PNAS RESEARCH ARTICLE | EVOLUTION

Dinucleotide preferences underlie apparent codon preference reversals in the *Drosophila melanogaster* lineage

Haruka Yamashita[a,b], Tomotaka Matsumoto[a,b,1], Kent Kawashima[a,2], Hassan Sibroe Abdulla Daanaa[a,b], Ziheng Yang[c], and Hiroshi Akashi[a,b,3]

Affiliations are included on p. 8.

Edited by Andrew Clark, Cornell University, Ithaca, NY; received September 27, 2024; accepted April 21, 2025

We employ fine-scale population genetic analyses to reveal dynamics among interacting forces that act at synonymous sites and introns among closely related *Drosophila* species. Synonymous codon usage bias has proven to be well suited for population genetic inference. Under major codon preference (MCP), translationally superior "major" codons confer fitness benefits relative to their less efficiently and/or accurately decoded synonymous counterparts. Our codon family and lineage-specific analyses expand on previous findings in the *Drosophila simulans* lineage; patterns in naturally occurring polymorphism demonstrate fixation biases toward GC-ending codons that are consistent in direction, but heterogeneous in magnitude, among synonymous families. These forces are generally stronger than fixation biases in intron sequences. In contrast, population genetic analyses reveal unexpected evidence of codon preference reversals in the *Drosophila melanogaster* lineage. Codon family-specific polymorphism patterns support reduced efficacy of natural selection in most synonymous families but indicate reversals of favored states in the four codon families encoded by NAY. Accelerated synonymous fixations in favor of NAT and greater differences for both allele frequencies and fixation rates among X-linked, relative to autosomal, loci bolster support for fitness effect reversals. The specificity of preference reversals to codons whose cognate tRNAs undergo wobble position queuosine modification is intriguing. However, our analyses reveal prevalent dinucleotide preferences for ApT over ApC that act in opposition to GC-favoring forces in both coding and intron regions. We present evidence that changes in the relative efficacy of translational selection and dinucleotide preference underlie apparent codon preference reversals.

nearly neutral evolution | codon usage | population genetics | major codon preference | *Drosophila*

Significance

Patterns in DNA sequence variation can reveal functional adaptation. Synonymous codon usage is a model system for studying nearly neutral evolution, and we combine genome-scale polymorphism data from natural populations of *Drosophila* species and reliable ancestral inference methods for "high resolution" (synonymous codon family- and lineage-specific) analyses. Population frequencies of synonymous mutations yield surprising signals of preference reversals from NAC to NAT codons in *Drosophila melanogaster*. We show that dinucleotide preferences (ApT over ApC) are a prevalent factor in these genomes in coding and noncoding regions and that dynamics in the balance of opposing codon and dinucleotide preferences underlie apparent codon preference reversals.

Codon usage bias is a well-developed system for studying evolutionary dynamics under nearly neutral genome evolution. A combination of sequence patterns and experimental findings support "major codon preference" (MCP), coadaptation between synonymous codon usage and cognate tRNA abundances and their modifications to allow efficient use of the translational machinery (i.e., greater elongation rates and reduced misincorporation) (reviewed in ref. 1). A wide range of microbes, as well as multicellular eukaryotes, show elevated usage of particular codons within synonymous families in highly expressed genes (reviewed in refs. 2, 3). Such codons are generally recognized by abundant and/or non-wobble-pairing cognate tRNAs and are referred to as "major" codons (4). Biochemical studies support both elevated elongation rates (5–7) and reduced misincorporations (8, 9) at major codons relative to their minor counterparts. Putative major codons can be identified through characterizations of tRNA pools (e.g., refs. 10, 11), experimental measures of translation rate (e.g., ref. 5), or by compositional trends among genes (e.g., refs. 12–18).

The main features of MCP can be captured in relatively simple, testable evolutionary models. Observed levels of codon usage bias are consistent with a balance among weak evolutionary forces including mutation, genetic drift, and weak natural selection (19, 20). Under MCP, mutations from minor to major codons confer small fitness benefits and mutations in the opposite direction are deleterious with similar magnitudes; population genetic approaches can test the model by contrasting evolutionary patterns between these predicted fitness classes (17). In addition, because translational effects of synonymous codon usage should be similar among genes, data can be pooled among loci to enhance the statistical power to detect minute fixation biases (i.e., forces that alter expected allele frequencies from generation to generation in a consistent direction). These features facilitate the study of evolutionary dynamics under a balance of weak forces and previous population genetic analyses have demonstrated directional forces that favor major codons over their synonymous counterparts (e.g., refs. 17, 21–24). GC-biased gene conversion can also underlie base compositional biases (25, 26), and distinguishing between natural

Author contributions: H.Y. and H.Akashi designed research; H.Y. and H.Akashi performed research; T.M., K.K., and Z.Y. contributed new reagents/analytic tools; H.Y., T.M., K.K., H.S.A.Daanaa, and H.Akashi analyzed data; and H.Y. and H.Akashi wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]Present address: SRL and Shizuoka Cancer Center Collaborative Laboratories Inc., Sunto-gun, Shizuoka 411-8777, Japan.

[2]Present address: Division of Evolutionary Studies of Complex Systems, Research Center for Integrative Evolutionary Science, The Graduate University for Advanced Studies, SOKENDAI, Hayama, Kanagawa 240-0193, Japan.

[3]To whom correspondence may be addressed. Email: hiakashi@nig.ac.jp.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2419696122/-/DCSupplemental.

Published XXXX.

selection and nonselective fixation biases has proven challenging. In addition, a number of empirical studies support beneficial effects of translational pauses in regulating protein folding or membrane insertion (reviewed in ref. 27). Major codons may be disadvantageous in some contexts, but such cases may be more strongly selected and relatively rare.

Here, we take advantage of genome-scale data from population samples of closely related *Drosophila* species to pursue evolutionary analyses at an increased resolution. We test fixation biases within individual synonymous families both among mutations segregating in populations and among fixations in ancestral lineages. Our analyses support MCP favoring predominantly G- and C-ending codons in *Drosophila simulans* but reveal appreciable heterogeneity in the magnitude of selective forces among synonymous families. In contrast, the *Drosophila melanogaster* lineage shows little evidence of fixation biases on synonymous mutations for most codon families. Population genetics theory predicts long-term stability of codon preferences under coadaptation between codon usage and tRNA pools, but surprisingly, we found that a subset of synonymous families, those encoded by NAY codons, show strong evidence for recent codon preference reversals in the *D. melanogaster* lineage. These findings, together with greater contrasts in both the SFS and fixation patterns at X-linked, relative to autosomal, loci make a compelling case for genome-wide shifts in NAY codon family adaptation. Intriguingly, all NAY codons, and no other codons, are recognized by cognate tRNAs that undergo a chemical (queuosine) modification within the anticodon region. The chemical precursor for this modification, queine, is obtained through the environment; diet- or habitat-based shifts in major codon definition have been suggested for other *Drosophila* lineages (28, 29) and seemed to be a compelling scenario to account for the patterns observed in this study.

However, NAT preference at NAY codons could also reflect dinucleotide ApT preference over ApC. We show patterns at synonymous codons as well as intron sites that support prevalent dinucleotide preference in coding and noncoding regions. ApT over ApC preference acts in opposition to GC fixation biases (including MCP) in both *D. melanogaster* and *D. simulans*. In the *D. melanogaster* lineage, reductions in the efficacy of GC fixation bias have shifted the balance of forces toward relatively stronger context-dependent T over C preference leading to *apparent* codon preference reversals at NAY codons. Although similar opposing forces act in *D. simulans*, the greater relative strength of GC-favoring factors, including MCP, can mask the signals of ApT over ApC preference. Our population genetic analyses reveal a dynamic balance among fixation biases underlying codon and base composition evolution among closely related *Drosophila* species.

Results and Discussion

We analyzed available population genomic DNA sequence data among closely related species from the *D. melanogaster* subgroup. We constructed CDS and intron sequence alignments for 21 *D. simulans* lines from a Madagascar population (24, 30) and 14 *D. melanogaster* lines from a Rwandan population (31). Reference sequences for *Drosophila yakuba* and *Drosophila erecta* (32), were used as outgroups (Fig. 1*A*). Classifying synonymous mutations by their predicted fitness effects, a key step for population genetic analyses of codon usage (17), requires inference of ancestral and derived states at polymorphic sites. The relatively short branch lengths in our tree allows reliable inference using a likelihood-based method that incorporates both biased and fluctuating base composition (33, 34; see *Materials and Methods*).

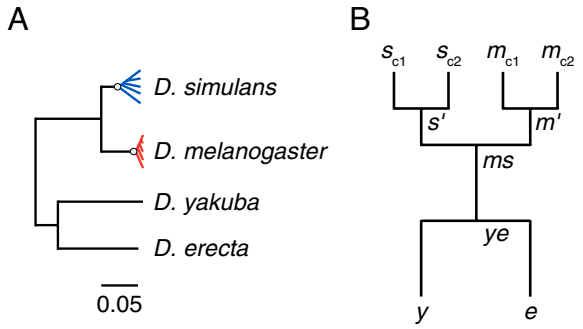

Fig. 1. Gene tree of *Drosophila melanogaster* subgroup species employed in this study. Population genetic analyses employed data from *Drosophila simulans* (*Dsim*), *D. melanogaster* (*Dmel*), and two outgroups, *Drosophila yakuba* (*Dyak*) and *Drosophila erecta* (*Dere*); only species included in our analyses are shown. (*A*) Genetic distances among DNA sequences from the four species. Branch lengths are numbers of nucleotide differences per site at autosomal short introns. The within-species gene trees are rough depictions based on the strong and weaker excesses of rare polymorphisms ("star-like" trees) compared to neutral equilibrium expectations in *Dsim* and *Dmel*, respectively. Gray circles indicate approximate positions of the MRCA within each population sample. If a single allele is sampled from a population, a proportion of the changes inferred on the terminal branch will occur at polymorphic sites; the post-MRCA distances reflect the expected numbers of such changes assuming independence among observed polymorphisms. (*B*) Tree topology employed for ancestral inference. Within-species variation was assigned to two sequences (s_{c1} and s_{c2} for *Dsim* and m_{c1} and m_{c2} for *Dmel*, see *Materials and Methods*). Ancestral nodes for within-species sequences are indicated as *s'* for *Dsim* and *m'* for *Dmel*. Ancestral nodes for between-species sequences are labeled *ms* and *ye* for the pairs *Dmel/Dsim* and *Dyak/Dere*, respectively. This topology is assumed for ancestral inference (note that branch lengths are estimated in the process).

Inferring Fixation Biases from *D. simulans and D. melanogaster* Polymorphism. Directional forces impact allele frequencies of mutations segregating within populations. We can infer such forces by examining distributions of allele frequencies among variants segregating in a population, the site frequency spectrum (SFS). Direct comparisons of SFS between classes of mutations that are interspersed within DNA (35, 36) can be a robust statistical approach that is sensitive to weak forces (21, 37) including natural selection and GC-biased gene conversion (gBGC).

SFS comparisons support directional forces favoring G/C over A/T at both twofold and fourfold redundant sites in *D. simulans* (*SI Appendix*, Table S5). The results expand on previous findings that combined data from twofold and fourfold synonymous families for smaller numbers of genes (21, 22) and that employed only fourfold synonymous families in larger datasets (24). These comparisons distinguish between mutational and fixation biases because the former affects the *numbers* of segregating mutations but not their frequencies within the population (assuming that mutation biases remain constant over the relevant time period). The results are also consistent with compositional trend analyses (refs. 13, 17, 38; see also *SI Appendix*, Figs. S11 and S17) that supported G- and/or C-ending major codons in all synonymous families in *D. melanogaster* (compositional trends are similar in *D. simulans*).

Genome-scale data allow us to refine the SFS analyses to a resolution of individual synonymous families. The following analyses will initially focus on twofold redundant sites and the term "synonymous" will refer to this class unless noted otherwise. SFS analyses for fourfold redundant sites in CDS are presented in *SI Appendix*, Fixation biases in fourfold synonymous families. In *D. simulans*, each of the 10 synonymous families (i.e., Phe, Asp, Asn, His, Tyr, Ser_2, Cys, Gln, Glu, and Lys) shows SFS differences similar to the general pattern of elevated within-population sample frequencies of GC-increasing (W → S) compared with AT-increasing (S → W)

mutations (Fig. 2*A* and *SI Appendix*, Table S6). The results are consistent with MCP predictions, but SFS for W → S mutations within short introns (SI) also indicate a GC fixation bias (*SI Appendix*, Table S3; this table also shows data for long introns, LI, >100 bp) as noted in previous studies (24, 39–41). GC fixation biases that are common to synonymous and intron mutations can include biased gene conversion and/or fitness differences unrelated to MCP.

We employ summary statistics to capture the magnitude of differences in SFS comparisons in order to distinguish among underlying directional forces. Glémin et al. (42) estimated a fixation-bias statistic (selection coefficient or conversion parameter scaled to population size) from SFS comparisons between forward and reverse mutations (*Materials and Methods*). We employ "W → S γ" as a measure of the magnitude of SFS differences between W → S and S → W mutations. For some analyses, we employ a difference statistic (W → S aDAF skew, where aDAF refers to average derived allele frequency) as an alternative measure that does not require an assumed neutrally evolving control class of mutations. The two measures are strongly correlated (*SI Appendix*, Fig. S1).

Directional forces favor G and/or C-ending codons across synonymous families in *D. simulans* (Fig. 3*A*) and the magnitude of GC fixation bias is heterogeneous among synonymous families. SI sequences also show GC fixation biases consistent with gBGC and/or natural selection. Larger W → S γ at synonymous sites than at SI sites supports additional directional forces favoring G and/or C-ending codons in coding regions (such as natural selection) as predicted under MCP. Kliman (39) found similar patterns among synonymous families as well as introns in polymorphism data from an autosome in *Drosophila pseudoobscura*.

The numbers of polymorphisms from genes on the X chromosome are sufficient for comparisons to autosomal loci (*SI Appendix*, Table S6) and stronger fixation biases at X-linked loci are a conspicuous feature in *D. simulans* that holds across synonymous families (Fig. 3*A*; Wilcoxon signed-rank test $P = 0.0051$). Potential causes are difficult to resolve, however, and include greater efficacy of natural selection acting on partially recessive fitness effects (43), GC elevation related to dosage compensation (44) or a greater impact of gBGC on the X chromosome.

In contrast to trends that generally conform to expectations under MCP in *D. simulans*, synonymous family-specific analyses reveal remarkable patterns in *D. melanogaster* polymorphism. Several synonymous families show SFS supporting fixation bias in the opposite direction to that observed in *D. simulans*: S → W mutations are segregating at higher frequencies than W → S mutations. The pattern holds for the four NAY synonymous families (codons encoding Asp, Asn, His, and Tyr) but is not found outside of these families (Fig. 3*B* and *SI Appendix*, Table S7). The Lys synonymous family (AAR) shows SFS supporting GC fixation bias; W → S mutations segregate at higher frequencies than S → W mutations but the magnitude of fixation bias is small (Fig. 2*B* and *SI Appendix*, Table S7). All other twofold synonymous families show indistinguishable SFS at both autosomal and X-linked loci (*SI Appendix*, Table S7) consistent with previous analyses of pooled data (17).

D. melanogaster polymorphisms reveal differences in evolutionary processes for X-linked vs. autosomal loci, an "X effect," for NAY families. Fig. 3*B* shows consistent support, across the four NAY synonymous families, for stronger fixation biases favoring NAT over NAC at X-linked than at autosomal genes. Thus, *D. simulans* and *D. melanogaster* both show elevated fixation biases at X-linked genes, but in opposing directions, i.e., favoring GC and AT, respectively. AT preference at NAY codons in *D. melanogaster* may reflect partially recessive fitness effects and/or other forces that override fixation biases toward GC (i.e., gBGC or dosage compensation constraints). The cause(s) of X chromosome vs. autosome differences in GC fixation biases in *D. simulans* are ambiguous as discussed above. However, the X effect for NAT fixation bias in *D. melanogaster* is difficult to explain in the absence of lineage-specific fitness benefits for NAT over NAC codons (interestingly, LI sites show an X effect in the opposite direction, GC preference. See *SI Appendix*, Fixation bias tests: mutation classes at intron sites). Mutation-driven scenarios require highly specified changes in mutational patterns (i.e., specific to NAY codons, greater on the X chromosome, and timed near the MRCA of the species polymorphism) to explain NAY polymorphism patterns in *D. melanogaster*.

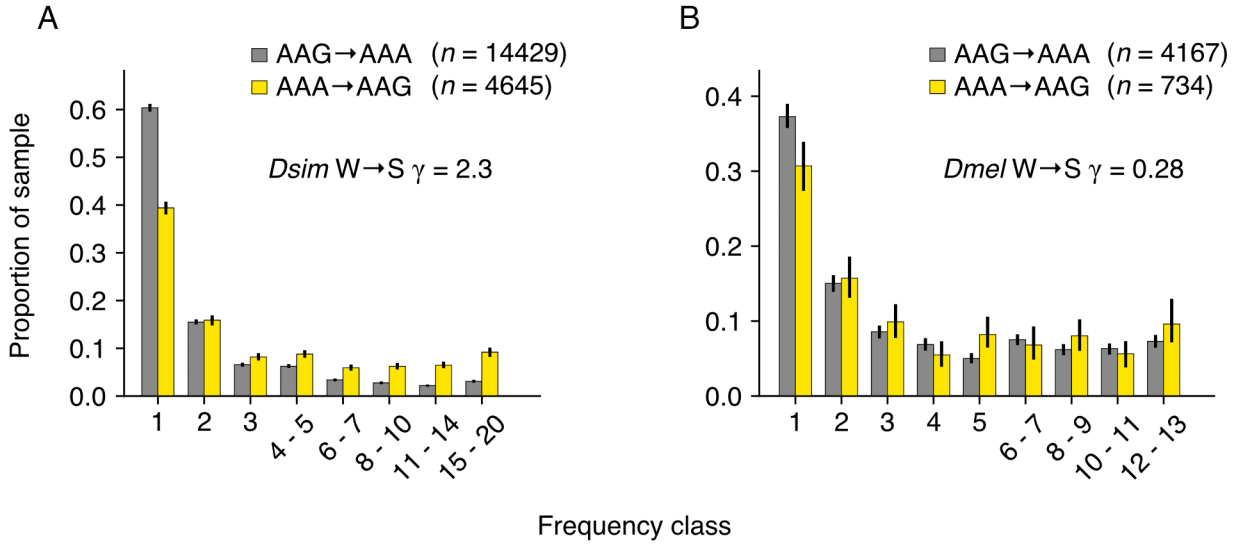

Fig. 2. Site frequency spectra and fixation bias estimates: an example from lysine codons in *D. simulans* and *D. melanogaster*. SFS for AAG → AAA (gray) is compared to that for AAA → AAG (yellow). (*A*) SFS for *D. simulans* (*Dsim*). (*B*) SFS for *D. melanogaster* (*Dmel*). See *SI Appendix*, Tables S5 and S6 for the results of statistical tests for *Dsim* and *Dmel*, respectively. Counts (*n*) for the two polymorphism classes are shown (values are rounded to integers). Data from some frequency classes are pooled as indicated on the x-axis labels. "W → S γ" indicate maximum likelihood estimates of G- and C-favoring fixation biases using the approach of Glémin et al. (42). SFS for intron GC-conservative changes are employed as a putatively neutral reference for the W → S γ estimation. Independent ancestral inference was performed for bootstrap replicates. Error bars indicate 95% CIs among 300 bootstrap replicates.

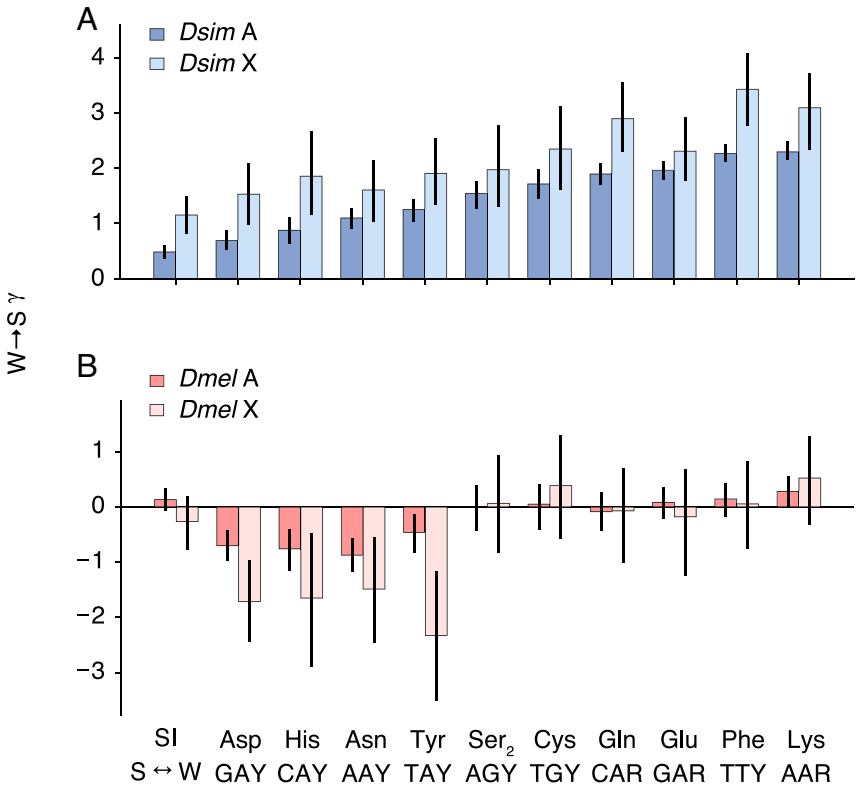

Fig. 3. Synonymous family-specific fixation biases inferred from *D. simulans* and *D. melanogaster* site frequency spectra. Fixation biases for GC-altering changes, W → S γ, estimated using SFS data are shown for (*A*) *D. simulans* (*Dsim*) and (*B*) *D. melanogaster* (*Dmel*). Negative W → S γ values indicate fixation bias favoring AT. "SI" refers to short introns. Autosomal (darker colors; "A") and X-linked (lighter colors; "X") loci are analyzed separately. SFS for GC-conservative changes within SI from autosomal and X-linked loci are employed as required neutral references for W → S γ estimation for the corresponding coding region data. Synonymous families are arranged in the order of γ values for *Dsim* autosomal loci. W → S γ values are presented in *SI Appendix*, Tables S5 and S6, for *Dsim* and *Dmel*, respectively. The x-axis labels are shared between the *Top* and *Bottom* panels. Independent ancestral inference was performed for each bootstrap replicate. Error bars indicate 95% CIs among 300 bootstrap replicates.

The analyses above support genome-wide differences in the direction of fixation biases at NAY codons in *D. melanogaster* and *D. simulans* but do not address the location of the preference shift within the evolutionary tree. Compositional trend analyses indicate NAC preference at all NAY families in species within the *melanogaster* group as well as in the more distantly related *obscura* group (refs. 21, 38; see also *SI Appendix*, Figs. S10–S17). SFS analyses further support NAC preference in *D. pseudoobscura* (39). The fixation bias patterns discussed above are most simply explained by a reversal of codon preference specific to NAY families in the *D. melanogaster* lineage after its split with *D. simulans*.

Fixation Biases in the Ancestral Lineages of *D. simulans* and *D. melanogaster*. The SFS analyses above focused on polymorphisms (i.e., mutations segregating within populations) in *D. simulans* and *D. melanogaster* and revealed strong evidence for AT fixation biases specific to NAY codons in *D. melanogaster*. We also examined patterns deeper in the evolutionary histories of these species. Fixations on the *D. simulans* ancestral (*ms-s'*) and *D. melanogaster* ancestral (*ms-m'*) branches can reveal changes in mutation and/or fixation biases in the relatively short lineages prior to the *s'* and *m'* nodes (Fig. 1*B*). Previous analyses of pooled synonymous families showed weak AT content increases in the *D. simulans* ancestral lineage (24, 45–48) and stronger AT content increases in the *D. melanogaster* ancestral lineage consistent with reduced GC fixation biases (17, 24, 46–50).

We employ a summary statistic to capture the direction and extent of departures from GC content equilibrium: W → S fixation skew, $(a - b)/(a + b)$, where $a = N_{W\to S}$, the number of W → S fixations and $b = N_{S\to W}$, the number of S → W fixations. Expected W → S fixation skew is zero at steady-state GC content and differs in sign, but is scaled symmetrically, for GC-increasing and GC-decreasing departures from equilibrium.

GC content decline is prevalent in the ancestral *D. simulans* lineage among mutation classes. However, W → S fixation skew is *positive* for our presumed best candidates for neutral evolution, SI with the lowest GC content (Fig. 4*A* and *SI Appendix*, Fig. S4). We attribute GC elevation at putatively neutrally evolving sites to an increase in mutational pressure toward GC. From coding regions, we analyzed synonymous changes within NAY and "non-NAY" families (synonymous families encoding Phe, Cys, Ser_2, Lys, Gln, and Glu) separately. For all three mutation classes (i.e., SI, non-NAY, and NAY), loss of GC intensifies with ancestral GC content (Fig. 4*A* and *SI Appendix*, Fig. S6*C*). Such near-linear trends are consistent with simple MCP scenarios of nonstationary mutation ratio and/or fixation bias (51). In particular, the *D. simulans* W → S fixation skew patterns appear roughly consistent with a combination of increase in GC mutation pressure and reduction in GC fixation bias (*SI Appendix*, Fig. S6; note the contrasting patterns for LI, however).

We tested whether GC content changes are consistent with shared parameter fluctuations among SI, non-NAY, and NAY mutation classes. Data were partitioned so that comparisons are conducted within similar ancestral GC-content ranges (*SI Appendix*, Table S11). SI and NAY classes show indistinguishable W → S fixation skews in *D. simulans* (*SI Appendix*, Table S11). Non-NAY codons show a weak signal toward less AT-biased fixation skews than SI and NAY codons (*SI Appendix*, Table S11), but low fixation counts on the *ms-s'* branch (Fig. 1) prevent firm conclusions.

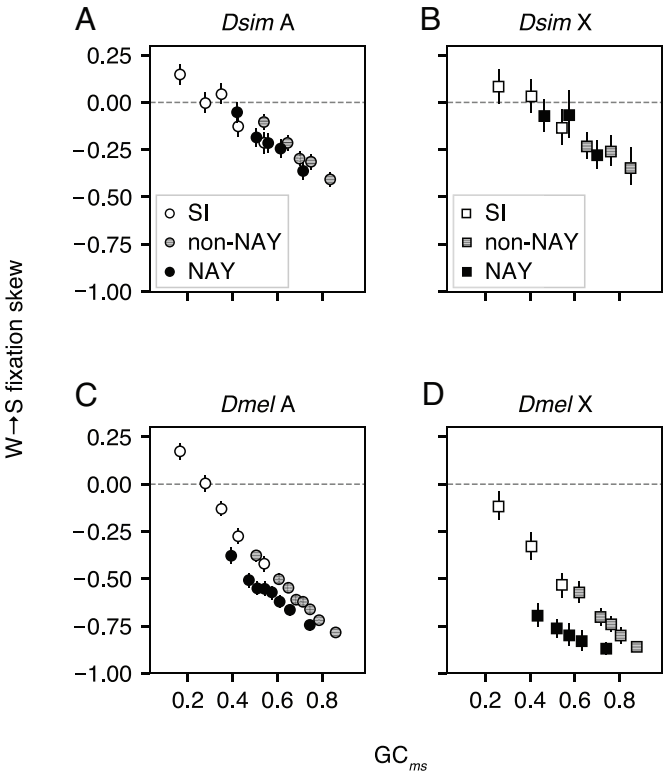

Fig. 4. GC content changes in the *D. simulans* and *D. melanogaster* ancestral lineages. W → S fixation skew = $(N_{W\to S} - N_{S\to W})/(N_{W\to S} + N_{S\to W})$, where $N_{W\to S}$ is W → S fixation count and $N_{S\to W}$ is S → W fixation count. This index measures the direction and magnitude of the departure from GC content equilibrium. Changes in GC content are plotted against GC content at *ms* node, GC_{ms}. Fixations are inferred within internal branches: (*A*) and (*B*) *ms-s'* for *D. simulans* fixation data and (*C*) and (*D*) *ms-m'* for *D. melanogaster* fixation data (Fig. 1*B*). (*A*) and (*C*) Autosomal and (*B*) and (*D*) X-linked loci are analyzed separately. The legends in (*A*) and (*B*) apply to (*C*) and (*D*), respectively. "SI" indicates short introns. Synonymous families are pooled into two classes: non-NAY (Gln, Glu, Lys, Phe, Cys, and Ser_2) and NAY (Asp, Asn, His, and Tyr). For binning, introns or CDS are ranked by GC_{ms} for the mutation class and assigned to bins with similar numbers of intron sites or codons, respectively. Statistical analyses are presented in *SI Appendix*, Table S12. W → S fixation skews are compared between autosomal and X-linked loci in *SI Appendix*, Fig. S7. Error bars indicate 95% CIs among 1,000 bootstrap replicates.

The ancestral lineage for *D. melanogaster* (Fig. 1) shows strong negative relationships between GC change and ancestral GC content for SI, non-NAY, and NAY classes. The slopes of these trends are steeper than in *D. simulans* (Fig. 4). Low-GC SI, our assumed neutrally evolving class, show an elevation of GC similar to the pattern in *D. simulans* (Fig. 4*B*) consistent with a shift in GC-elevating mutation bias that predates the *ms* node. Stronger negative W → S fixation skew slopes can be explained by larger fixation bias reductions (smaller $N_e s$) in the ancestral *D. melanogaster* lineage (51). Notably, NAY codons show accelerated AT content shifts compared to non-NAY codons within this lineage (*SI Appendix*, Table S11). Extents of GC content change were statistically indistinguishable between SI sites and non-NAY codons (*SI Appendix*, Table S11). Such patterns are predicted under a combination of fixation biases *favoring* S → W mutations at NAY codons *and* reduced GC fixation biases at SI sites and non-NAY codons.

X effects are a compelling feature in the SFS analyses (see above) and W → S fixation skew comparisons between X-linked vs. autosomal genes reveal similarly notable patterns. We focus on *D. melanogaster* because the low number of fixations on the *D. simulans* lineage limits the exploration of X effects (findings are discussed in *SI Appendix*). In *D. melanogaster*, non-NAY, NAY, and fourfold synonymous classes as well as SI mutation classes show small, but statistically significant, elevations of AT gains at X-linked loci compared to autosomal loci (*SI Appendix*, Fig. S7 *E–H* and Table S12). Notably, NAY codons show significantly greater AT gains at X-linked, than at autosomal loci, than other mutation classes (*SI Appendix*, Fig. S7*H* and Table S13). Enhanced efficacy of NAT codon preference at X-linked, relative to autosomal, loci for both polymorphism (post-*m'* node) and fixations (*ms-m'* nodes) in the *D. melanogaster* lineage provides a simple explanation for this pattern and is consistent with partially recessive fitness advantages for NAC to NAT mutations.

tRNA Modification: A Potential Factor in *D. melanogaster* Preference Reversals at NAY Codons. Shared properties of NAY codons raise several possibilities for the biological cause(s) of the reversal of fitness benefits of NAT over NAC codons in the *D. melanogaster* lineage. These codons have common nucleotide composition at the 2nd and 3rd codon positions, show relatively weak fixation biases in *D. simulans* in SFS analyses (Fig. 3*A*), and are recognized by cognate tRNAs that undergo a particular chemical modification: queuosine (Q) modification at the anticodon 1st (wobble) position G nucleoside. Q modification is observed in the majority of eukaryotes (52–55) and is not known to occur for cognate tRNAs for other synonymous families (56). Many eubacteria can synthesize the Q precursor, queine, but eukaryotes lack the required biochemical pathways and sequester the micronutrient from their diet and/or gut microbiota. Because queine levels are strongly dependent on diet in *Drosophila* (29, 57, 58), lineage-specific modification levels (related to nutrient availability) have been proposed to explain codon preference differences (29, 59).

Fitness benefits of Q modification are clear from the retention of this pathway in a wide range of eukaryotes, but the functional basis of the benefit remains difficult to determine and may differ among taxa (reviewed in ref. 56). Such benefits could lie outside of impacts on translational elongation/accuracy; for example, tRNA-modification affects rates of breakdown of isoacceptors into tRNA-derived small RNAs that function in gene regulation and stress response (60, 61). Experimental studies of protein synthesis support similar translation rates at NAC and NAT codons for unmodified cognate tRNAs and higher translational elongation rates at NAC codons for Q-modified tRNAs (62). Changes in Q modification can have measurable effects on translation rates at both cognate and noncognate codons and tissue- and sex-specific effects in mice (63). Importantly, compositional trend analyses consistently support translational preference of NAC over NAT in organisms that employ Q modification [*Escherichia coli* (18), *Bacillus subtilis* (18), *Saccharomyces pombe* (64), and *Caenorhabditis elegans* (16, 65)] as well as in those that lack the modification [*Saccharomyces cerevisiae* (66), *Candida albicans* (15), and *Arabidopsis thaliana* (65, 67)]. These patterns argue against a simple relationship between Q modification levels and major codon identity; in the limited number of established cases of loss of Q modification, translation selection appears to favor NAC codons.

We tested for signals for major codon reassignment (reversals in translational advantages) as a cause of NAT preference in *D. melanogaster*. Greater fitness benefits in highly translated genes are a hallmark feature of MCP (10, 68). We expect elevated fixation biases for minor to major codon mutations among genes that show the strongest patterns of ancestral MCP. In *D. simulans*, both NAY and non-NAY codons show greater GC fixation biases in genes with higher major codon usage at the *ms* node (*SI Appendix*, Fig. S20). If NAT preference in *D. melanogaster* is driven by translational selection, then we expect similarly larger absolute

magnitudes of fixation bias in genes under stronger translational selection. We detected no such trend; magnitudes of fixation bias are indistinguishable between high and low major codon usage genes (*SI Appendix*, Fig. S20). We consider potential advantages of NAT codon usage that fall outside of the standard MCP model.

Testing ApT vs. ApC Dinucleotide Preference. Context-dependent fixation biases can arise from dinucleotide preferences that act in opposition to MCP-related GC fixation biases. ApT vs. ApC dinucleotide preferences could, in principle, account for NAY family-specific polymorphism and fixation patterns and may show less association with gene expression level than MCP. Interactions between dinucleotide preferences and translational selection have been proposed in previous *Drosophila* studies (69, 70). We take advantage of features within the genetic code to distinguish specific codon and dinucleotide preference scenarios. Leu is encoded by six codons and, T → C transitions at the first position are synonymous among YTR codons. ApT dinucleotide preference predicts context-dependent fixation biases toward TTR codons for Leu codons downstream of a 5' A (A|TTR vs. A|CTR where pipe symbols indicate codon boundaries). We discuss *D. melanogaster* patterns before comparisons to *D. simulans* in order to assess causes of NAT preference. Fig. 5*A* confirms the predicted signal within *D. melanogaster* polymorphism; SFS is skewed toward elevated frequencies for C → T compared to T → C transitions among A|TTR ↔ A|CTR synonymous polymorphism. The magnitude of the fixation bias (aDAF skew is strongly correlated with γ; *SI Appendix*, Fig. S1) appears roughly similar to NAT preference at NAY codons. This pattern is consistent with ApT over ApC dinucleotide preference but could also simply reflect translational selection favoring TTR over CTR Leu codons. We distinguish these possibilities by examining context (5' nearest neighbor) effects for synonymous changes at Leu first codon positions within *D. melanogaster* polymorphism. Differences in SFS for T → C and reverse changes are specific to the 5' A context; B|TTR ↔ B|CTR synonymous polymorphism show similar aDAF (ambiguity character B denotes C, T, or G nucleotides; Fig. 5*A*). Permutation tests support context-dependent T → C fixation bias; W → S aDAF skew is lower for A|TTR → A|CTR than for B|TTR → B|CTR ($P < 0.005$, 10^5 iterations). These patterns together support ApT over ApC dinucleotide preference and, because Leu cognate tRNAs do not undergo queuosine modification, call into question the tRNA modification-driven codon preference reversal scenario.

We examined first codon position synonymous changes at Leu codons to test for ApT vs. ApC preference within *D. simulans* polymorphism. Fig. 5*B* illustrates context-dependent SFS differences among T → C and reverse synonymous changes at Leu first positions. SFS differences indicate a relatively strong GC fixation bias among B|TTR ↔ B|CTR synonymous polymorphism. However, the same nucleotide transition in a 5' A context shows more similar SFS; permutation tests support lower W → S aDAF skew for A|TTR → A|CTR than for B|TTR → B|CTR ($P < 0.0001$, 10^5 iterations). Viewed in isolation, this difference could be interpreted most simply as support for context-dependent weakening of first codon position C preference (i.e., reduced MCP) at Leu codons. However, support for fixation biases favoring A|TTR over A|CTR in *D. melanogaster* (Fig. 5*A*) lead us to posit that this reduction in the magnitude of aDAF skew in *D. simulans* (Fig. 5*B*) reflects a common ApT over ApC preference acting in opposition to a stronger CTR-favoring force (presumably MCP) than in *D. melanogaster*. We employ similar reasoning in the analyses below.

We tested for dinucleotide preferences for ApT vs. GpT, the reverse complement (RC) of ApT vs. ApC within coding regions. We consider synonymous A ↔ G transitions in the context of 3' neighboring T vs. V (C, A, or G nucleotides). Overall trends toward aDAF skew differences expected under ApT over GpT preference are strong among 13 synonymous families (*D. melanogaster*, $P < 0.0008$; *D. simulans*, $P < 0.0003$; *SI Appendix*, Fig. S8 and Tables S14 and S15). We propose a scenario of MCP-related GC pressure balanced by an opposing ApT over ApC pressure at NAY and Leu codons in both *D. melanogaster* and *D. simulans*. Synonymous changes within a larger number of families support

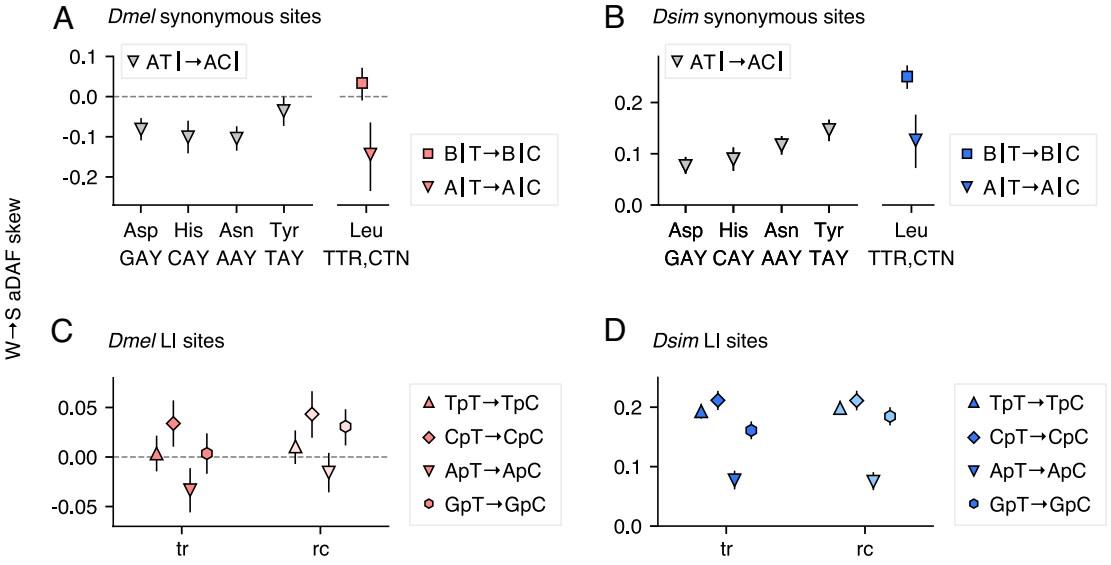

Fig. 5. ApT vs. ApC dinucleotide preference inferred from *D. melanogaster* and *D. simulans* site frequency spectra. W → S aDAF skew measures the direction and magnitude of SFS difference between "forward" and "reverse" context-dependent nucleotide changes. Arrows specify the forward direction. Positive W → S aDAF skew indicates higher frequencies for forward changes than reverse changes. W → S aDAF skews for ApT → ApC are compared to those for BpT → BpC ("B" indicates "non-A" or T, C, or G nucleotides). ApT → ApC and BpT → BpC changes at synonymous sites in (*A*) *D. melanogaster* (*Dmel*) and (*B*) *D. simulans* (*Dsim*). Pipes ("|") indicate codon borders. ApT → ApC and other context-dependent nucleotide changes at long intron (LI) sites in (*C*) *Dmel* and (*D*) *Dsim*. Context-dependent nucleotide changes on the transcribed (tr) strand are indicated in the in-figure legends. Light colors indicate reverse complement (RC) changes (e.g., ApA → GpA is the RC of TpT → TpC). Error bars indicate 95% CIs among 1,000 bootstrap replicates. Ancestral reconstructions are resampled in units of introns or CDS.

this dinucleotide preference on the nontranscribed strand. Reduced efficacy of MCP reveals the action of this dinucleotide preference at *D. melanogaster* NAY codons but nearest-neighbor-dependent SFS in other synonymous families reveal similar forces, acting with different relative intensities, in *D. simulans*.

We expanded the data for analyses of ApT vs. ApC preference by including polymorphisms within LI. Although constraints within LI are not well understood (71), larger numbers of segregating sites in these data (*SI Appendix*, Tables S16 and S17) allow us to examine fixation biases in specific contexts. In *D. melanogaster*, aDAF skew indicates T over C preference in a 5' A context which is absent, or acting in the opposing direction, for other 5' contexts (Fig. 5*C* and *SI Appendix*, Table S16). CI are large in the *D. melanogaster* data, but context-dependent SFS differences appear to be similar on transcribed and nontranscribed strands (Fig. 5*C*). Higher polymorphism levels in the *D. simulans* data help to reveal strong context dependence in T → C fixation biases. BpT → BpC mutations are strongly favored but this W → S fixation bias is clearly reduced for ApT → ApC changes (Fig. 5*D*). These contrasting patterns of SFS are similar for RC changes. SFS analyses of polymorphisms within LI support ApT over ApC dinucleotide preference in noncoding regions and confirm that evolutionary forces act similarly in the transcribed and nontranscribed strands.

These features of ApT over ApC preference may be important in considering the biological function(s) underlying dinucleotide preferences. Dinucleotides can impact biophysical properties of DNA such as bendability that can affect chromatin structure, DNA repair, and protein binding (69), reviewed in (72). Highly ApT-rich regions can be prone to chromosomal mutations (73). Dinucleotides can influence secondary structure formation in mRNAs (74) which can affect the translation process (75). However, the similarity of preferences in coding and noncoding regions and on transcribed and nontranscribed strands is more consistent with DNA constraints. Langley and coworkers (76) found enrichment of ApA, TpT, and GpC dinucleotides in *D. melanogaster* intergenic and intron regions. These dinucleotides occur in (roughly 10 bp) periodic patterns consistent with nucleosome positioning constraints. Population genetic analyses support fixation biases underlying these patterns. However, ApT dinucleotides are not overrepresented in their analyses (76) or in previous compositional analyses within coding regions (69, 77), but see ref. 78. The functional bases for ApT over ApC dinucleotide preferences remain unclear.

Genome Dynamics Under Near Neutrality. Sensitivity of rates of evolution and levels of adaptation to effective population size is central features of nearly neutral evolution. N_e variation on the time-scale of evolutionary processes (and corresponding modulations in selection intensity, $N_e s$) are often invoked to explain patterns in genome variation (e.g., refs. 47, 79, 80). Our findings challenge whether N_e fluctuations are sufficient to account for evolutionary dynamics across mutation classes in either the *D. melanogaster* or the *D. simulans* lineage.

In *D. melanogaster*, reduced historical N_e could explain patterns of declining GC usage among classes of synonymous sites in coding regions (non-NAY and fourfold synonymous families), SI, and LI (*SI Appendix*, Fig. S6*D*). However, the same factor should also have attenuated the efficacy of selection for preferred dinucleotides. Changes in the relative fitness effects (i.e., an increase in selection coefficients favoring ApT vs. ApC dinucleotides relative to those favoring major codons and intron GC-elevation) appear necessary to account for our observations.

In *D. simulans*, proportionate changes in N_e are consistent with declines in GC among several classes of synonymous change (fourfold and NAY synonymous families; *SI Appendix*, Fig. S6*C*) as well as SI. However, LI show a distinct pattern of GC content stability across GC classes in the ancestral lineage; this suggests an elevation in selection coefficients favoring GC at LI sites relative to similar forces at SI and synonymous sites. Our findings are difficult to explain in the absence of changes in the relative magnitudes of fitness effects among mutation classes in both focal evolutionary lineages. More generally, our findings indicate that pleiotropy plays an important role in synonymous site and noncoding DNA evolution and that small fitness effects may not reflect weak functional effects; in particular, smaller fixation biases at NAY codons compared to other synonymous families do not necessarily indicate smaller translational effects of synonymous changes in these families.

Our study demonstrates how increasing the resolution (including codon-family specificity and short time scales) of molecular evolutionary studies can reveal context effects and interactions among opposing forces in genome adaptation. Between-species contrasts of polymorphism and fixation patterns were critical for identifying dinucleotide preferences that we hope will motivate functional genomic analyses.

Materials and Methods

DNA Sequences. We employed available genome data for isofemale lines established from natural populations of *D. melanogaster* and *D. simulans*. Data from a Rwanda population of *D. melanogaster* (31) were downloaded from http://www.dpgp.org/dpgp2/DPGP2.html (last downloaded on 22nd July 2014). Sequence analyses support a relatively stable effective population size (81) and lower inversion frequencies (81, 82) for this population compared to other populations for which population genomic data are available. Among 22 lines reported in the Pool *et al*. study, 14 lines were employed for this analysis (RG2, RG3, RG5, RG9, RG18N, RG19, RG22, RG24, RG25, RG28, RG32N, RG34, RG36, and RG38N). We excluded five lines that show evidence for a high proportion of admixture with European populations [RG10, RG11N, RG15, RG21N, and RG35; (31)]. In addition, we excluded three lines that contain relatively high proportions of ambiguous nucleotides (RG4N, RG7, and RG33). We extracted predicted CDS and intron sequences from the genomic sequences for a total of 13,691 protein-coding genes using FlyBase r6.24 *D. melanogaster* genome annotations (ftp://ftp.flybase.net/genomes/Drosophila_melanogaster/dmel_r6.24_FB2018_05/; last downloaded on 12th July 2019; *SI Appendix*, *Methods*).

We reconstructed genome sequences for *D. simulans* within-species samples using available data (https://drive.google.com/drive/folders/0B4O-acc8EJwheS1HZ1hnWkpOQlE; last downloaded on 1st March 2017; *SI Appendix*, *Methods*). We analyzed the DNA variant information for 21 lines from a Madagascar population (24, 30); 10 lines (MD06, MD105, MD106, MD15, MD199, MD221, MD233, MD251, MD63, and MD73) were sequenced by Rogers et al. (30), and 11 lines (MD03, MD146, MD197, MD201, MD224, MD225, MD235, MD238, MD243, MD255, and MD72) were sequenced by Jackson et al. (24). These sequences show high nucleotide diversity relative to data from other population samples (83). These 10 and 11 lines were sampled from the same localities in Madagascar by Rogers et al. (30) and William Ballard, respectively, as described by Jackson et al. (24). We extracted predicted CDS and intron sequences from the genome sequences for a total of 13,831 protein-coding genes using FlyBase r2.02 *D. simulans* genome annotations (ftp://ftp.flybase.net/genomes/Drosophila_simulans/dsim_r2.02_FB2017_04; last downloaded on 9th March 2020).

Genome sequences from species within the *D. melanogaster* subgroup, *D. yakuba* and *D. erecta* were employed as outgroup data. We extracted predicted CDS and intron sequences from the genome sequences using genome annotations for FlyBase r1.05 *D. yakuba* (ftp://ftp.flybase.net/genomes/Drosophila_yakuba/dyak_r1.05_FB2016_05/; last downloaded on 12th July 2019) and FlyBase r1.05 *D. erecta* (ftp://ftp.flybase.net/genomes/Drosophila_erecta/dere_r1.05_FB2016_05/; last downloaded on 12th July 2019).

We constructed CDS and intron sequence alignments including *D. melanogaster* and *D. simulans* within-species samples and outgroup sequences (*SI Appendix, Methods*). Custom Python codes used in the alignment pipeline are available at https://github.com/nigevogen. The dataset includes 9,361 CDS and 26,534 (17,244 for SI and 9,290 for LI) intron alignments for autosomal loci and 1,613 CDS and 4,305 (2,564 for SI and 1,741 for LI) intron alignments for X-linked loci which are available on Zenodo (DOI: 10.5281/zenodo.15274324).

Inference of Polymorphic Changes. We define "synonymous family" as a group of synonymous codons that can be interchanged in single nucleotide steps; serine coding codons were split into a twofold (AGY, referred to as Ser_2) and a fourfold (TCN, referred to as Ser_4) family. We analyzed 10 synonymous families (Phe, Asp, Asn, His, Tyr, Ser_2, Cys, Gln, Glu, and Lys) with twofold redundancy and six families (Ala, Gly, Val, Thr, Pro, and Ser_4) with fourfold redundancy.

We estimated probabilities of nucleotides at ancestral nodes in the gene tree shown in Fig. 1*B*. Although the sequences examined are relatively closely related, ancestral inference under simple substitution models such as maximum parsimony can be unreliable when character states are biased and/or composition is changing within the gene tree (34, 84). In addition, our analyses require inference of ancestral and derived states at segregating sites in recombining regions where gene trees may differ among sites. In order to address these issues, we employed a likelihood-based approach that incorporates uncertainty in ancestral inference. We inferred ancestral nucleotides for both within- and between-species variation using a combination of BASEML (85) and the Bifurcating Tree with Weighting (BTW) approach (34). For BASEML ancestral inference, we employed the GTR-NH_b nucleotide substitution model (33, 86) with a newly implemented option (available on BASEML in PAML ver4.9) that allows user-defined branches to share transition parameters. Here, we set parameters to be shared within (but not between) collapsed sequence pairs in *D. melanogaster* and *D. simulans* (terminal branches leading to m_{c1}, m_{c2} and s_{c1}, s_{c2}, respectively, in Fig. 1*B*). Ancestral inference was conducted separately for data from autosomal and X-linked loci. We refer to inferred changes on the m'-m_{c1} and m'-m_{c2} branches as polymorphisms in *D. melanogaster* and those on the s'-s_{c1} and s'-s_{c2} branches as polymorphisms in *D. simulans*. We treated probabilities of changes as counts for the numbers of polymorphic mutations for each of 12 mutation classes (*SI Appendix, Methods*). Ancestral inference for context dependence analysis is also described in the *SI Appendix*.

Polymorphism Analysis. We analyzed SFS for forward and reverse mutations between pairs of nucleotides (e.g., A → G vs. G → A). We use the term "forward" mutations for W → S changes (where W indicates A or T, and S indicates G or C) among GC-altering mutations and for T → A and C → G among GC-conservative mutations. Mutations in the opposite direction are termed "reverse" mutations (designations of the terms "forward" and "reverse" are arbitrary). Among the six possible pairs, four are GC-altering (i.e., W ↔ S) and two are GC-conservative (i.e., A ↔ T and G ↔ C). Mann–Whitney U (MWU) tests were employed to test for SFS differences among synonymous and intron mutations. Direct SFS comparisons between mutation classes that are physically interspersed within DNA (35, 36) attempt to control for effects of linked selection and demographic history in the inference of fixation biases. This approach can be employed to test weak selection models of synonymous codon usage bias (21, 37).

We estimated the magnitude of fixation biases using a maximum likelihood method (40) that fits observed SFS to theoretical expectations (87, 88). SFS of putatively neutral mutations (here, intron GC-conservative mutations; see also *SI Appendix*, Table S3) were employed to adjust for possible departures from steady-state SFS caused by demographic history and linked selection (89). We employed the M1 model in the anavar software package (42, 90) to estimate W → S γ, the fixation bias statistic. Positive and negative values of W → S γ indicate fixation biases that elevate and reduce GC content, respectively. This statistic is an estimate of the product of $4N_e$ and either the selection coefficient (s), or the intensity of the conversion bias (b) in selection and biased gene conversion models, respectively. In our analyses, γ estimates are strongly correlated with simple difference statistics between average derived allele frequencies (aDAF skew; *SI Appendix*, Fig. S1). We employ these indices as summary statistics for the magnitude of difference between SFS in comparisons across mutation classes and X-linked vs. autosomal loci.

Inference of Fixations in Coding Regions. We inferred synonymous and replacement changes within the *D. simulans* ancestral (ms-s') and *D. melanogaster* ancestral (ms-m') lineages by combining an approach introduced by Akashi et al. (47) with the method described above (BTW using the GTR-NH_b model). This approach reconstructs internal node codon configurations (INCC) by combining separate estimations of internal node nucleotide configurations (INNC) for three codon positions, assuming independent processes among positions (47).

We employed the alignments described in *SI Appendix, Methods*. We inferred ancestral codons at internal nodes in the tree shown in Fig. 1*B*. A pair of ancestral and derived codons that differ at a single position will be referred to as a "codon change" (e.g., AAA at the ms node and AAG at the m' node at a given codon position in the aligned CDS is an AAA → AAG codon change in the ms-m' branch). The probability of INCC is used as a "count" of the change. For the case of codon pairs that differ at two or three positions, we calculated the relative probability for each possible minimal step path from ancestral and derived codon. These probabilities were weighted by the numbers of synonymous and nonsynonymous changes involved in a path (91) using the nonsynonymous-to-synonymous substitution rate ratio calculated from codon pairs that differ at a single position ($d_N/d_S = 0.1234$ for autosomal loci; $d_N/d_S = 0.1434$ for X-linked loci). We used the Nei and Gojobori (91) method to count the numbers of synonymous and nonsynonymous sites.

In our analyses, "fixations" refer to changes inferred on the ms-s' and ms-m' branches and are thus "fixed in the sample" and may include mutations segregating at high frequencies within the populations (these changes do not overlap with data for the SFS analysis). We also note that our approach of fixation counting does not consider multiple changes at a site within a branch. This can cause underestimation of fixation counts, but the effect is expected to be minor given the low divergence levels between *D. simulans* and *D. melanogaster* (33, 34).

W → S Fixation Skew Analysis. To infer changes in fixation bias and/or mutation rate bias, we compared $d_{WS, SW}$ among genes that differ in GC content at the ms node (GC_{ms}). GC_{ms} is the sum of probabilities for INNC with G or C nucleotides at the ms node divided by the number of "sites" (*i.e.*, the total probabilities for INNC) for introns and serves as a proxy for ancestral GC fixation bias. We used a similar approach to calculate GC_{ms} at synonymous sites, the proportion of G- and/or C-ending codons at the ms node. GC_{ms} are used for binning introns and CDS into nonoverlapping bins with similar numbers of sites (nucleotide positions for intron and codons for CDS). We filtered introns and CDS that include fewer than 10 aligned sites. Ancestral reconstructions were resampled in units of intron and CDS. Statistical approaches to detect differences in W → S fixation skews between mutation classes are described in *SI Appendix*, GC content evolution in the *D. simulans* and *D. melanogaster* ancestral lineages.

Data, Materials, and Software Availability. Python codes and analysis data have been deposited in GitHub (https://github.com/nigevogen) and Zenodo (https://zenodo.org/communities/evogen-akashi-lab/), respectively. Previously published data were used for this work (24, 30–32).

ACKNOWLEDGMENTS. We are grateful to Tomoko Ohta, Takahiro Sakamoto, two anonymous reviewers, and the handling editor for their comments and advice that helped to considerably improve this manuscript. We also thank Hiroko Mochizuki for technical assistance. Portions of the paper were developed from the PhD thesis of H.Y.

Author affiliations: [a]Laboratory of Evolutionary Genetics, Department of Genomics and Evolutionary Biology, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan; [b]Department of Genetics, The Graduate University for Advanced Studies, SOKENDAI, Mishima, Shizuoka 411-8540, Japan; and [c]Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, United Kingdom

1. G. E. Andersson, C. G. Kurland, An extreme codon preference strategy: Codon reassignment. *Mol. Biol. Evol.* **8**, 530–544 (1991).
2. H. Akashi, Gene expression and molecular evolution. *Curr. Opin. Genet. Dev.* **11**, 660–666 (2001).
3. L. Duret, Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* **12**, 640–649 (2002).
4. T. Ikemura, Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* **2**, 13–34 (1985).
5. S. Varenne, J. Buc, R. Lloubes, C. Lazdunski, Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J. Mol. Biol.* **180**, 549–576 (1984).

6. J. F. Curran, M. Yarus, Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. *J. Mol. Biol.* **209**, 65–77 (1989).
7. A. Dana, T. Tuller, The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res.* **42**, 9171–9181 (2014).
8. J. Precup, J. Parker, Missense misreading of asparagine codons as a function of codon identity and context. *J. Biol. Chem.* **262**, 11351–11355 (1987).
9. E. B. Kramer, P. J. Farabaugh, The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA* **13**, 87–96 (2007).
10. T. Ikemura, Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* **151**, 389–409 (1981).
11. T. Ikemura, Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes: Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. *J. Mol. Biol.* **158**, 573–597 (1982).
12. R. Grantham, C. Gautier, M. Gouy, M. Jacobzone, R. Mercier, Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res.* **9**, r43–r74 (1981).
13. D. C. Shields, P. M. Sharp, D. G. Higgins, F. Wright, Silent sites in *Drosophila* genes are not neutral: Evidence of selection among synonymous codons. *Mol. Biol. Evol.* **5**, 704–716 (1988).
14. P. M. Sharp, K. M. Devine, Codon usage and gene expression level in *Dictyosteiium discoidtum*: Highly expressed genes do "prefer" optimal codons. *Nucleic Acids Res.* **17**, 5029–5040 (1989).
15. A. T. Lloyd, P. M. Sharp, Evolution of codon usage patterns: The extent and nature of divergence between *Candida albicans* and *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **20**, 5289–5295 (1992).
16. M. Stenico, A. T. Lloyd, P. M. Sharp, Codon usage in *Caenorhabditis elegans*: Delineation of translational selection and mutational biases. *Nucleic Acids Res.* **22**, 2437–2446 (1994).
17. H. Akashi, Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. *Genetics* **139**, 1067–1076 (1995).
18. S. Kanaya, Y. Yamada, Y. Kudo, T. Ikemura, Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: Gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* **238**, 143–155 (1999).
19. W.-H. Li, Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. *J. Mol. Evol.* **24**, 337–345 (1987).
20. M. Bulmer, The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**, 897–907 (1991).
21. H. Akashi, S. W. Schaeffer, Natural selection and the frequency distributions of "silent" DNA polymorphism in *Drosophila*. *Genetics* **146**, 295–307 (1997).
22. R. M. Kliman, Recent selection on synonymous codon usage in *Drosophila*. *J. Mol. Evol.* **49**, 343–351 (1999).
23. P. M. Sharp, L. R. Emery, K. Zeng, Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **365**, 1203–1212 (2010).
24. B. C. Jackson, J. L. Campos, P. R. Haddrill, B. Charlesworth, K. Zeng, Variation in the intensity of selection on codon bias over time causes contrasting patterns of base composition evolution in *Drosophila*. *Genome Biol. Evol.* **9**, 102–123 (2017).
25. G. Marais, Biased gene conversion: Implications for genome and sex evolution. *Trends Genet. TIG* **19**, 330–338 (2003).
26. L. Duret, N. Galtier, Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.* **10**, 285–311 (2009).
27. A. A. Komar, A code within a code: How codons fine-tune protein folding in the cell. *Biochem. Biokhimiia* **86**, 976–991 (2021).
28. J. R. Powell, E. Sezzi, E. N. Moriyama, J. M. Gleason, A. Caccone, Analysis of a shift in codon usage in *Drosophila*. *J. Mol. Evol.* **57**, S214–S225 (2003).
29. J. M. Zaborske *et al.*, A nutrient-driven tRNA modification alters translational fidelity and genome-wide protein coding across an animal genus. *PLoS Biol.* **12**, e1002015 (2014).
30. R. L. Rogers *et al.*, Landscape of standing variation for tandem duplications in *Drosophila yakuba* and *Drosophila simulans*. *Mol. Biol. Evol.* **31**, 1750–1766 (2014).
31. J. E. Pool *et al.*, Population genomics of sub-saharan *Drosophila melanogaster*: African diversity and non-African Admixture. *PLoS Genet.* **8**, e1003080 (2012).
32. Drosophila 12 Genomes Consortium, Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218 (2007).
33. T. Matsumoto, H. Akashi, Z. Yang, Evaluation of ancestral sequence reconstruction methods to infer nonstationary patterns of nucleotide substitution. *Genetics* **200**, 873–890 (2015).
34. T. Matsumoto, H. Akashi, Distinguishing among evolutionary forces acting on genome-wide base composition: Computer simulation analysis of approximate methods for inferring site frequency spectra of derived mutations in recombining regions. *G3* **8**, 1755–1769 (2018).
35. M. G. Bulmer, Protein polymorphism. *Nature* **234**, 410–411 (1971).
36. S. A. Sawyer, D. E. Dykhuizen, D. L. Hartl, Confidence interval for the number of selectively neutral amino acid polymorphisms. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6225–6228 (1987).
37. H. Akashi, Inferring the fitness effects of DNA mutations from polymorphism and divergence data: Statistical power to detect directional selection under stationarity and free recombination. *Genetics* **151**, 221–238 (1999).
38. S. Vicario, E. N. Moriyama, J. R. Powell, Codon usage in twelve species of *Drosophila*. *BMC Evol. Biol.* **7**, 1–17 (2007).
39. R. M. Kliman, Evidence that natural selection on codon usage in *Drosophila pseudoobscura* varies across codons. *G3 Genes Genomes Genet.* **4**, 681–692 (2014).
40. B. Jackson, B. Charlesworth, Evidence for a force favoring GC over AT at short intronic sites in *Drosophila simulans* and *Drosophila melanogaster*. *G3* **11**, jkab240 (2021).
41. B. Yıldırım, C. Vogl, The influence of GC-biased gene conversion on nonadaptive sequence evolution in short introns of *Drosophila melanogaster*. *J. Evol. Biol.* **37**, 383–400 (2024).
42. S. Glémin *et al.*, Quantification of GC-biased gene conversion in the human genome. *Genome Res.* **25**, 1215–1228 (2015).
43. B. Charlesworth, J. A. Coyne, N. H. Barton, The relative rates of evolution of sex chromosomes and autosomes. *Am. Nat.* **130**, 113–146 (1987).
44. A. A. Alekseyenko *et al.*, Sequence-specific targeting of dosage compensation in *Drosophila* favors an active chromatin context. *PLoS Genet.* **8**, e1002646 (2012).
45. D. J. Begun, The frequency distribution of nucleotide variation in *Drosophila simulans*. *Mol. Biol. Evol.* **18**, 1343–1352 (2001).
46. A. D. Kern, D. J. Begun, Patterns of polymorphism and divergence from noncoding sequences of *Drosophila melanogaster* and *D. simulans*: Evidence for nonequilibrium processes. *Mol. Biol. Evol.* **22**, 51–62 (2005).
47. H. Akashi *et al.*, Molecular evolution in the *Drosophila melanogaster* species subgroup: Frequent parameter fluctuations on the timescale of molecular divergence. *Genetics* **172**, 1711–1726 (2006).
48. N. D. Singh, P. F. Arndt, A. G. Clark, C. F. Aquadro, Strong evidence for lineage and sequence specificity of substitution rates and patterns in *Drosophila*. *Mol. Biol. Evol.* **26**, 1591–1605 (2009).
49. H. Akashi, Molecular evolution between *Drosophila melanogaster* and *D. simulans*: Reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. *Genetics* **144**, 1297–1307 (1996).
50. Y.-P. Poh, C.-T. Ting, H.-W. Fu, C. H. Langley, D. J. Begun, Population genomic analysis of base composition evolution in *Drosophila melanogaster*. *Genome Biol. Evol.* **4**, 1245–1255 (2012).
51. H. Akashi, P. Goel, A. John, Ancestral inference and the study of codon bias evolution: Implications for molecular evolutionary analyses of the *Drosophila melanogaster* subgroup. *PloS One* **2**, e1065 (2007).
52. F. Harada, S. Nishimura, Possible anticodon sequences of tRNAHis, tRNAAsn, and tRNAAsp from *Escherichia coli* B. Universal presence of nucleoside Q in the first position of the anticodons of these transfer ribonucleic acids. *Biochemistry* **11**, 301–308 (1972).
53. B. N. White, G. M. Tener, J. Holden, D. T. Suzuki, Activity of a transfer RNA modifying enzyme during the development of *Drosophila* and its relationship to the su(s) locus. *J. Mol. Biol.* **74**, 635–651 (1973).
54. H. Kasai, Y. Kuchino, K. Nihei, S. Nishimura, Distribution of the modified nucleoside Q and its derivatives in animal and plant transfer RNA's. *Nucleic Acids Res.* **2**, 1931–1939 (1975).
55. R. Zallot *et al.*, Plant, animal, and fungal micronutrient Queuosine is salvaged by members of the DUF2419 protein family. *ACS Publ.* **9**, 1812–1825 (2014).
56. C. Fergus, D. Barnes, M. A. Alqasem, V. P. Kelly, The queuine micronutrient: Charting a course from microbe to man. *Nutrients* **7**, 2897–2929 (2015).
57. B. N. White, G. M. Tener, J. Holden, D. T. Suzuki, Analysis of tRNAs during the development of *Drosophila*. *Dev. Biol.* **33**, 185–195 (1973).
58. K. B. Jacobson, W. R. Farkas, J. R. Katze, Presence of queuine in *Drosophila melanogaster*: Correlation of free pool with queuosine content of tRNA and effect of mutations in pteridine metabolism. *Nucleic Acids Res.* **9**, 2351–2366 (1981).
59. Y. Chiari, K. Dion, J. Colborn, A. Parmakelis, J. R. Powell, On the possible role of tRNA base modifications in the evolution of codon usage: Queuosine and *Drosophila*. *J. Mol. Evol.* **70**, 339–345 (2010).
60. X. Wang *et al.*, Queuosine modification protects cognate tRNAs against ribonuclease cleavage. *RNA* **24**, 1305–1313 (2018).
61. S. Muthukumar, C.-T. Li, R.-J. Liu, C. Bellodi, Roles and regulation of tRNA-derived small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **25**, 359–378 (2024).
62. F. Meier, B. Suter, H. Grosjean, G. Keith, E. Kubli, Queuosine modification of the wobble base in tRNAHis influences "in vivo" decoding properties. *EMBO J.* **4**, 823–827 (1985).
63. C. Cirzi *et al.*, Queuosine-tRNA promotes sex-dependent learning and memory formation by maintaining codon-biased translation elongation speed. *EMBO J.* **42**, e112507 (2023).
64. S. Kanaya, Y. Yamada, M. Kinouchi, Y. Kudo, T. Ikemura, Codon usage and tRNA genes in Eukaryotes: Correlation of codon usage diversity with translation efficiency and with CG-Dinucleotide usage as assessed by multivariate analysis. *J. Mol. Evol.* **53**, 290–298 (2001).
65. L. Duret, D. Mouchiroud, Expression pattern and surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4482–4487 (1999).
66. H. Akashi, Translational selection and yeast proteome evolution. *Genetics* **164**, 1291–1303 (2003).
67. S. I. Wright, C. B. K. Yau, M. Looseley, B. C. Meyers, Effects of gene expression on molecular evolution in *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Mol. Biol. Evol.* **21**, 1719–1726 (2004).
68. T. Ikemura, Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* **146**, 1–21 (1981).
69. M. A. Antezana, M. Kreitman, The nonrandom location of synonymous codons suggests that reading frame-independent forces have patterned codon preferences. *J. Mol. Evol.* **49**, 36–43 (1999).
70. P. P. Kokate, S. M. Techtmann, T. Werner, Codon usage bias and dinucleotide preference in 29 *Drosophila* species. *G3 Bethesda Md* **11**, jkab191 (2021).
71. D. L. Halligan, A. Eyre-Walker, P. Andolfatto, P. D. Keightley, Patterns of evolutionary constraints in intronic and intergenic DNA of *Drosophila*. *Genome Res.* **14**, 273–279 (2004).
72. A. Sievers *et al.*, Moderation of structural DNA properties by coupled dinucleotide contents in Eukaryotes. *Genes* **14**, 755 (2023).
73. M. Irony-Tur Sinai *et al.*, AT-dinucleotide rich sequences drive fragile site formation. *Nucleic Acids Res.* **47**, 9685–9695 (2019).
74. C. Workman, A. Krogh, No evidence that mRNAs have lower folding free energies than random sequences with the same dinucleotide distribution. *Nucleic Acids Res.* **27**, 4816–4822 (1999).
75. J.-R. Yang, X. Chen, J. Zhang, Codon-by-codon modulation of translational speed and accuracy via mRNA folding. *PLoS Biol.* **12**, e1001910 (2014).
76. S. A. Langley, G. H. Karpen, C. H. Langley, Nucleosomes shape DNA polymorphism and divergence. *PLoS Genet.* **10**, e1004457 (2014).
77. A. J. Gentles, S. Karlin, Genome-scale compositional comparisons in eukaryotes. *Genome Res.* **11**, 540–546 (2001).
78. A. Fedorov, S. Saxonov, W. Gilbert, Regularities of context-dependent codon bias in eukaryotic genes. *Nucleic Acids Res.* **30**, 1192–1197 (2002).
79. T. Ohta, An examination of the generation-time effect on molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10676–10680 (1993).
80. N. A. Moran, Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2873–2878 (1996).
81. Q. D. Sprengelmeyer *et al.*, Recurrent collection of *Drosophila melanogaster* from Wild African environments and genomic insights into species history. *Mol. Biol. Evol.* **37**, 627–638 (2020).
82. J. B. Lack *et al.*, The *Drosophila* genome nexus: A population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics* **199**, 1229–1241 (2015).
83. M. D. Dean, J. W. O. Ballard, Linking phylogenetics with population genetics to reconstruct the geographic origin of a species. *Mol. Phylogenet. Evol.* **32**, 998–1009 (2004).
84. T. M. Collins, P. H. Wimberger, G. J. P. Naylor, Compositional bias, character-state bias, and character-state reconstruction using parsimony. *Syst. Biol.* **43**, 482–496 (1994).

85. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
86. S. Tavaré, Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* **17**, 57–86 (1986).
87. R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1930).
88. S. Wright, The distribution of gene frequencies under irreversible mutation. *Proc. Natl. Acad. Sci. U.S.A.* **24**, 253–259 (1938).
89. A. Eyre-Walker, M. Woolfit, T. Phelps, The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* **173**, 891–900 (2006).
90. A. Muyle, L. Serres-Giardi, A. Ressayre, J. Escobar, S. Glémin, GC-biased gene conversion and selection affect GC content in the *Oryza* genus (rice). *Mol. Biol. Evol.* **28**, 2695–2706 (2011).
91. M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426 (1986).