The Role of the Effective Population Size in Compensatory Evolution

Robert Piskol*,† and Wolfgang Stephan

Section of Evolutionary Biology, Ludwig-Maximilian University, Munich, Germany

†Present address: Department of Genetics, Stanford University, Stanford, California

*Corresponding author: E-mail: piskol@stanford.edu.

Accepted: 4 June 2011

Abstract

The impact of the effective population size (N_e) on the efficacy of selection has been the focus of many theoretical and empirical studies over the recent years. Yet, the effect of N_e on evolution under epistatic fitness interactions is not well understood. In this study, we compare selective constraints at independently evolving (unpaired) and coevolving (paired) sites in orthologous transfer RNAs (tRNA molecules for vertebrate and drosophilid species pairs of different N_e. We show that patterns of nucleotide variation for the two classes of sites are explained well by Kimura's one- and two-locus models of sequence evolution under mutational pressure. We find that constraints in orthologous tRNAs increase with increasing N_e of the investigated species pair. Thereby, the effect of N_e on the efficacy of selection is stronger at unpaired sites than at paired sites. Furthermore, we identify a "core" set of tRNAs with high structural similarity to tRNAs from all major kingdoms of life and a "peripheral" set with lower similarity. We observe that tRNAs in the former set are subject to higher constraints and less prone to the effect of N_e, whereas constraints in tRNAs of the latter set show a large influence of N_e. Finally, we are able to demonstrate that constraints are relaxed in X-linked drosophilid tRNAs compared with autosomal tRNAs and suggest that N_e is responsible for this difference. The observed effects of N_e are consistent with the hypothesis that evolution of most tRNAs is governed by slightly to moderately deleterious mutations (i.e., $|N_e s| \leq 5$).

Key words: effective population size, tRNA, compensatory evolution, selective constraints.

Introduction

The effective population size (N_e) is a fundamental quantity in population genetics. It is essential in shaping neutral nucleotide variation in a population and crucial for determining the efficacy of selection (Kimura 1983; Charlesworth 2009). The rate of molecular evolution may decrease, remain unchanged, or increase with increasing N_e, depending on whether mutations are deleterious, (nearly) neutral, or beneficial in nature, respectively (Gillespie 1999). For independently evolving sites, the rate depends on the product of N_e and the selection coefficient s as well as the scaled mutation rate ($\theta = 4N_e\mu$). Therefore, a mutation that is slightly deleterious in a species of large N_e might have a neutral effect in a species with small N_e (Chamary et al. 2006). This role of N_e in the evolution of independently evolving sites has been studied extensively from a theoretical point of view (Kimura and Ohta 1969; Ohta 1972; Kimura 1983) and has been empirically confirmed (Weinreich and Rand 2000; Woolfit and Bromham 2003, 2005; Előry et al. 2010; Andolfatto et al. 2011). However, the relation between the speed of evolution due to N_e at independent nucleotide sites and positions that evolve under epistasis is much less clear.

To study the evolution of sites that are involved in epistatic interactions, a model with at least two loci is needed. Kimura (1985) introduced a two-locus model of compensatory neutral mutations in molecular evolution (fig. 1*a*). He assumed that mutations at a pair of loci may be individually deleterious but neutral in certain combinations. Given two loci with wild-type alleles *A* and *B* at the first and second locus, respectively, he studied the expected fixation time ($\bar{T}_{coev}$) for the double-mutant *ab* under the assumption that selection against individual mutants is strong (and thus the mutation process is nearly irreversible). Specifically, he assumed that the intermediate configurations of alleles (*Ab*, *aB*) suffer the same disadvantage *s* and that the wild-type *AB* and double-mutant *ab* have the same fitness (i.e., the process does not lead to adaptation but only compensation). Under such conditions, *ab* may rise to fixation without

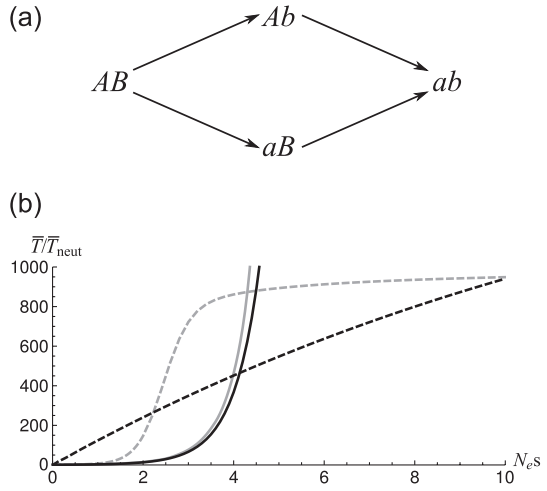

Fig. 1—(*a*) Kimura's (1985) two-locus model of sequence evolution, which assumes unidirectional mutation. The model is described in the main text. (*b*) Expected ratio of waiting times until fixation of deleterious ($\bar{T}$) and selectively neutral mutations ($\bar{T}_{neut}$) at independently evolving (solid lines) and coevolving sites (dashed lines). Black lines describe fixation times in Kimura's unidirectional models (eq. 13 from Kimura 1980 and eq. 16 from Kimura 1985). Gray lines were obtained by taking back mutations into account using equations (5a and 6) in Innan and Stephan (2001) for coevolving sites and simulations of the Wright–Fisher process for independent sites. Results are given for mutation rate $\mu = 2.5 \times 10^{-8}$ and selection coefficient $s = 10^{-4}$.

prior fixation of any of the deleterious intermediates (stochastic tunneling) (Iwasa et al. 2004). Subsequently, Kimura's model was extended by incorporating different reductions in fitness (s_1, s_2) for the intermediates *Ab*, *aB* (Stephan 1996) and also allowing for weak purifying selection such that back mutations may be possible (Innan and Stephan 2001). In this case, fixation of *ab* can be preceded by a fixation of any of the deleterious intermediates (Ohta 1973). Fixation times in the two-locus case were also investigated in diploid populations (Ichinose et al. 2008) and for double mutations that lead to adaptation (Lynch 2010; Weissman et al. 2010). All these models have in common that for most parameter combinations, $\bar{T}_{coev}$ was found to increase with increasing $N_e s$. Furthermore, for weak selection, faster fixation at independently evolving sites is expected, whereas it was shown that in the case of strong selection against deleterious intermediates, evolution proceeds faster at coevolving sites (fig. 1*b*) (Kimura 1985).

The role of compensatory mutations has been investigated in the case of protein evolution (Brown et al. 2010), but also RNA molecules provide a great opportunity to directly compare evolution at independently evolving and coevolving sites as they are composed of unpaired nucleotides and nucleotides that form Watson–Crick (WC) base pairs. Previous studies have shown that compensatory mutations are the main driving force of evolution in paired regions of RNA molecules (Parsch et al. 1997; Chen et al. 1999; Chen and Stephan 2003; Meer et al. 2010) and that the rate of compensation depends on structural features of the molecule. Specifically, this rate can be related to the length of the pairing region (helix), the position of the pairing nucleotide within the helix, and the GC content of the helix (Parsch et al. 2000; Piskol and Stephan 2008). Furthermore, population genetic parameters such as the recombination rate between pairing sites were shown to influence the rate of coevolution in RNA molecules (Kirby et al. 1995). Here we investigate how another population genetic parameter, N_e, shapes RNA evolution. We are especially interested how it influences the rate of evolution at independently evolving and coevolving sites. Therefore, we focus on transfer RNAs (tRNAs)—a class of noncoding RNAs with well-studied structure and function. We present a rigorous analysis of selective constraints in tRNA molecules with particular focus on the difference between selective constraints for paired and unpaired nucleotides and interpret the results in the light of theoretical predictions for fixation times of deleterious mutations. We use Kimura's (1980, 1985) unidirectional models to describe sequence evolution at independently evolving and coevolving sites (fig. 1*a*). In our analysis, the range of moderate and weak purifying selection ($|N_e s| \leq 5$) is of particular interest as evolution of paired sites in noncoding RNA molecules was shown to take place in this parameter range (Piskol and Stephan 2011).

Materials and Methods

Sequence Data

Sequence data were obtained from the University of California Santa Cruz (UCSC) Genome Browser FTP server (Kent et al. 2002) in form of axt pairwise alignments for the following vertebrate species pairs: human/macaque (hg19/rheMac2), macaque/marmoset (rheMac2/calJac3), dog/cat (canFam2/felCat3), and chicken/zebra finch (galGal3/taeGut1). The assemblies of these genomes are the same as used by Rfam (Gardner et al. 2009) for the annotation of noncoding RNA families. The pairwise genomic alignment of mouse/rat available at UCSC is based on different assemblies than the Celera assemblies (Mural et al. 2002) used by Rfam. Therefore, the Celera assemblies of the mouse and rat genomes were aligned following the same protocol that was used to produce the UCSC alignments. The vertebrate alignments served as a source for orthologous tRNAs and neutrally evolving sequences. Annotations of tRNAs were downloaded from Rfam (Release 10.0) for human, macaque, mouse, rat, dog, and chicken. The UCSC Drosophila multiple alignment, which consists of up to 12 species, was analyzed for the species pairs *Drosophila melanogaster/D. simulans* and *D. melanogaster/D. yakuba*. It was used to determine neutrally evolving regions only. The annotations of orthologous Drosophila tRNAs were taken from Rogers et al.

Table 1
Composition of tRNA Data Sets for Different Species Pairs

Species Pair	N_e	all tRNAs			Peripheral tRNAs			Core tRNAs		
		# tRNAs	GC Content		# tRNAs	GC Content		# tRNAs	GC Content	
			Paired	Unpaired		Paired	Unpaired		Paired	Unpaired
Human/macaque	8.9×10^4	277 (2)	0.5138	0.3105	151 (2)	0.4976	0.3142	126 (0)	0.5316	0.3059
Macaque/marmoset	1.7×10^5	268 (1)	0.5172	0.3165	144 (1)	0.4915	0.3173	124 (0)	0.5441	0.3156
Dog/cat	5.2×10^5	259 (0)	0.5256	0.3080	134 (0)	0.5206	0.3124	125 (0)	0.5298	0.3033
Chicken/zebra finch	6.5×10^5	114 (1)	0.7029	0.4123	63 (1)	0.7149	0.4169	51 (0)	0.6884	0.4062
Mouse/rat	$\approx 10^6$	106 (0)	0.5552	0.3074	46 (0)	0.5850	0.3271	60 (0)	0.5356	0.2920
Drosophila melanogaster/D. yakuba	$>10^6$	277 (21)	0.6963	0.3827	95 (5)	0.6788	0.4019	182 (16)	0.7061	0.3720
D. melanogaster/D. simulans	$>10^6$	229 (13)	0.6956	0.3822	83 (2)	0.6770	0.4025	146 (11)	0.7071	0.3700

NOTE.—Numbers of X-linked tRNAs are shown in parentheses. The GC content is given for non–CpG-prone positions only.

(2010), and corresponding sequences were downloaded in batch from Flybase (Tweedie et al. 2009). Annotations of protein coding genes were acquired from the refGene tracks of the UCSC Genome Browser for all species except mouse and rat. The locations of ancestral repeats (ARs), that is, repetitive sequences common to both species in a pair were determined according to RepeatMasker annotations available in the "rmsk" tables of the UCSC Genome Browser (downloaded on 18 December 2010). Protein coding gene annotations in mouse and rat were obtained from GenBank, and repeats in the Celera mouse and rat assemblies were annotated using RepeatMasker 3.2.9 (Smit et al. 1996–2010) based on mouse/rat-specific repeat libraries RM-20090604 (Jurka et al. 2005).

Effective Population Sizes

Estimates of long-term effective population sizes were obtained from the literature (table 1) for chicken/zebra finch (Jennings and Edwards 2005), mouse/rat (Baines and Harr 2007), and Drosophila (Li and Stephan 2006). N_e for macaque/marmoset and dog/cat were taken from Piganeau and Eyre-Walker (2009), assuming that the ratio $N_{e-autosomes}$:$N_{e-mitochondria}$ is 4:1. In most of these studies, long-term N_e for the pairs were calculated as averages of single-species N_e, which were obtained from polymorphism data. Because no estimate of N_e existed for the pair human/macaque, we averaged over N_e for the two species (Eyre-Walker et al. 2002; Evans et al. 2010). However, due to the heterogeneity of the data sources employed for the calculation of N_e, the absolute values were not directly used in our analysis. Estimates of N_e merely served to establish the following semiquantitative relationship between species pairs: N_e (human/macaque) $<$ N_e (macaque/marmoset) $<$ N_e (dog/cat) $<$ N_e (chicken/zebra finch) $<$ N_e (mouse/rat) $<$ N_e (*D. melanogaster/D. yakuba*) $\approx$ N_e (*D. melanogaster/D. simulans*).

tRNA Alignments and Structures

Orthologous vertebrate tRNA sequences and structures for all species pairs were determined based on the pairwise species alignments and Rfam annotations. Thereby, if tRNA Rfam annotations existed for both species in a pair, overlapping orthologs were identified and the corresponding sequences extracted from the pairwise alignment. If Rfam annotations existed only for the reference species, then sequences of the other species ("query" species) that were aligned to the reference in the annotated regions were evaluated against the Rfam tRNA covariance model using cmsearch from the INFERNAL package (version 1.0.2) (Nawrocki et al. 2009) to obtain a bit score S and corresponding E value for each sequence. The score S indicates how well a given sequence matches the Rfam tRNA covariance model, which describes a consensus tRNA structure that is based on a seed alignment of tRNAs from 967 species. These cover all major kingdoms of life (bacteria, fungi, plants, and animals). The score represents the $\log_2$ odds ratio between the probability of the target sequence under the covariance model and its probability to be a random sequence. For instance, a bit score of 35 symbolizes a 2^{35} higher probability of the target sequence to be a tRNA, compared with a random sequence. Higher S values indicate larger similarity of the target sequence to the covariance model and therefore higher structural similarity to a core set of tRNAs common to all species, whereas low E values describe a small probability that the sequence occurred only by chance in a database of random sequences. Therefore, selecting for higher bit scores allows us to choose tRNAs that share a high similarity with tRNAs from a wide range of other species. For all pairwise vertebrate and Drosophila alignments, only tRNA annotations with an INFERNAL bit score of $S > 35$ in both species and an E value < 0.01 were retained for further analysis. Furthermore, we discarded cases where the query sequence aligned to more than one location in the reference genome and only considered cases where both aligned tRNA annotations were located either on the X chromosome or on the autosomes in the two species. Subsequently, each pair of orthologous sequences was realigned using cmalign (Nawrocki et al. 2009). To rule out the influence of alignment and structure prediction on observed selective constraints and to avoid

problems with the alignment of unpaired regions, we also created alternative alignments using mlocarna (Will et al. 2007) for a structure-based alignment and a combination of muscle (Edgar 2004) and RNAalifold (Hofacker et al. 2002) where alignment and structure are determined separately from each other. Both mlocarna and RNAalifold rely on thermodynamic predictions of the secondary structure. In some cases, thermodynamic prediction may fail to determine the correct topology of tRNA molecules. Therefore, we informed mlocarna and RNAalifold by providing the cmalign structures as constraints for either both sequences or the reference sequence, respectively. Orthologous drosophilid tRNAs from Rogers et al. (2010) were scored with cmsearch and subsequently aligned using the same three methods as for vertebrate tRNAs. Here, we present results based on the mlocarna alignments. Estimates of selective constraints obtained with muscle and cmalign are shown in supplementary tables S2 and S3 (Supplementary Material online), respectively. They only differ quantitatively, whereas qualitative predictions are the same for all three methods.

Neutrally Evolving Sequences

ARs, that is, repetitive sequences common to both species in a pair, served as indicators for neutral evolution in vertebrates (Előry et al. 2010). Only ARs that reside in intergenic locations were considered. ARs were excluded if the pairwise alignment contained less than 50% of aligned nucleotides. Similar to previous studies (Előry et al. 2010; Piskol and Stephan 2011), only long terminal repeats, DNA transposons, short interspersed elements, long interspersed elements, and other repeats were considered, whereas simple repeats, low complexity regions, and microsatellites were excluded from the analysis. Neutral evolution in drosophilids was based on positions 8–30 in short introns of protein coding genes (Parsch et al. 2010). Thereby only introns of single transcript genes were analyzed to ensure that the sequence is exclusively located in an intron and does not overlap with exons of other splice forms. Introns in genes with overlapping gene annotations on the same or opposite strand were discarded.

Selective Constraints

The strength of selection on a sequence of interest in a species was estimated by calculating the amount of selective constraint C ($=1-\frac{N_{obs}}{N_{neut}}$), where N_{obs} is the number of observed nucleotide substitutions between two closely related species and N_{neut} is the number of substitutions in a neutrally evolving region of the same length. We obtained N_{obs} in tRNAs for each species pair by concatenating all single tRNA orthologs. Therefore, positions of the alignment were classified into paired and unpaired states according to the consensus tRNA structure for the two sequences. Subsequently, N_{obs} was obtained separately for paired and unpaired alignment positions. The estimation of constraints may be confounded by several factors. Usually, the rate of substitutions in mammals is increased for dinucleotides in a CpG context through an elevation of the C → G transversion rates after the methylation of cytosine (Siepel and Haussler 2004). For that reason, all CpG-prone sites were excluded from the analysis by removal of all sites that are preceded by a C or followed by a G in the mammalian sequences (Gaffney and Keightley 2008). Furthermore, it was shown before that the GC content of the sequence and its deviation from the equilibrium GC content (GC*) will lead to increased rates of substitutions (Piganeau et al. 2002; Piskol and Stephan 2008). Therefore, differences in GC content between species pairs were accounted for by replacing N_{neut} with the expected number of substitutions (N_{exp}) that was calculated from ARs following the method of Halligan et al. (2004). Thereby, substitution rates that change the GC content were adjusted according to GC*, which was assumed to be 0.37 (Halligan et al. 2004; Khelifi et al. 2006; Duret and Arndt 2008). In all cases, 95% confidence intervals (CIs) for constraints were obtained by bootstrapping the tRNA alignments by column (while ensuring that the number of paired and unpaired columns in the bootstrapped alignment remained the same).

Results and Discussion

Expected Selective Pressures in RNA Molecules

We used the selective constraint (C) defined by Halligan et al. (2004) as a proxy for the level of selection on tRNA molecules. C describes the portion of deleterious mutations that are removed from the sequence due to purifying selection and is defined as $C=1-\frac{N_{obs}}{N_{exp}}$ (see Materials and Methods). $\frac{N_{obs}}{N_{exp}}$ is equal to $\frac{\bar{T}_{neut}}{\bar{T}}$, where $\bar{T}$ and $\bar{T}_{neut}$ are the expected fixation times for deleterious and neutral mutations, respectively (Innan and Stephan 2001; Piskol and Stephan 2011). Therefore, the expected values for C can be described in terms of the theoretical predictions for the fixation times as

$$C = 1 - \frac{\bar{T}_{neut}}{\bar{T}}. \quad (1)$$

Due to the dependence of the fixation times on θ and $N_e s$, also C will be influenced by these parameters. The resulting relationship between selective constraints and $N_e s$ (fig. 2) for coevolving sites (C_{coev}) and independently evolving sites (C_{ind}) can be obtained by using the expected fixation times ($\bar{T}_{coev}$ and $\bar{T}_{ind}$) and their neutral analogs in equation (1), respectively. We used Kimura's unidirectional models for $\bar{T}_{coev}$ and $\bar{T}_{ind}$ (Kimura 1980, 1985) because they are directly comparable in terms of model assumptions and parameters. However, the predictions made here are qualitatively the same as for models that take reversibility of the mutation process into account. Assuming that s is constant between species, the comparison of C_{coev} and C_{ind} allows for three main predictions in the case of weak purifying selection against new mutations:

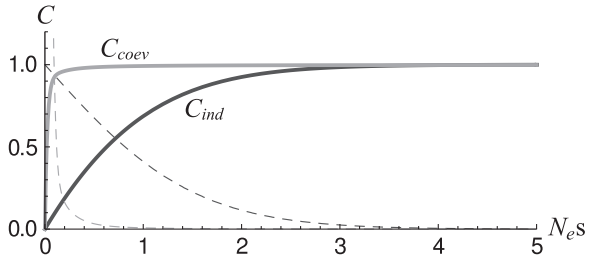

FIG. 2.—Expected selective constraints at independently evolving sites (C_{ind}) and coevolving sites (C_{coev}) as a function of the scaled selection coefficient $N_e s$. Dashed lines indicate the corresponding slopes. There exists a range of $N_e s$ in which C_{ind} increases more rapidly than C_{coev}. Therefore, the steeper slope for C_{ind} results in a larger difference in constraints at independently evolving sites than at coevolving sites between species with different N_e. The trajectories for C_{ind} and C_{coev} were obtained from Kimura's unidirectional models for the expected fixation times of mutant alleles in a population (eq. 13 from Kimura 1980 and eq. 16 from Kimura 1985) for a mutation rate $\mu = 2{:}5 \times 10^{-8}$ and selection coefficient $s = 10^{-4}$.

1. Coevolving sites are under stronger selective constraints than independently evolving sites (i.e., $C_{coev} > C_{ind}$),
2. Constraints increase with increasing effective population size (i.e., $C_{ind}\ (N_{e1}) < C_{ind}\ (N_{e2})$ and $C_{coev}\ (N_{e1}) < C_{coev}\ (N_{e2})$ for $N_{e1} < N_{e2}$), and
3. There exists a range of $N_e s$ in which N_e has a stronger effect on the evolution at independently evolving than on coevolving sites (i.e., $|C_{ind}\ (N_{e1}) - C_{ind}\ (N_{e2})| > |C_{coev}(N_{e1}) - C_{coev}(N_{e2})|$).

These general observations are independent of differences in scaled mutation rates (supplementary fig. S1, Supplementary Material online) and also imply that a change in N_e will result in small differences between C for large $N_e s$ but in large differences if $N_e s$ is small (fig. 2).

We can use tRNA molecules to test these predictions by assuming that tRNA positions, which are not involved in secondary structure formation (here denoted as "unpaired" positions), evolve under the independent model, whereas changes at nucleotide positions that are involved in WC pair formation with other partners within the sequence ("paired" positions) will be subject to coevolutionary dynamics. It is important for the analysis that N_e differs between tRNA molecules. This can be achieved by comparing orthologous tRNAs between species pairs of different long-term N_e but can also be tested within species pairs through the comparison of constraints between X chromosomal and autosomal tRNAs that differ in N_e.

Data Set

To investigate the effect of N_e on selective constraints in tRNAs, we collected data sets of orthologous tRNAs in seven species pairs of different N_e (table 1). We were able to extract approximately the same numbers of orthologous tRNAs for all pairs (a list of genomic positions is available from the authors upon request). Only for murids and birds, a smaller number of tRNAs was available. Although it might be expected that the amount of identifiable orthologous tRNAs will decrease with increasing divergence between species, we did not observe such a correlation. However, for all species, only relatively small numbers of tRNAs (if any) were identified on the X chromosome compared with the autosomes. This is not due to a low rate of detection of orthologs on the X chromosome but rather due to a significant underrepresentation of tRNA annotations on the X chromosomes. For instance, the initial set of tRNA annotations in the human genome contained 13 annotations on the X chromosome but 543 on the autosomes. Considering the contribution of the X chromosome to the complete genetic material, 28 tRNAs would have been expected to be located on the X chromosome and 528 on the autosomes, which constitutes a significant deviation from the observed numbers ($\chi^2 = 8.265$, $P = 0.004$). The same is true for drosophilid tRNAs. In general, the GC content in the paired portion of tRNA molecules is larger than for unpaired nucleotides. In particular, paired regions in Drosophila and birds show elevated levels of GC nucleotides. For these species, no specific increase in mutations due to CpG dinucleotides was expected. Therefore, we did not apply the procedure of Gaffney and Keightley (2008), which usually removes a large portion of guanines and cytosines from the sequence and resulted in a lower number of G and C nucleotides in vertebrates.

Core and Peripheral Sets of tRNAs

The total sets of orthologous tRNAs consisted only of molecules that fit the Rfam tRNA covariance model with high probability (relative to a null model that assumes no structure), which was reflected in INFERNAL bit scores $S > 35$ (see Materials and Methods). We noticed that the distribution of S for most vertebrate pairs is bimodal with a valley at $S \approx 60$ (supplementary fig. S2, Supplementary Material online). Therefore, the initial set was separated into two subsets according to this value. tRNAs with very high scores ($S \geq 60$) were denoted as a "core" set because they share great structural similarity with tRNAs from other species in various kingdoms of life. The second ("peripheral") set consisted of tRNAs with lower similarity to the consensus structure of a tRNA ($35 < S < 60$). This partitioning was performed because we suspected that tRNAs in the core set are under stronger selective constraints, whereas constraints in peripheral tRNAs are more relaxed. We assumed that under these circumstances, N_e will have stronger influence in the peripheral set and will result in more pronounced differences between C, as expected for slightly deleterious mutations. Here, our notion of a core set is based on the structural similarity of tRNAs and differs from the definition of Rogers et al. (2010) who defined a core set based on the conservation of tRNAs throughout the Drosophila genus. According

to tRNAScan-SE (Lowe and Eddy 1997), our data set contains tRNAs that encode 20 amino acids (supplementary fig. S3, Supplementary Material online). Most tRNAs that encode a specific amino acid are present in both core and peripheral sets. We did not find a preference of tRNAs for either the core or the peripheral set depending on their potential to encode essential or nonessential amino acids ($\chi^2 = 0.0781$, $P = 0.7799$). Although several species pairs contain pseudogenized tRNAs in the peripheral sets, this amount is comparatively small (human/macaque = 6, macaque/marmoset = 2, dog/cat = 6, mouse/rat = 1) and will only have a minor impact on the effect of N_e in our analysis. Also the length of the variable region that allows to discriminate between Class I tRNAs (short variable region of 4–5 nt) and Class II tRNAs (long variable region of > 10 nt, which is supposed to form a short helix) (Rich and RajBhandary 1976) does not correlate with the presence of tRNAs in core and peripheral sets. In that respect, although $tRNA^{Leu}$ and $tRNA^{Ser}$ are both representatives of Class II and contain a long variable region, the former is more abundant in the peripheral set, whereas the latter occurs more often in the core set. Vice versa, several Class I tRNAs are more abundant in the core set, whereas others are present in higher numbers in the peripheral set. This also suggests that the variable region only plays a minor role in our classification of tRNAs according to the Rfam bit score. Unusually high (or low) bit scores (and therefore classification of tRNAs in core or peripheral set) may have also been caused by a biased nucleotide composition. However, we did not observe any indication that high scores in our data were related to an exceptionally high or low GC content (supplementary fig. S4, Supplementary Material online).

The Influence of the Effective Population Size on Constraints in Nuclear-Encoded tRNAs

To test the predictions that are based on Kimura's models for sequence evolution at independently evolving and coevolving sites under continued mutation pressure (Kimura 1980, 1985), we calculated selective constraints at paired (C_{paired}) and unpaired ($C_{unpaired}$) positions in orthologous tRNAs for all species pairs (fig. 3*a*; supplementary table S1, Supplementary Material online). Thus, we related the rate of molecular evolution in tRNAs to evolutionary rates obtained from the corresponding neutral standard (supplementary table S5, Supplementary Material online). Depending on the species pair, the obtained values for C_{paired} and $C_{unpaired}$ fall into the ranges of (0.884, 0.996) and (0.698, 0.982), respectively, and thus surpass constraints at nonsynonymous sites in protein coding genes of hominids, murids, and drosophilids (Eőry et al. 2010; Parsch et al. 2010). For each species pair, we were able to observe significantly higher C_{paired} than $C_{unpaired}$ values (CIs do not overlap), as was expected from the comparison of independently evolving and coevolving

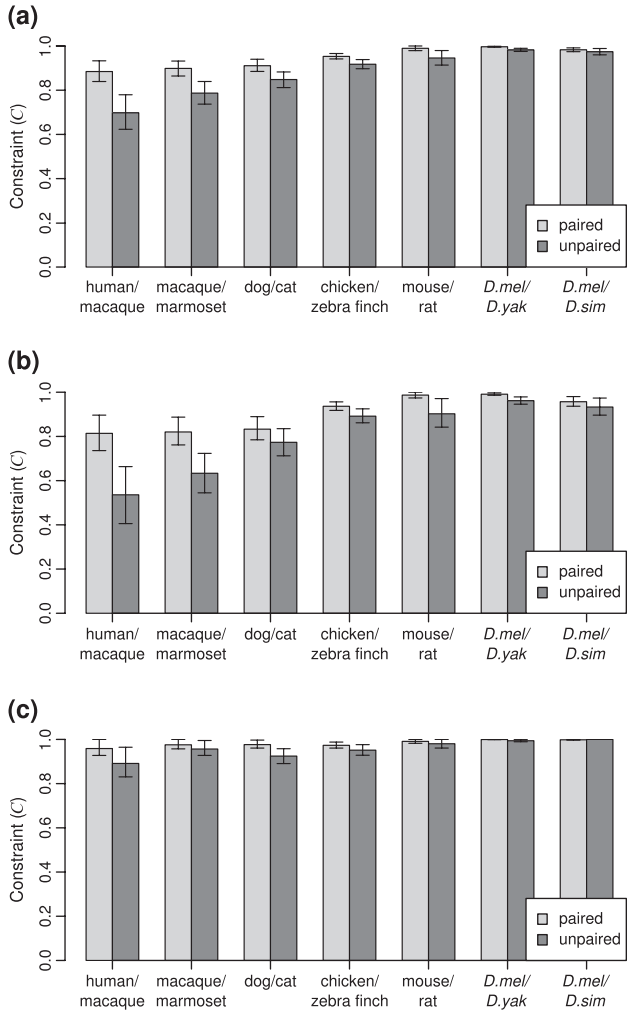

Fig. 3.—Constraint (*C*) for paired (light gray) and unpaired (dark gray) positions in orthologous tRNAs of different species pairs for (*a*) the whole data set, (*b*) peripheral set, and (*c*) core set.

sites under Kimura's models. The larger C_{paired} can be explained by the requirement for paired nucleotides to maintain their conformation and thus to preserve the secondary structure of the molecule.

Further examination of figure 3*a*, in which species pairs were arranged by increasing N_e from left to right, indicates that constraints also increase in the same order and verifies that C_{paired} and $C_{unpaired}$ increase with increasing N_e—the second prediction that followed from Kimura's models. For instance, the species pairs human/macaque, chicken/zebra finch, and *D. melanogaster*/*D. yakuba*, in that order, have significantly increasing N_e in the ranges of 10^4, 10^5, and 10^6, respectively. At the same time, the corresponding values of C_{paired} increase from (0.839, 0.933) to (0.942, 0.966) and (0.994, 0.999) and thus significantly differ as well. The same relationship also exists at unpaired sites to an even larger extent. This observation immediately

results in the third prediction of Kimura's models, which stated that constraints at independently evolving sites are affected by changes in N_e to a larger extent than at coevolving sites. As a result, larger differences can be observed in constraints at unpaired sites between species of different N_e than at paired sites. For example, the difference in $C_{unpaired}$ between primates and murids $\Delta C_{unpaired}$ (prim/mur) = 0.248, whereas at paired positions ΔC_{paired} (prim/mur) = 0.105 and thus much smaller. The same is true for most comparisons between other species pairs. However, as was expected, with increasing N_e (and thus also increasing C), the discrepancies between constraints in different species become smaller. Furthermore, the stronger effect of N_e on $C_{unpaired}$ also manifests itself in comparisons within species pairs through a decrease in the difference $|C_{paired} - C_{unpaired}|$ with increasing N_e. In general, these particular patterns in selective constraints are caused by the interplay of $N_e s$ and the influence of mutation rates (supplementary fig. S1, Supplementary Material online). Thereby, the relationship between C_{ind} is mostly independent of θ, whereas increased C_{coev} is expected when θ is low. This is particularly apparent in the pair human/macaque, which has the lowest θ value (=0.001) in our analysis.

When comparing constraints between species pairs with different divergence (k), it might have been expected that C increases with increasing k because only orthologous tRNAs with higher conservation can be identified for distant sequences. However, similar to the nonsignificant relation between k and the number of identified tRNAs (n), the relationships between k and C (Kendall's $\tau = -0.43$, $P = 0.24$) as well as n and C (Kendall's $\tau = -0.39$, $P = 0.22$) are not significant. Therefore, we can exclude an influence of divergence and number of identified tRNAs on estimates of constraints in our data. Even if we assume that divergence between chicken/zebra finch and *D. melanogaster*/*D. yakuba* is of a magnitude such that multiple hits cannot be safely ignored (which would result in an underestimation of C for these species), the general pattern persists. For instance, our hypothesis still holds if we replace the *D. melanogaster*/*D. yakuba* pair by *D. melanogaster*/*D. simulans* (which has much smaller divergence and thus lower probability for multiple hits).

Stronger Constraints in Core tRNA Genes

It is also of interest to determine whether the effect of N_e on C is influenced by the overall strength of purifying selection in tRNAs. Therefore, constraints in tRNAs were analyzed after splitting the data into core and peripheral sets. If selection in the former set is strong, then the effect of N_e on C in this set should be low and vice versa. If, on the other hand, N_e is not responsible for the pattern observed above, the core and peripheral sets should both show signs of approximately equally reduced constraints in species of small N_e. However, the latter assumption can be clearly rejected based on figure 3*b* and *c*. Consistent with the assumption that selection pressure is higher in tRNAs belonging to the core set, we are able to observe higher constraints in tRNAs from the core set compared with the peripheral set for all species pairs. It is more important, however, that the increase in C with increasing N_e is strong in the peripheral set of tRNAs (fig. 3*b*) and only very weak (but still present) in the core set of tRNAs as shown in figure 3c. The latter effect is better visible after applying the transformation $\log(1 - C)$ (supplementary fig. S5c, Supplementary Material online). Therefore, we can assume that selective constraints in tRNAs are most likely influenced by N_e and that this effect is strong if selection is weak, whereas in the case of strong selection, our observations follow theoretical predictions, which show that the fixation time of compensatory double mutants is rather independent of N_e (eq. 8c in Stephan 1996). Even though a separation of the data according to a single score may be crude, our results show that it allows us to distinguish between two sets of tRNAs that seem to be under different selective constraints. Further evidence for this hypothesis comes from the observed GC contents at non–CpG-prone nucleotides of the two sets. Compared with the peripheral sets, the core sets show higher GC contents in the paired portion of tRNA molecules for most species pairs (table 1). A higher GC content was shown to be associated with an increased substitution rate (Előry et al. 2010; Piskol and Stephan 2011). Therefore, if tRNAs in the core set were subject to the same constraints as tRNAs in the peripheral set, more substitutions would have been expected in tRNAs belonging to the core set. However, the exact opposite is observed, which justifies our separation of the data in two sets and confirms higher constraints in the core set.

To understand how a difference in GC content between neutrally evolving regions and regions of interest may influence selective constraints, we resampled the neutrally evolving regions such that they match the GC content at paired and unpaired nucleotides in vertebrate tRNAs. Subsequently, we repeated the calculation of selective constraints for all tRNAs, as well as for the peripheral sets and core sets of tRNAs separately (supplementary table S4, Supplementary Material online). Although the effect of N_e on selective constraints is slightly weakened after resampling of neutral regions, the same overall pattern as before persists (compare supplementary tables S1 and S4, Supplementary Material online). Constraints at paired positions only change marginally after resampling of neutral sites. This is presumably due to the high overall selective pressure at paired sites. However, the resampling of neutral regions to match the GC content at unpaired sites showed a larger effect on constraint estimates. It led to a decrease in GC content and therefore an increase of substitution rates at neutral positions, which ultimately resulted in higher estimated selective constraints at unpaired sites. This effect was expected due to the overall

negative correlation between substitution rates and GC content (Előry et al. 2010) for GC contents between 30% and 50%. These findings suggest that our estimates of selective constraint at paired sites, which were obtained without resampling of neutral standards (fig. 3 and supplementary table S1, Supplementary Material online), are robust against differences in GC content. For unpaired sites, our estimates without resampling are conservative and an adjustment through resampling leads to higher estimated constraints.

Differences in Selective Constraints between Autosomes and X Chromosome

Apart from the differences in N_e between species and their effect on nucleotide variation, effects of N_e on C might also be expected within species. If the contribution of genetic material to the next generation is equal for males and females, the expected ratio of X to autosomal N_e ($\frac{N_{eX}}{N_{eA}}$) is 0.75, due to the presence of the X chromosome in a single copy in males. However, this assumption is not always met. It was reported previously (Hutter et al. 2007) that in a European population of *D. melanogaster* $\frac{N_{eX}}{N_{eA}} = 0.49$ and thus lower than expected, whereas in an African (ancestral) population of *D. melanogaster* $\frac{N_{eX}}{N_{eA}} = 0.90$. Other studies also suggest that $\frac{N_{eX}}{N_{eA}}$ in ancestral populations may be larger than expected (Andolfatto 2001; Connallon 2007; Singh et al. 2007). Therefore, the efficacy of selection may differ between X chromosome and autosomes and may lead to different selective constraints.

To test whether differences in constraints are observed between the X chromosome and autosomes, we divided the 277 orthologous tRNAs for the *D. melanogaster/D. yakuba* pair according to their genomic location into 21 X-linked and 256 autosomal tRNAs and obtained selective constraints separately for these two sets. It was shown before (Betancourt et al. 2002) that evolutionary rates do not differ between chromosomes in *D. melanogaster*. Nonetheless, we avoided any confounding effects due to systematic differences in mutation rates between X chromosome and autosomes by using introns that were exclusively located on the X chromosome or autosomes as neutral standards for the evolution of X and autosomes, respectively (supplementary table S5, Supplementary Material online). Table 2A shows constraints in paired and unpaired regions for all X-linked and autosomal tRNAs. Again, paired positions are subject to significantly higher evolutionary constraints than positions that are not involved in the formation of WC base pairs for both autosomes and the X chromosome. More interestingly, lower constraints can be observed on the X chromosome (C_x) than on the autosomes (C_A) (presumably due to the smaller N_e of the X chromosome). The difference in constraints between autosomes and X chromosome ($|C_A - C_x|$) is particularly apparent in unpaired portions of tRNAs and is in accordance with theoretical predictions that N_e will have a large impact on evolution at independently evolving sites. Lower constraints on the X chromosome might have also been observed due to reduced evolutionary rates in the neutral standard on the X chromosome rather than increased rates of fixation in tRNAs. However, our neutral divergence estimate for the X chromosome is slightly larger than for autosomes (supplementary table S5, Supplementary Material online) and hence cannot be held accountable for lower constraints in X-linked tRNAs but suggests that a lower C_x at paired and unpaired sites is indeed due to a higher number of fixed differences in tRNAs on the X chromosome. Similar patterns of higher divergence on the X chromosome have also been observed at nonsynonymous sites in the *D. melanogaster* and *D. yakuba* lineages (Begun et al. 2007). Given the estimates of $\frac{N_{eX}}{N_{eA}} > 0.75$ for the ancestral population of *D. melanogaster* from previous studies and assuming that mutations in tRNAs will be mostly slightly deleterious, we would have expected that the rate of fixation on the X chromosome was reduced compared with the autosomes (Vicoso and Charlesworth 2009; Mank et al. 2010). However, the slightly lower constraints on the X chromosome suggest faster fixations of mildly deleterious mutations in X-linked tRNAs (compared with autosomal tRNAs) and point to a long-term $\frac{N_{eX}}{N_{eA}}$, which is smaller than 0.75 for tRNAs in the *D. melanogaster/D. yakuba* pair (see fig. 3 in Vicoso and Charlesworth 2009).

In addition, we confirmed that the lower constraints in X-linked tRNAs are in fact significant and did not arise simply by chance due to the small sample size of tRNAs on the X chromosome. For this reason, we generated 1000 data sets by randomly splitting the 277 Drosophila tRNAs into sets of 21 and 256 instances (resembling the sizes of X and autosomal data). For all repetitions, we calculated constraints at paired and unpaired sites in the large and small sets, respectively, and thus obtained distributions for $|C_A - C_x|$ that would be expected at random (fig. 4). Indeed, the observed values of $|C_A - C_x|$ are significantly larger than in the randomly assembled sets. This is true for paired regions ($|C_A - C_x| = 0.0144$, $P = 0.031$*) and to a larger extent in the unpaired portion of tRNAs ($|C_A - C_x| = 0.0505$, $P = 0.004$**).

When repeated separately for tRNAs grouped in core and peripheral sets, the same analysis also supports our previous conjecture that effects of N_e on the difference in constraints between X chromosome and autosomes are large if selection is weak (table 2, B) but much smaller when selection on the tRNA molecule is overall strong (table 2, C). This becomes apparent through significant values of $|C_A - C_x|$ in the peripheral set, whereas no significant differences are observed in the core set of tRNAs (table 2 and supplementary figs. S6 and S7, Supplementary Material online).

Table 2
Selective Constraints for Paired (C_{paired}) and Unpaired ($C_{unpaired}$) Positions in Drosophilid tRNAs Located on the Autosomes and the X Chromosome for (A) the Whole Data Set, (B) Peripheral Set, and (C) Core Set

		C_{paired}	(95% CI)	$C_{unpaired}$	(95% CI)	$\lvert C_{paired} - C_{unpaired} \rvert$
A.	Autosomes	0.9977	(0.9961,0.9996)	0.9862	(0.9804,0.9932)	0.0115
	X chromosome	0.9833	(0.9707,1.000)	0.9357	(0.8900,0.9879)	0.0467
	$C_A - C_x$	0.0144	0.031*	0.0505	0.004**	
B.	Autosomes	0.9937	(0.9894,0.9989)	0.9698	(0.9547,0.9860)	0.0239
	X chromosome	0.9472	(0.8944,1.000)	0.8369	(0.7073,0.9750)	0.1103
	$C_A - C_x$	0.0455	0.015*	0.1329	0.001**	
C.	Autosomes	1.000	(1.000,1.000)	0.9961	(0.9922,1.000)	0.0059
	X chromosome	0.9945	(0.9891,1.000)	0.9744	(0.9488,1.000)	0.0201
	$C_A - C_x$	0.0055	0.085	0.0217	0.067	

NOTE.—$C_A - C_x$ is the difference in constraints between tRNAs encoded on the autosomes and X chromosome for paired and unpaired sites. In this case, values in the 95% CI column give the *P* value for the difference. Significance levels: $*P < 0.05$, $**P < 0.01$.

Conclusions

We showed that divergence patterns in nuclear-encoded tRNA molecules of vertebrate and drosophilid species follow general theoretical predictions for sequence evolution under mutational pressure. Larger selective constraints can be observed with increasing N_e. This effect is weaker at coevolving sites than at independently evolving sites. The influence of N_e on nucleotide variation is not exclusive to tRNAs but seems to be universal in RNA molecules as microRNAs exhibit a similar increase of selective constraints with increasing N_e (supplementary table S6, Supplementary Material online).

Here, we did not take the effect of recombination on $\bar{T}_{coev}$ into account. It was shown previously that recombination may retard the rate of fixation of compensatory double mutants in RNA molecules even when the distance in sequence (d) between paired nucleotides is small ($50 < d < 250$) (Piskol and Stephan 2008). For mildly deleterious single mutants, recombination also has the potential to combine individual mutant alleles thus leading to complex adaptations (Lynch 2010; Weissman et al. 2010). However, usually fixation times of double mutants are only moderately affected by recombination.

Although we cannot completely rule out that some of the substitutions investigated in our study are of an adaptive nature, we assumed that the vast majority of mutations in tRNAs are deleterious. Given that tRNA molecules have a well-defined function, mutations will most likely alter the structure and original conformation of the molecule in space thus potentially changing its functionality and leading to a decrease in fitness. Very important for our analysis was the assumption that WC base pairs, which form the secondary structure of the tRNA, are subject to coevolutionary dynamics, whereas other nucleotides in the tRNA, whether involved in non-WC pairs or completely unpaired, may evolve independently. This was shown to be the case in bacterial ribosomal RNAs (Dutheil et al. 2010) and is also directly applicable to tRNAs due to the universality of base pairs (Leontis and Westhof 2001).

In a recent study (Piskol and Stephan 2011), we reported that selective constraints in computationally predicted noncoding RNAs that are encoded in the nuclear genomes of drosophilids and hominids differ in their magnitude between the two genera. We suggested that N_e is responsible for this difference and results in stronger selective constraints in drosophilids. In general, the definition of neutral evolution and the distinction between neutrality and

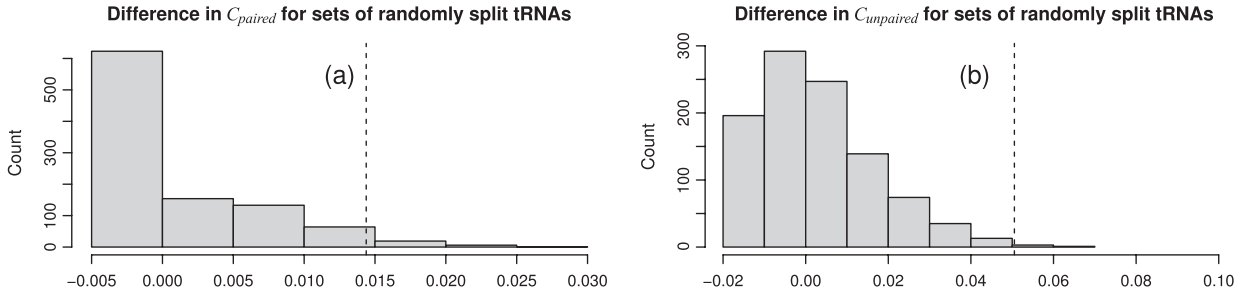

FIG. 4.—Histogram of differences in constraints at (*a*) paired and (*b*) unpaired positions between sets of 256 and 21 tRNAs that were created by randomly splitting 277 orthologous tRNAs of *Drosophila melanogaster* and *D. yakuba* 1000 times. The dashed lines represent the observed values of $|C_A - C_x|$ taken from table 2, A.

purifying selection in terms of N_e is complicated and has been the topic of many controversies (Nei et al. 2010). Even though the definition of neutrality may have changed over the years (Ohta and Gillespie 1996; Nei 2005), our present results demonstrate that N_e indeed can be held accountable for differences in the efficacy of selection and does so by affecting coevolving and independently evolving sites to different degrees. We suggest that there exists a set of peripheral tRNAs for which mutations are slightly deleterious and scaled selective coefficients are only of a moderate size ($|N_e s| \leq 5$). For this regime, the pattern of increasing constraints is strongly influenced by the effective population size (fig. 3*b*) and follows the theoretical predictions for fixation times of deleterious mutations in Kimura's one- and two-locus models (Kimura 1980, 1985). The remaining (core) tRNAs might be subject to stronger evolutionary restrictions, and thus, divergence patterns in these molecules are less susceptible to differences in N_e. Although the reason for the existence of different constraints in tRNAs is largely unknown, the expression of tRNAs and the usage of optimal codons may play a role in the maintenance of certain levels of selective pressures in these molecules (Moriyama and Powell 1997; Carlini and Stephan 2003; Hense et al. 2010).

Our results have also direct consequences for the inference of phylogenetic relationships between taxa that differ in their long-term effective population size. If the estimation of branch lengths is performed using independently evolving sites that are subject to weak purifying selection (e.g., synonymous sites), then the length of branches leading to taxa with large N_e might be underestimated to a larger extent than for taxa with small N_e.

Supplementary Material

Supplementary tables S1–S6 and figures S1–S7 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors would like to thank John Parsch for helpful discussion as well as Olivier Tenaillon and Fyodor Kondrashov for comments on the manuscript. This work was supported by grant Ste 325/8 to W.S. from the Deutsche Forschungsgemeinschaft.

Literature Cited

Andolfatto P. 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol Biol Evol. 18(3):279–290.

Andolfatto P, Wong KM, Bachtrog D. 2011. Effective population size and the efficacy of selection on the X chromosomes of two closely related Drosophila species. Genome Biol Evol. 3:114–128.

Baines JF, Harr B. 2007. Reduced X-linked diversity in derived populations of house mice. Genetics. 175(4):1911–1921.

Begun DJ, et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol. 5(11):e310.

Betancourt AJ, Presgraves DC, Swanson WJ. 2002. A test for faster X evolution in Drosophila. Mol Biol Evol. 19(10):1816–1819.

Brown KM, et al. 2010. Compensatory mutations restore fitness during the evolution of dihydrofolate reductase. Mol Biol Evol. 27(12): 2682–2690.

Carlini DB, Stephan W. 2003. In vivo introduction of unpreferred synonymous codons into the Drosophila *Adh* gene results in reduced levels of ADH protein. Genetics. 163(1):239–243.

Chamary JV, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat Rev Genet. 7(2): 98–108.

Charlesworth B. 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 10(3):195–205.

Chen Y, Stephan W. 2003. Compensatory evolution of a precursor messenger RNA secondary structure in the *Drosophila melanogaster Adh* gene. Proc Natl Acad Sci U S A. 100(20):11499–11504.

Chen Y, et al. 1999. RNA secondary structure and compensatory evolution. Genes Genet Syst. 74(6):271–286.

Connallon T. 2007. Adaptive protein evolution of X-linked and autosomal genes in Drosophila: implications for faster-X hypotheses. Mol Biol Evol. 24(11):2566–2572.

Duret L, Arndt PF. 2008. The impact of recombination on nucleotide substitutions in the human genome. PLoS Genet. 4(5):e1000071.

Dutheil JY, Jossinet F, Westhof E. 2010. Base pairing constraints drive structural epistasis in ribosomal RNA sequences. Mol Biol Evol. 27(8):1868–1876.

Edgar RC. 2004. Muscle: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 5:113.

Előry L, Halligan DL, Keightley PD. 2010. Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes. Mol Biol Evol. 27(1):177–192.

Evans BJ, Pin L, Melnick DJ, Wright SI. 2010. Sex-linked inheritance in macaque monkeys: implications for effective population size and dispersal to Sulawesi. Genetics. 185(3):923–937.

Eyre-Walker A, Keightley PD, Smith NGC, Gaffney D. 2002. Quantifying the slightly deleterious mutation model of molecular evolution. Mol Biol Evol. 19(12):2142–2149.

Gaffney DJ, Keightley PD. 2008. Effect of the assignment of ancestral CpG state on the estimation of nucleotide substitution rates in mammals. BMC Evol Biol. 8:265.

Gardner PP, et al. 2009. Rfam: updates to the RNA families database. Nucleic Acids Res. 37(Database issue):D136–D140.

Gillespie JH. 1999. The role of population size in molecular evolution. Theor Popul Biol. 55(2):145–156.

Halligan DL, Eyre-Walker A, Andolfatto P, Keightley PD. 2004. Patterns of evolutionary constraints in intronic and intergenic DNA of Drosophila. Genome Res. 14(2):273–279.

Hense W, Anderson N, Hutter S, Stephan W, Parsch J, Carlini DB. 2010. Experimentally increased codon bias in the drosophila *Adh* gene leads to an increase in larval, but not adult, alcohol dehydrogenase activity. Genetics. 184(2):547–555.

Hofacker IL, Fekete M, Stadler PF. 2002. Secondary structure prediction for aligned RNA sequences. J Mol Biol. 319(5):1059–1066.

Hutter S, Li H, Beisswanger S, de Lorenzo D, Stephan W. 2007. Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosomewide single nucleotide polymorphism data. Genetics. 177(1):469–480.

Ichinose M, Iizuka M, Kado T, Takefu M. 2008. Compensatory evolution in diploid populations. Theor Popul Biol. 74(2):199–207.
Innan H, Stephan W. 2001. Selection intensity against deleterious mutations in RNA secondary structures and rate of compensatory nucleotide substitutions. Genetics. 159(1):389–399.
Iwasa Y, Michor F, Nowak MA. 2004. Stochastic tunnels in evolutionary dynamics. Genetics. 166(3):1571–1579.
Jennings WB, Edwards SV. 2005. Speciational history of Australian grass finches (Poephila) inferred from thirty gene trees. Evolution. 59(9): 2033–2047.
Jurka J, et al. 2005. Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 110(1–4):462–467.
Kent WJ, et al. 2002. The human genome browser at UCSC. Genome Res. 12(6):996–1006.
Khelifi A, Meunier J, Duret L, Mouchiroud D. 2006. GC content evolution of the human and mouse genomes: insights from the study of processed pseudogenes in regions of different recombination rates. J Mol Evol. 62(6):745–752.
Kimura M. 1980. Average time until fixation of a mutant allele in a finite population under continued mutation pressure: studies by analytical, numerical, and pseudo-sampling methods. Proc Natl Acad Sci U S A. 77(1):522–526.
Kimura M. 1983. The neutral theory of molecular evolution. Cambridge (UK): Cambridge University Press.
Kimura M. 1985. The role of compensatory neutral mutations in molecular evolution. J Genet. 64:7–19.
Kimura M, Ohta T. 1969. The average number of generations until fixation of a mutant gene in a finite population. Genetics. 61(3):763–771.
Kirby DA, Muse SV, Stephan W. 1995. Maintenance of pre-mRNA secondary structure by epistatic selection. Proc Natl Acad Sci U S A. 92(20):9047–9051.
Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. RNA. 7(4):499–512.
Li H, Stephan W. 2006. Inferring the demographic history and rate of adaptive substitution in Drosophila. PLoS Genet. 2(10):e166.
Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25(5):955–964.
Lynch M. 2010. Scaling expectations for the time to establishment of complex adaptations. Proc Natl Acad Sci U S A. 107(38):16577–16582.
Mank JE, Vicoso B, Berlin S, Charlesworth B. 2010. Effective population size and the faster-X effect: empirical results and their interpretation. Evolution. 64(3):663–674.
Meer MV, Kondrashov AS, Artzy-Randrup Y, Kondrashov FA. 2010. Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. Nature. 464(7286):279–282.
Moriyama EN, Powell JR. 1997. Codon usage bias and tRNA abundance in Drosophila. J Mol Evol. 45(5):514–523.
Mural RJ, et al. 2002. A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. Science. 296(5573): 1661–1671.
Nawrocki EP, Kolbe DL, Eddy SR. 2009. Infernal 1.0: inference of RNA alignments. Bioinformatics. 25(10):1335–1337.
Nei M. 2005. Selectionism and neutralism in molecular evolution. Mol Biol Evol. 22(12):2318–2342.
Nei M, Suzuki Y, Nozawa M. 2010. The neutral theory of molecular evolution in the genomic era. Annu Rev Genomics Hum Genet. 11:265–289.
Ohta T. 1972. Population size and rate of evolution. J Mol Evol. 1(3): 305–314.
Ohta T. 1973. Slightly deleterious mutant substitutions in evolution. Nature. 246(5428):96–98.
Ohta T, Gillespie JH. 1996. Development of neutral and nearly neutral theories. Theor Popul Biol. 49(2):128–142.
Parsch J, Braverman JM, Stephan W. 2000. Comparative sequence analysis and patterns of covariation in RNA secondary structures. Genetics. 154(2):909–921.
Parsch J, Novozhilov S, Saminadin-Peter SS, Wong KM, Andolfatto P. 2010. On the utility of short intron sequences as a reference for the detection of positive and negative selection in Drosophila. Mol Biol Evol. 27(6):1226–1234.
Parsch J, Tanda S, Stephan W. 1997. Site-directed mutations reveal long-range compensatory interactions in the *Adh* gene of *Drosophila melanogaster*. Proc Natl Acad Sci U S A. 94(3):928–933.
Piganeau G, Eyre-Walker A. 2009. Evidence for variation in the effective population size of animal mitochondrial DNA. PLoS One. 4(2): e4396.
Piganeau G, Mouchiroud D, Duret L, Gautier C. 2002. Expected relationship between the silent substitution rate and the GC content: implications for the evolution of isochores. J Mol Evol. 54(1):129–133.
Piskol R, Stephan W. 2008. Analyzing the evolution of RNA secondary structures in vertebrate introns using Kimura's model of compensatory fitness interactions. Mol Biol Evol. 25(11):2483–2492.
Piskol R, Stephan W. 2011. Selective constraints in conserved folded RNAs of drosophilid and hominid genomes. Mol Biol Evol. 28(4): 1519–1529.
Rich A, RajBhandary UL. 1976. Transfer RNA: molecular structure, sequence, and properties. Annu Rev Biochem. 45:805–860.
Rogers HH, Bergman CM, Griffiths-Jones S. 2010. The evolution of tRNA genes in Drosophila. Genome Biol Evol. 2:467–477.
Siepel A, Haussler D. 2004. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. Mol Biol Evol. 21(3):468–488.
Singh ND, Macpherson JM, Jensen JD, Petrov DA. 2007. Similar levels of X-linked and autosomal nucleotide variation in African and non-African populations of *Drosophila melanogaster*. BMC Evol Biol. 7:202.
Smit A, Hubeley R, Green P. 1996–2010. Repeat Masker Open-3.0 [Internet]. [cited 2011 January 15]. Available from: http://www.repeatmasker.org.
Stephan W. 1996. The rate of compensatory evolution. Genetics 144(1):419–426.
Tweedie S, et al. 2009. FlyBase: enhancing Drosophila gene ontology annotations. Nucleic Acids Res. 37(Database issue):D555–D559.
Vicoso B, Charlesworth B. 2009. Effective population size and the faster-X effect: an extended model. Evolution 63(9):2413–2426.
Weinreich DM, Rand DM. 2000. Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. Genetics 156(1):385–399.
Weissman DB, Feldman MW, Fisher DS. 2010. The rate of fitness-valley crossing in sexual populations. Genetics 186(4):1389–1410.
Will S, Reiche K, Hofacker IL, Stadler PF, Backofen R. 2007. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. PLoS Comput Biol. 3(4):e65.
Woolfit M, Bromham L. 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol Biol Evol. 20(9):1545–1555.
Woolfit M, Bromham L. 2005. Population size and molecular evolution on islands. Proc Biol Sci. 272(1578):2277–2282.

Associate editor: Marta Wayne