R. Soc. Lond. B **242:** 149–156.

Fu, Y. X., 1994 A phylogenetic estimator of effective population size or mutation rate. Genetics **136:** 685–692.

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

Gillespie, J. H., 1991 *The Causes of Molecular Evolution*. Oxford University Press, Oxford.

Gillespie, J. H., 1994 Substitution processes in molecular evolution. II. Exchangeable models from population genetics. Evolution **48:** 1101–1113.

Gosling, E. M., 1992 Systematics and geographical distribution of *Mytilus*, pp. 1–20 in *The Mussel Mytilus: Ecology, Physiology, Genetics and Culture*, edited by E. M. Gosling. Elsevier, Amsterdam.

Graur, D., and W. H. Li, 1991 Neutral mutation hypothesis test. Nature **354:** 114–115.

Higgins, D. G., J. D. Thompson and T. J. Gibson, 1996 Using Clustal for multiple sequence alignments. Methods Enzymol. **266:** 383–402.

Hoeh, W. R., K. H. Blakley and W. M. Brown, 1991 Heteroplasmy suggests limited biparental inheritance of *Mytilus mitochondrial* DNA. Science **251:** 1488–1498.

Hoeh, W. R., D. T. Stewart, B. W. Sutherland and E. Zouros, 1996 Cytochrome c oxidase sequence comparisons suggest an unusually high rate of mitochondrial DNA evolution in *Mytilus* (Mollusca: Bivalvia). Mol. Biol. Evol. **13:** 418–421.

Hoffmann, R. J., J. L. Boore and W. M. Brown, 1992 A novel mitochondrial genome organization for the blue mussel *Mytilus edulis*. Genetics **131:** 397–412.

Kaneko, M., Y. Satta, E. T. Matsura and S. Chigusa, 1993 Evolution of the mitochondrial ATPase 6 gene in *Drosophila*: unusually high level of polymorphism in *D. melanogaster*. Genet. Res. **61:** 195–204.

Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge.

Kumar, K., K. Tamura and M. Nei, 1993 *MEGA: Molecular Evolutionary Genetics Analysis*. Version 1.02. Pennsylvania State University, University Park, PA.

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

Nachman, M. W., S. N. Boyer and C. F. Aquadro, 1994 Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. Proc. Natl. Acad. Sci. USA **91:** 6364–6368.

Nachman, M. W., W. M. Brown, M. Stoneking and C. F. Aquadro, 1996 Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics **142:** 953–963.

Ohta, T., 1972 Population size and rate of evolution. J. Mol. Evol. **1:** 305–314.

Ohta, T., 1992 The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. **23:** 263–286.

Quesada, H., 1993 Estructura genética de poblaciones naturales del mejillón *Mytilus galloprovincialis* Lmk. Servicio de Publicaciones de la Universidad de Santiago de Compostela, Santiago de Compostela, Spain.

Quesada, H., C. M. Beynon and D. O. F. Skibinski, 1995a A mitochondrial DNA discontinuity in the mussel *Mytilus galloprovincialis* Lmk: pleistocene variance biogeography and secondary intergradation. Mol. Biol. Evol. **12:** 521–524.

Quesada, H., R. Wenne and D. O. F. Skibinski, 1995b Differential introgression of mitochondrial DNA across species boundaries within the marine mussel genus *Mytilus*. Proc. R. Soc. Lond. B **262:** 51–56.

Quesada, H., C. Zapata and G. Alvarez, 1995c A multilocus allozyme discontinuity in the mussel *Mytilus galloprovincialis*: the interaction of ecological and life-history factors. Mar. Ecol. Prog. Ser. **116:** 99–115.

Quesada, H., D. A. G. Skibinski and D. O. F. Skibinski, 1996 Sex-biased heteroplasmy and mitochondrial DNA inheritance in the mussel *Mytilus galloprovincialis* Lmk. Curr. Genet. **29:** 423–426.

Quesada, H., C. Gallagher, D. A. G. Skibinski and D. O. F. Skibinski, 1998 Patterns of polymorphism and gene flow of gender-associated mitochondrial DNA lineages in European mussel populations. Mol. Ecol. (in press).

Rand, D. M., 1994 Thermal habit, metabolic rate and the evolution of mitochondrial DNA. TREE **9:** 125–131.

Rand, D. M., M. Dorfsman and L. M. Kann, 1994 Neutral and nonneutral evolution of *Drospholia* mitochondrial DNA. Genetics **138:** 741–756.

Rand, D. M., and L. M. Kann, 1996 Excess amino acid polymorphism in mitochondrial DNA: contrast among genes from *Drosophila*, mice and humans. Mol. Biol. Evol. **13:** 735–748.

Rawson, P. D., and T. J. Hilbish, 1995 Evolutionary relationships among the male and female mitochondrial DNA lineages in the *Mytilus edulis* species complex. Mol. Biol. Evol. **12:** 893–901.

Rawson, P. D., C. L. Secor and T. J. Hilbish, 1996 The effects of

natural hybridization on the regulation of doubly uniparental mtDNA inheritance in blue mussels (*Mytilus* spp.). Genetics **144:** 241–248.

Roff, D. A. and P. Bentzen, 1989 The statistical analysis of mitochondrial DNA polymorphisms: X^2 and the problem of small samples. Mol. Biol. Evol. **6:** 539–545.

Rozas, J. and R. Rozas, 1997 DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. Comput. Appl. Biosci. **13:** 307–311.

Saavedra, C., D. T. Stewart, R. R. Stanwood and E. Zouros, 1996 Species-specific segregation of gender-associated mitochondrial DNA types in an area where two mussel species (*Mytilus edulis* and *Mytilus trossulus*) hybridize. Genetics **143:** 1359–1367.

Skibinski, D. O. F., C. Gallagher and C. M. Beynon, 1994a Mitochondrial DNA inheritance. Nature **368:** 817–818.

Skibinski, D. O. F., C. Gallagher and C. M. Beynon, 1994b Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis.* Genetics **138:** 801–809.

Stewart, D. T., C. Saavedra, R. R. Stanwood, A. O. Ball and E. Zouros, 1995 Male and female mitochondrial DNA lineages in the blue mussel (*Mytilus edulis*) species group. Mol. Biol. Evol. **12:** 735–747.

Stewart, D. T., E. R. Kenchington, R. K. Singh and E. Zouros, 1996 Degree of selective constraint as an explanation of the different rates of evolution of gender-specific mitochondrial DNA lineages in the mussel *Mytilus.* Genetics **143:** 1349–1357.

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

Whittam, T. S. and M. Nei, 1991 Neutral mutation hypothesis test. Nature **354:** 115–116.

Zouros, E., K. R. Freeman, A. O. Ball and G. H. Pogson, 1992 Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel *Mytilus.* Nature **359:** 412–414.

Zouros, E., A. O. Ball, C. Saavedra and K. R. Freeman, 1994a Mitochondrial DNA inheritance. Nature **368:** 817–818.

Zouros, E., A. O. Ball, C. Saavedra and K. R. Freeman, 1994b An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus.* Proc. Natl. Acad. Sci. USA **91:** 7463–7467.

Communicating editor: A. G. Clark

APPENDIX 1

Variable nucleotides from F and M *COIII* mtDNA in three Mytilus taxa

	1 4 5 6 7 9	1 1 2	1 1 3	1 1 5	1 2 0	2 2 1	2 2 4	2 2 5	2 3 6	3 3 0	3 3 3	3 3 6	3 3 8	3 4 9	4 4 2	4 4 5	4 5 8	5 5 1	5 5 2	5 5 5	5 5 7	6 6 8	6 6 0	6 6 7	7 7 2	7 7 5	7 6
Consen.	G C A T A G	G G	A	G	T T	G	G	G	A	C	C	A	G	G	A	C	C	C	T	G	G	G	G	C	G	C	G	A
ED.27M	. . . . .	. .	.	.	. .	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
ED.13M	A	. .	.	.	C .	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.
ED.6M	. . . . .	. .	.	.	. .	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.5M	. . . . .	. .	.	.	. .	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.14R	. . . . .	. .	.	.	. .	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.8R	. . . . .	. .	.	.	. .	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
ED.5R	. . . . .	. .	.	.	. .	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
ED.3R	. . . . .	. .	.	.	. .	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.
ED.2R	. . . . .	. .	.	.	. .	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
AG.64V	. . . . .	T .	.	.	. .	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.60V	. . . . .	. .	.	.	C .	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.58V	. . . . .	. .	.	.	. .	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.57V	. . . . .	. .	.	.	. .	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.53V	. . . . .	. .	.	.	. .	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
AG.52V	. . . . .	. .	.	.	. .	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.23C	. A	. .	.	.	. .	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
MG.11C	. . . . .	C .	.	.	. .	.	C	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
MG.2C	. . . . .	. .	.	.	. .	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
MG.57P	. . . . .	. .	T	.	. .	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
MG.56P	A	. .	T	.	. .	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
MG.55P	A	. .	?	.	. C	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.54P	. . . . .	. .	?	.	. C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
MG.53P	. . . . .	. .	.	.	. C	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
MG.52P	. . . . .	. .	.	.	C .	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
ED.30M	. . . . .	A .	A	.	. .	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	C	G
ED.26M	. . . . .	A .	A	.	. .	A	A	.	.	T	T	G	.	C	C	T	T	T	.	T	.	A	A	T	C	T	.	G
ED.25M	. . . . .	A .	A	.	. .	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	C	G
ED.21M	. . . . .	A .	A	.	. .	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	C	G
ED.25R	. . . . .	A .	A	G	. C	.	A	A	.	.	T	.	.	.	C	C	T	T	T	.	T	.	A	A	T	T	T	T	G
ED.23R	. . . . .	A .	A	.	. C	.	A	A	.	.	T	T	G	.	C	C	T	T	.	.	T	.	T	.	T	T	T	C	G
ED.20R	. . . . .	A .	A	.	. C	.	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	C	G
ED.13R	. . . . .	A .	A	.	. C	.	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	C	G
ED.10R	. . . . .	A .	A	.	. .	.	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	C	G
ED.4R	. . . . .	A .	A	.	. .	.	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	C	G
AG.28V	. . . . .	A .	A	.	. .	.	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	.	T	T	C	G
AG.25V	. . . . .	A .	A	.	. C	.	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	C	G
AG.24V	. . . . .	A .	A	.	. C	.	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	T	G
AG.20V	. . T . G	. .	T	G	. C	.	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	T	G
AG.18V	. . . . .	A .	A	.	. .	.	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	C	G
AG.16V	. . . . .	A .	A	.	. C	.	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	T	A	A	T	T	T	T	G
MG.29C	. . . . .	A .	A	.	. C	.	A	A	.	.	T	T	G	.	C	C	T	T	T	.	T	.	A	A	T	T	T	T	G
MG.19C	. . . . .	A .	A	.	. C	.	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	T	G
MG.18C	. . . . .	A .	A	.	. C	.	A	A	.	.	T	T	G	.	C	C	T	T	T	.	T	.	A	A	T	T	T	T	G
MG.6C	. . . C .	A .	A	G	A C	.	A	A	.	.	T	T	C	.	C	T	T	T	T	.	C	.	A	A	.	T	T	A	G
MG.26P	. . . C .	A .	A	G	A C	.	A	A	.	.	T	T	C	.	C	C	T	T	T	.	C	.	A	A	T	T	T	A	G
MG.24P	. . . . .	A .	A	G	A C	.	A	A	.	.	T	T	C	.	C	T	T	T	T	.	C	.	A	A	T	T	T	A	G
MG.22P	. . . . .	A .	A	.	. .	.	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	C	G
MG.21P	. . . . .	A .	A	.	. .	.	A	A	.	.	T	T	G	.	T	C	T	T	T	.	T	.	A	A	T	T	T	C	G
MG.15P	. . . . .	A .	A	.	. C	.	A	A	.	.	T	T	G	.	C	C	T	T	T	.	T	.	A	A	T	T	T	C	G
MG.14P	C	A A	A	.	. C	.	A	A	A	.	T	T	G	C	C	C	T	T	T	.	T	A	A	A	T	T	T	.	.
Type of change	R R R S R	S R	S	R	S S	R	S	R	R	R S	S	S	R	S	S	S	S	S	S	R	R	R	S	S	R	S	S	R
								S												S_M	S				S_M			

APPENDIX 1

Continued

	7 8	7 9	8 0	8 1	8 2	8 4	8 7	8 9	9 0	9 3	9 7	9 9	1 0 1	1 0 2	1 0 4	1 0 5	1 0 8	1 1 1	1 1 4	1 1 5	1 1 7	1 2 0	1 2 1	1 2 2	1 2 3	1 2 4	1 2 5	1 2 6	1 2 9	1 3 0	1 3 2	1 3 5	1 3 8	1 4 1	1 4 4
Consen.	T	G	T	G	T	T	C	T	G	A	C	G	G	G	A	A	T	T	A	A	C	C	T	A	T	A	C	T	T	G	A	T	G	T	T
ED.27M	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.13M	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.6M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.5M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.14R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.8R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.
ED.5R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.3R	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	.	.	.	.	.	.	C	T	A	T	.	.	.	.	.	.	.	.	.
ED.2R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.64V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
AG.60V	.	C	G	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.58V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.57V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.53V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
AG.52V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.23C	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
MG.11C	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	C	.	C	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
MG.2C	C	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	C	C	.	.
MG.57P	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
MG.56P	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
MG.55P	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.54P	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C
MG.53P	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
MG.52P	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
ED.30M	.	.	.	A	.	C	T	.	A	.	G	A	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	C	.
ED.26M	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
ED.25M	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
ED.21M	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
ED.25R	.	.	.	A	.	C	T	C	A	.	G	.	.	A	.	G	C	.	.	.	.	T	.	.	.	.	G	G	.	.	.	.	T	.	.
ED.23R	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
ED.20R	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
ED.13R	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	C	.	G	G	.	.	.	C	T	.	.
ED.10R	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	.	T	.	.
ED.4R	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	.	.	.	.	C	T	.	.
AG.28V	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
AG.25V	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	T	.	C	T	.	.
AG.24V	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
AG.20V	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
AG.18V	.	.	.	A	.	C	T	.	A	.	T	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
AG.16V	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
MG.29C	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
MG.19C	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
MG.18C	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
MG.6C	.	.	.	A	.	C	T	.	A	G	G	.	.	A	.	G	C	.	G	.	.	.	.	.	C	.	G	G	.	.	.	C	T	.	.
MG.26P	.	.	.	A	.	C	T	.	A	G	G	.	.	A	.	G	C	.	G	.	.	.	.	.	C	.	G	G	C	.	.	C	T	.	.
MG.24P	.	.	.	A	.	C	T	.	A	G	G	.	.	A	.	G	.	C	G	.	.	.	.	.	C	.	G	G	.	.	.	C	T	.	.
MG.22P	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
MG.21P	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
MG.15P	.	.	.	A	.	C	T	.	A	.	G	.	.	A	.	G	.	.	.	.	.	T	.	.	.	.	G	G	.	.	.	C	T	.	.
MG.14P	.	.	.	A	.	C	T	.	A	.	.	C	.	A	.	G	C	.	.	.	.	T	.	.	.	.	G	.	.	.	.	C	T	.	.
Type of change	S	R	R	S	R	S	S	R	S	S	R S	S	R	S	R	S	S	S	S	R	S	S	R	R	R S	R	R	S	R S	R	S	S	S	S	S

APPENDIX 1

Continued

	1 4 7	1 5 0	1 5 3	1 5 5	1 5 6	1 5 9	1 6 3	1 6 8	1 7 1	1 7 4	1 7 6	1 7 7	1 8 0	1 8 1	1 8 2	1 8 3	1 8 6	1 8 9	1 9 2	1 9 8	1 9 9	2 0 0	2 0 1	2 0 2	2 0 4	2 0 8	2 1 2	2 1 3	2 1 6	2 1 9	2 2 0	2 2 2	2 2 3	2 2 4	2 2 5
Consen.	T	G	T	T	T	C	T	T	A	C	A	C	G	A	T	A	T	T	C	G	A	C	T	A	T	C	T	G	A	G	T	A	G	T	T
ED.27M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.13M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.6M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.5M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.14R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.8R	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.5R	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.3R	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.2R	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.64V	.	A	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.60V	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.58V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.57V	.	.	.	.	C	.	.	.	.	.	C	.	.	.	.	.	.	.	.	A	C	T	A	C	.	.	.	.	.	.	.	.	.	.	.	.
AG.53V	C	.	.	.	C	T	C	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.52V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.23C	.	.	.	.	C	T	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.11C	.	.	.	.	C	T	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.2C	C	.	.	C	C	T	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.57P	.	.	.	.	C	T	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.56P	C	.	.	.	C	T	C	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.55P	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.54P	C	.	.	.	.	T	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.53P	C	.	.	.	C	T	C	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.52P	C	.	.	.	C	T	C	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.30M	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	.	T	.	A	.	T	C	G	.	C	C
ED.26M	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
ED.25M	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	.	T	.	A	.	T	C	G	.	C	.
ED.21M	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
ED.25R	.	A	.	.	C	T	.	.	T	T	.	T	.	G	C	.	.	A	A	.	.	.	C	T	C	T	.	A	.	C	C	G	.	C	.
ED.23R	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
ED.20R	C	A	C	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
ED.13R	C	A	C	.	.	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
ED.10R	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
ED.4R	C	A	.	.	C	T	.	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
AG.28V	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	.	G	.	C	.
AG.25V	C	A	.	.	C	T	C	.	G	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	.	T	.	A	.	T	C	G	.	C	.
AG.24V	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	G	C	.	.	A	.	T	C	G	.	.	.
AG.20V	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	A	G	.	C	.
AG.18V	.	A	.	.	C	T	.	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
AG.16V	C	A	C	.	C	T	.	.	T	.	.	T	.	G	G	G	.	G	A	.	.	.	.	G	C	T	.	A	.	T	C	G	.	C	.
MG.29C	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
MG.19C	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
MG.18C	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
MG.6C	.	A	C	.	.	T	.	C	G	T	.	T	.	G	C	.	.	G	.	.	.	.	.	.	T	C	A	G	T	.	.	.	.	.	.
MG.26P	.	A	.	.	.	T	.	C	G	.	.	T	A	G	C	.	.	G	.	.	.	.	.	.	T	G	A	G	T	.	.	A	G	.	
MG.24P	.	A	.	.	.	T	.	C	G	.	.	T	A	G	C	.	.	G	.	.	.	.	.	.	T	C	A	G	T	.	.	.	.	.	.
MG.22P	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
MG.21P	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
MG.15P	C	A	.	.	C	T	C	.	C	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
MG.14P	C	A	.	.	C	T	C	.	T	.	.	T	.	G	C	G	.	G	A	.	.	.	.	.	C	T	.	A	.	T	C	G	.	C	.
Type of change	S	S	S	R	S	S	S	S	S	S	R	S	S	R	R	S	S	S	S	S	R	R	S	R	S	S	R	S	S	S	S R	S	R	R	S

APPENDIX 1

Continued

	2 2 8	2 2 9	2 3 4	2 3 7	2 4 0	2 4 6	2 5 2	2 5 5	2 5 8	2 6 1	2 6 4	2 6 7	2 7 6	2 7 9	2 8 0	2 8 6	2 8 8	2 9 4	2 9 7	2 9 8	3 0 0	3 0 1	3 0 2	3 0 3	3 0 6	3 0 9	3 1 2	3 1 6	3 1 7	3 1 8	3 1 9	3 2 4	3 2 5	3 2 6	3 2 7
Consen.	A	C	A	C	G	T	T	A	A	C	C	G	T	T	A	T	C	T	T	G	A	G	A	T	G	A	A	G	C	A	T	G	T	G	C
ED.27M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.13M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.6M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.5M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.14R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.8R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.5R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.3R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.2R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.64V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.60V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.58V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.57V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.53V	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.
AG.52V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.23C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.11C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.2C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.57P	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
MG.56P	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.55P	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.54P	.	.	.	.	.	.	A	.	.	T	.	.	.	.	.	T	.	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
MG.53P	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.52P	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.30M	C	T	G	T	A	C	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
ED.26M	C	T	G	T	A	.	C	.	C	T	T	T	.	C	.	.	T	.	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
ED.25M	C	T	G	T	A	C	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
ED.21M	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
ED.25R	C	T	G	T	A	C	C	.	C	T	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
ED.23R	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
ED.20R	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	C	A	.	.	A	G	G	C	.	.	T	T
ED.13R	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
ED.10R	C	T	G	T	A	.	C	.	T	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
ED.4R	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	G	A	G	G	C	A	.	T	T
AG.28V	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
AG.25V	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
AG.24V	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
AG.20V	C	T	G	T	A	.	C	.	.	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
AG.18V	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
AG.16V	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	C	A	.	.	A	G	G	C	A	.	T	T
MG.29C	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
MG.19C	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
MG.18C	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	A	.	.	A	.	.	A	G	G	C	A	.	T	T
MG.6C	C	T	G	T	A	.	C	.	C	T	T	.	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	.	G	.	A	C	T	T
MG.26P	C	T	G	T	A	.	C	.	C	T	T	.	C	C	.	.	T	C	.	.	G	.	G	.	A	.	.	A	.	G	.	A	C	T	T
MG.24P	C	T	G	T	A	.	C	.	C	T	T	.	.	C	.	.	T	C	.	.	G	A	.	.	A	.	.	A	.	G	.	A	C	T	T
MG.22P	C	T	G	T	A	.	C	C	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	.	C	A	.	T	T
MG.21P	C	T	G	T	A	.	C	C	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	.	C	A	.	T	T
MG.15P	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	G	.	.	.	A	.	.	A	G	G	C	A	.	T	T
MG.14P	C	T	G	T	A	.	C	.	C	.	T	T	.	C	.	.	T	C	.	.	.	.	.	.	A	.	.	A	G	G	C	A	.	T	T
Type of change	S	S	S	S	S	S	S	R	S	S	S	S	S	S	R	R	S	S	S	R	S	R	R	S	S	S	S	R	R	S	S	S	R	R	S

APPENDIX 1

Continued

	3 2 8	3 2 9	3 3 0	3 3 3	3 3 5	3 3 6	3 3 7	3 3 9	3 4 8	3 5 1	3 5 2	3 5 5	3 5 7	3 6 0	3 6 3	3 6 4	3 7 2	3 7 5	3 7 7	3 7 8	3 8 0	3 8 1	3 8 4	3 8 7	3 8 8	3 9 0	3 9 1	3 9 2	3 9 3	3 9 6
Consen.	T	T	A	G	A	T	G	G	A	A	T	T	A	T	G	T	T	G	T	A	G	G	T	G	G	T	G	T	T	C
ED.27M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.13M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.6M	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.
ED.5M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.14R	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.8R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.5R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.3R	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ED.2R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.64V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.60V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.58V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.57V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.53V	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AG.52V	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.23C	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
MG.11C	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	T
MG.2C	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	T
MG.57P	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	T
MG.56P	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.55P	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MG.54P	C	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	C	.	.
MG.53P	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
MG.52P	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
ED.30M	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	.	.	.	A	C	T	.	C	A	.	.	T
ED.26M	G	.	G	.	.	C	.	.	A	G	G	.	.	.	C	.	.	.	A	A	.	.	A	C	T	.	.	A	.	.	T
ED.25M	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	.	.	.	A	C	T	.	.	A	.	.	T
ED.21M	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	.	.	.	A	C	T	.	C	A	.	.	T
ED.25R	G	.	G	A	.	.	.	.	A	G	G	.	.	.	C	.	.	C	A	.	.	.	A	.	T	.	C	A	.	.	.
ED.23R	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	.	.	.	A	C	T	.	C	A	.	.	T
ED.20R	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	A	.	.	A	C	T	.	C	A	.	.	.
ED.13R	G	.	G	.	.	.	C	A	G	G	.	.	.	C	.	.	.	A	.	T	.	A	C	T	.	C	A	.	.	T
ED.10R	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	.	.	.	A	C	T	.	C	A	.	.	T
ED.4R	G	A	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	A	.	.	A	C	T	.	.	A	.	.	T
AG.28V	G	.	G	.	.	.	.	.	A	G	G	.	.	.	.	.	.	.	A	.	.	.	A	C	T	.	C	A	.	.	-
AG.25V	G	.	.	.	.	.	.	.	A	G	G	.	C	.	C	.	G	.	A	.	.	.	A	C	T	.	C	A	.	.	T
AG.24V	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	.	.	.	A	C	T	.	C	A	.	.	T
AG.20V	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	.	.	.	A	C	T	.	C	A	.	.	T
AG.18V	G	.	G	.	.	.	.	.	A	G	T	.	.	.	C	.	.	.	A	.	.	.	A	C	T	.	C	A	.	.	T
AG.16V	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	.	.	.	A	C	T	.	C	A	.	.	.
MG.29C	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	.	.	.	A	C	T	.	C	A	.	.	T
MG.19C	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	A	.	.	A	C	T	.	C	A	.	.	T
MG.18C	G	A	T	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	.	.	.	A	C	T	.	C	A	.	.	T
MG.6C	G	.	G	.	.	C	.	A	.	G	.	.	G	.	A	.	.	A	.	.	.	.	A	C	T	.	C	A	.	.	.
MG.26P	G	G	G	A	.	C	.	A	.	G	.	.	G	C	A	.	.	A	.	T	.	A	C	T	.	.	A	.	.	.
MG.24P	G	.	G	A	.	C	.	A	.	T	A	.	T	C	A	.	.	A	.	.	.	A	C	T	.	-	A	.	.	-
MG.22P	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	.	.	.	A	C	T	.	-	A	.	.	-
MG.21P	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	.	T	.	A	C	T	.	C	A	.	.	T
MG.15P	G	.	G	.	.	.	.	.	A	G	G	.	C	.	C	.	.	.	A	.	.	.	A	C	T	.	.	A	.	.	T
MG.14P	G	.	G	.	.	.	.	.	A	G	G	.	.	.	C	.	.	.	A	.	.	.	A	C	T	.	C	A	.	.	T
Type of change	S	R	S	S	R	S	R	S	S	S	R	S	S	S	S	R	S	S	R	R	R	S	S	S	R	S	R	R	S	S
	R												R																	

R, replacement; S, synonymous; S_M, synonymous within M lineage.
F sequences appear in regular type; M sequences appear in bold type.