Supplementary Materials

Natural selection shaped the rise and fall of passenger pigeon genomic diversity

Gemma G. R. Murray, André E. R. Soares, Ben J. Novak, Nathan K. Schaefer, James A. Cahill, Allan J. Baker, John R. Demboski, Andrew Doll, Rute R. Da Fonseca, Tara L. Fulton, M. Thomas P. Gilbert, Peter D. Heintzman, Brandon Letts, George McIntosh, Brendan L. O'Connell, Mark Peck, Marie-Lorraine Pipes, Edward S. Rice, Kathryn M. Santos, A. Gregory Sohrweide, Samuel H. Vohr, Russell B. Corbett-Detig, Richard E. Green, Beth Shapiro.

Contents

4. References

1. Materials and Methods

1.1. DNA & RNA extraction, library preparation, and sequencing

We extracted DNA from toe pads or bone samples of 84 passenger pigeons (table S1), targeting individuals of known age and geographic origin and maximizing geographic and temporal range. We selected the two best preserved passenger pigeon samples (ROM 34.3.23.2 and ROM 40360) for high coverage nuclear genome sequencing. In addition, we assembled nuclear genomes from published short read data from two passenger pigeons: BMNH1149 (SRA SRX555773(*35*)) and BMNH749 (SRA SRX555813 (*4*)).

We extracted DNA from four band-tailed pigeons: a captive-bred female (BTP2013; Exotic Wings International Aviary, Hemet, CA), an ethanol-preserved muscle sample from a separate individual (AMNH DOT 14025), and two embryonic fibroblast cell cultures (BTP2014 and BTP2015; established by Advanced Cell Technologies, Inc., now Ocata Therapeutics). Both cell cultures were from fertile eggs laid by BTP2013, and were used for the purposes of generating high molecular weight (HMW) DNA libraries.

We extracted DNA from the passenger pigeon samples following standard procedures for working with ancient DNA (*36*), including working in a purpose-built, positive air-pressure clean room, using sterile reagents, supplies, and full-body protective clothing, and processing two negative controls alongside every eight samples. We extracted DNA from bones using protocols optimised for ancient bone (*37*, *38*), and from toe pads using the Qiagen Blood and Tissue Kit, with ancient DNA-specific modifications (*39*). We purified the digested samples using either the Qiagen DNeasy extraction protocol (*39*), Qiagen Nucleotide Removal Kit, or "in-house" silica columns (*37*).

For the first 62 samples processed, we characterized preservation by amplifying a 136 base pair (bp) fragment of the nuclear intron 7 of the fibrinogen beta chain (*39*) and a 136 bp fragment of the mitochondrial cytochrome *b* gene, using the primers 5'-CAAAGAAACCTGAAACACAGG (*40*) (forward) and 5'-GGGACAGCCGAGAATAGGTT (reverse). We performed PCR following (*41*), but with an annealing temperature of 48˚C for cytochrome *b*. We cleaned PCR products using the MagNA bead protocol (*42*), and assessed damage and potential contamination via molecular cloning and Sanger sequencing (*41*). We sequenced the resulting fragments at the University of California Berkeley DNA Sanger sequencing facility. For the 49 extracts from which passenger pigeon DNA could be PCR-amplified, we prepared Illumina sequencing libraries following (*43*). We

cleaned the libraries using MagNA beads as above, and sequenced them at the UCSC Paleogenomics Lab on an Illumina MiSeq using v3 2x75 bp chemistry.

For an additional 22 passenger pigeon samples, including many of the older specimens that are not expected to retain >100-bp fragments, we prepared and screened libraries as above but without a PCR-testing phase.

After screening the above-described 71 libraries, we selected 36 that had either high proportions of endogenous DNA and high complexity or were the oldest in the collection (up to 4000 years BP). We pooled and sequenced these 36 libraries at three Illumina sequencing facilities using: (1) the HiSeq 2500 with 2x100bp paired-end chemistry at the UCSF Center for Advanced Technology, (2) the HiSeq 2500 with 1x100bp single-end chemistry at the Centre of GeoGenetics, Denmark, and (3) the HiSeq 2000 with 1x50 or 1x100bp single-end chemistry at the University of Toronto, Canada. We aimed to recover at least 20-fold mitochondrial genomes and 40-fold nuclear genomes.

We extracted DNA from the band-tailed pigeons AMNH-DOT-14025 and BTP2013 using the Qiagen Blood & Tissue kit, and HMW DNA from BTP2014 and BTP2015 using the Qiagen DNeasy Midi Kit, following the manufacturer's protocols. We sheared the resulting DNA to ~1,000 bp fragments and transformed the extracts into sequencing libraries as above. We pooled the libraries in equimolar ratios and sequenced the pool on two lanes of Illumina MiSeq (v3 chemistry, 2x75bp) at UCSC, and two lanes of Illumina HiSeq 2500 (2x100bp) at UCSF.

1.2. Mitochondrial genome assembly and analysis

To assemble the 36 sequenced passenger pigeon mitochondrial genomes, we first removed sequencing adapters using SeqPrep (http://github.com/jstjohn/SeqPrep). We mapped the reads to the published reference mitochondrial genome of passenger pigeons (GenBank accession KX902243). We used MIA (https://github.com/mpieva/mapping-iterative-assembler), an iterative assembler that uses an ancient DNA-specific substitution matrix. We required a minimum of three independent reads to call a base at each site, and 67% agreement between mapped reads. Sites not meeting these criteria were called as "N" and interpreted as missing data. The average mitochondrial coverage for 19th century passenger pigeons was 78x (20x to 692x), and average coverage ranged from 10-48X for the ~4,000 year-old passenger pigeons (table S2).

In addition to the 36 newly sequenced and assembled passenger pigeon mitochondrial genomes, we downloaded three previously published passenger pigeon mitochondrial genomes (KX902243, KX902244, and KY260683) (*44*), and assembled mitochondrial genomes as above from published short read data for BMNH1149 (SRA SRX555773)(*35*) and BMNH749 (SRA SRX555813) (*4*).

We aligned all 41 passenger pigeon mitochondrial genomes using MUSCLE (*45*) as implemented in SeaView v.4 (*46*). The resulting alignment is 16,948 bp long and contains 255 segregating sites, with an average pairwise difference of 23 sites and average pairwise distance of 0.001 differences per site.

We estimated the coalescent history of the mitochondrial genomes using the Bayesian genealogical inference package, BEAST v1.8.1 (*47*). We assumed the HKY+Γ nucleotide substitution model and the skyline plot model of the coalescent process (*48*, *49*). Because there were few segregating sites and the data was from a single species, we assumed a strict molecular clock. No fossil calibration is available to inform the molecular clock, but a recent study estimated lineage-specific evolutionary rates for all sites in the mitochondrial genome, and for third codon positions only, for 437 different bird species, including passenger pigeons (*14*). We used both rate estimates for passenger pigeons (SM2.1), and report in the main text the results using the third codon position rates, as this is likely not to result in an overestimation of N_e. We ran two MCMC chains for 30 million iterations, discarding the first 10% as burn-in. We visualized convergence of the MCMC chains by eye using Tracer v1.6 (*50*) and calculated the maximum clade credibility tree using TreeAnnotator v1.8 (*51*).

1.3. Draft band-tailed pigeon genome assembly and annotation

We generated a high-coverage and high-contiguity band-tailed pigeon nuclear genome for use as a reference genome onto which to map sequences from the passenger pigeon and remaining band-tailed pigeon libraries. We first Illumina shotgun sequenced BTP2013 to a target of 20x coverage, based on a 1.08 Gb genome size, and used these data to build a *de novo* contig assembly using MERACULOUS (*52*). We provided HMW DNA from BTP2014 to Dovetail Genomics (Santa Cruz, CA), who prepared and sequenced a Chicago library (*53*) from this extract. They used the resulting data to scaffold our *de novo* assembly with their software, HiRise (*53*). This resulted in a genome with an estimated physical coverage of 131x and a total length of 1,089.5 Mb on 9,843 scaffolds, the largest of which is 78.5 Mb. Ninety percent of the genome is contained in 85 scaffolds (minimum length of 1.79 Mb), and at least 50% of the genome is contained in 17 scaffolds with a minimum length of 20.0 Mb.

To annotate the band-tailed pigeon genome, we harvested embryonic brain, heart, lung, liver, muscle, skin, and ovary tissues from an 18-day old band-tailed pigeon embryo at Crystal Bioscience, Inc (San Francisco). We extracted RNA from each tissue using the Direct-zol Kit (Zymo Research), and captured mRNA from these extracts using the Next poly(A) mRNA Magnetic Isolation Module (New England Biolabs). We prepared samples for sequencing using the Ultra Library Prep kit (NEB), with blunt-ending, ligation, and fill-in steps performed with cDNA bound to magnetic SPRI beads with Y adapters. We sequenced these libraries on one lane at the UCSD IGM Genomics Center on an Illumina HiSeq 4000 (2x150). We trimmed the resulting reads using SeqPrep with options "-M 0.05 -N 0.75 -m 0.8 -n 0.02 -X 0.25 -Z 26", and aligned the trimmed reads to the band-tailed pigeon genome using the default options in tophat2 (*54*). We then used these alignments to generate intron annotation hints using the bam2hints script in the AUGUSTUS package with default options(*55*). We generated gene predictions with AUGUSTUS with the options "--alternatives-from-evidence=true --species=chicken --allow_hinted_splicesites=atac --genemodel=complete --noInFrameStop=true", and assigned names to predicted genes based on protein sequence orthology with refseq proteins from the chicken (*Gallus gallus*, galGal4) (*15*) and rock dove (*Columba livia*) (*56*) genomes. We performed reciprocal BLASTP (*57*) of chicken and band-tailed pigeon protein sequences and of rock dove and band-tailed pigeon protein sequences, and assigned names to band-tailed pigeon genes with reciprocal best BLASTP hits in either the rock dove or chicken genomes, giving the rock dove genome precedence. We deposited the annotated assembly and RNA-seq reads used for annotation under NCBI Bioproject PRJNA308039.

1.4. Nuclear genome assembly, genotyping, and alignment

Using the *de novo* band-tailed pigeon as a reference, we assembled nuclear genomes for four passenger pigeons (two for which we generated data and two for which we downloaded data from the SRA) and one additional band-tailed pigeon (AMNH DOT 14025). For all five pigeons, we removed sequencing adapters with SeqPrep as above, and mapped the reads to the draft genome of the band-tailed pigeon described above using BWA-MEM 0.7.10 with default parameters (*58*). We sorted, indexed, and removed duplicates using samtools (*59*) and calculated genome coverage statistics using genomeCoverageBed (*60*).

We genotyped each individual using GATK 3.3 UnifiedGenotyper (*61*). We used PicardTools AddOrReplaceReadGroups to add read group labels to bam files, GATK RealignerTargetCreator and IndelRealigner for indel realignment, and GATK UnifiedGenotyper to call variants.

To mitigate the effects of potential DNA damage and artefacts due to mapping short reads, we used VCFTools (*62*) and a custom Python program to filter our variant call set, in an approach similar to Prüfer *et al. (63)*. We excluded variants for which the root mean squared mapping quality was less than 30 or for which variant quality was below 50, and excluded genotypes for which genotype quality was below 30 or coverage depth was below 5×. Because we observed a small number of sites with extremely high coverage in each individual, we calculated the 97.5th percentile of coverage for each individual after removing sites with greater than 100-fold coverage, and excluded genotypes for which coverage depth exceeded this threshold. To minimize the effect of spurious mappings, we also limited our data set to positions where reference genome 35-mers were found to be at least 50% mappable, using seqbility (https://github.com/lh3/misc/tree/master/seq/seqbility), which uses BWA (*59*) alignments to measure k-mer mappability.

We created two pseudo-haploid reference-based genomes for each individual using our filtered variant call set. At each position in the genome, each haplotype was assigned either a reference or alternate allele, or "N" when a position or genotype did not pass filters. At heterozygous sites, alleles were randomly assigned to one or the other haplotype sequence. To simplify downstream analysis, we excluded indels from these pseudo-haplotype sequences.

We downloaded the published raw data for the rock dove (SRA+ SRR516969), which is the most closely related species to have a high-quality genome assembly (*56*). We mapped the rock dove data to both the band-tailed pigeon genome and the passenger pigeon genome, and performed genotyping as described above. We masked sites that differed across these mappings. We used these data to infer the ancestral states of polymorphisms within passenger pigeons, polymorphisms within band-tailed pigeons, and fixed differences between the species.

We created synteny maps for the band-tailed pigeon and chicken genomes using MUMmer version 3.1 (*64*). We aligned the draft band-tailed pigeon genome to the chicken reference genome (one of the few bird genomes assembled to chromosome level) using the nucMER algorithm. We used the band-tailed pigeon/chicken coordinates from mummerplot to position the band-tailed pigeon scaffolds in relation to the chicken genome in all plots presented here.

1.5. Estimation of neutral diversity and divergence

We estimated diversity within passenger pigeons and within band-tailed pigeons for 5 Mb windows along the scaffolds, ordered according to our mapping to the chicken genome. We excluded sites that were not successfully genotyped for every genome. We used a K80 evolutionary model in *R (65)* to estimate pairwise distances between genomes. We also estimated heterozygosity within each individual. We did this both for all sites and for only sites that were outside of annotated genes, with similar results. We also estimated the frequency of biallelic polymorphic sites that are transitions and transversions separately, in order to investigate the impact of ancient DNA damage.

We estimated substitution rates for derived mutations along both the passenger pigeon and band-tailed pigeon lineages, using the rock dove genome to infer the most likely ancestral state. We omitted sites that were not successfully genotyped for every sample and sites where the ancestral state was ambiguous (i.e. where the rock dove variant differed from the passenger pigeon and band-tailed pigeon variants, or could not be called). We also generated counts of different types of nucleotide base substitution, and used these to infer differences in substitution biases across the two lineages.

1.6. Estimation of diversity and divergence within genes

We extracted alignments of protein-coding genes using the most probable transcript in our annotation of the band-tailed pigeon genome, as determined by AUGUSTUS (*55*). We estimated the numbers of synonymous and non-synonymous sites in each gene. We counted synonymous and nonsynonymous fixed differences in each gene between our passenger and band-tailed pigeon samples, and inferred substitution rates along both the passenger pigeon and band-tailed pigeon lineages using the rock dove genome, as above. We counted the number of polymorphic sites in each gene and estimated diversity at synonymous and nonsynonymous sites within the two species, as above. We also counted the numbers of polymorphic sites at different frequencies in each species, using the rock dove to polarize variants. Only biallelic variants were included in this analysis, and ancestral states were inferred as the variant that matched the majority of variants in passenger or band-tailed pigeon, and the rock dove. For the passenger pigeon samples this was done both for all four individuals, and for only two, in order to facilitate comparisons with the band-tailed pigeon data. We also separately counted substitutions and polymorphisms for different types of nucleotide base change, in order to compare relative rates across the two lineages.

1.7. Selection statistics and tests

We estimated pN/pS, dN/dS, the proportion of nonsynonymous substitutions that were adaptive, $\alpha = 1 - (pN/pS)/(dN/dS)$, and the rate of adaptive substitution relative to the rate of

neutral substitution $\omega_a = \alpha(dN/dS)$ (*66*, *67*) for protein-coding regions of the genome, summing counts either over 5 Mb windows (Fig. 3) or over high- and low-diversity regions of the genome (fig. S3). We used this window size, and did not estimate these statistics for individual genes, in order to ensure accurate estimates given the level of divergence and diversity, and the number of sites that were excluding due to missing data. On average, each 5 Mb window has 18,272 synonymous sites (range: 1,846 - 49,108) that could be called for all individuals in our data set. Passenger pigeons and band-tailed pigeons are estimated to have diverged around 12 millions years ago (*44*), and this has resulted in an average of 2.5% divergence at synonymous sites across their genome. We compared these estimates across passenger pigeons and band-tailed pigeons and across high- and low-diversity regions of the genome, which we defined as regions that have higher or lower diversity in passenger pigeons than in band-tailed pigeons (approximately dividing the genome in half). We also estimated these statistics separately for different types of nucleotide base changes to compare biases across lineages.

We also estimated the direction of selection (DoS = $dN/(dN+dS) - pN/(pN+pS)$)(*68*) for individual genes, since DoS is more robust than other statistics when data are sparse and summing over multiple loci can be affected by biases (*69*). We only estimated DoS for genes with 5 or more synonymous polymorphisms within both the passenger pigeon and band-tailed pigeon samples, and 5 or more synonymous substitutions along both the passenger pigeon and band-tailed pigeon lineages, because DoS estimates are likely to be inaccurate for genes with fewer variable sites. We compared the DoS of genes in high- and low-diversity regions and along the passenger pigeon and band-tailed pigeon lineages. We also tested for evidence of adaptive substitutions in individual genes using a McDonald-Kreitman test (*70*). We implemented this test using a two-tailed Fisher's exact test in *R*.

2. Supplementary Text

2.1. Analyses of the mitochondrial genomes

We analysed 41 passenger pigeon mitochondrial genomes to reconstruct the population history of passenger pigeons. In the main text we report a Bayesian skyline analysis of demographic history. Here, we report the results of additional analyses relevant to assessing whether the rate estimate used in this analysis was appropriate, and analyses that use other rate estimates, to determine how robust our conclusions are.

Minimum Spanning Network

To visualize the relationships between mitogenomic haplotypes, we calculated a minimum spanning network (MSN) using the Tamura-Nei substitution model in Arlequin v3.5 (*71*), which we visualised using HapStar v0.7 (*72*). The MSN displays star radiations, with four to sixteen nucleotide differences between haplotypes (fig. S9), which is consistent with a population expansion.

The structure present in this network and in the phylogeny estimated as part of the dating analysis (Fig. 1B) suggests the presence of two clades, separated by 16 nucleotide differences. These two clades do not correlated with geography (Fig. 1A). This is consistent the absence of geographic structure in the passenger pigeon population. In particular, for two of the locations from which we had multiple passenger pigeon specimens (Troy, New York and Flint, Michigan), samples from the same location fell within different clades.

We also estimated Tajima's D (-2.56) and Fu's F_S (-23.36) using Arlequin v3.5 (*71*). A negative Tajima's D value indicates an excess of low frequency polymorphisms, which could result from population growth or purifying selection. Similarly, a negative F_S value indicates an excess number of alleles, which could result from population growth or genetic hitchhiking.

Testing for selection in the passenger pigeon mitochondrial genome

Due to the small number of segregating sites (255 SNPs), we used all sites in our Bayesian skyline analysis of the passenger pigeon mitochondria. Since selection can affect evolutionary rates, we looked for evidence of selection on nonsynonymous sites in the mitochondrial genomes by comparing counts of synonymous and nonsynonymous polymorphisms and fixed differences between passenger pigeons and a band-tailed pigeon (the previously published mitochondrial genome of KX902240 (*44*)) within protein-coding genes. We tested for differences in the ratios of nonsynonymous to synonymous

polymorphisms and fixed differences using a Fisher's exact test in *R*. We found evidence of an elevation in the ratio of nonsynonymous to synonymous polymorphisms relative to the ratio of nonsynonymous to synonymous fixed differences (Fisher's Exact Test, two-tailed: $p = 0.04$; table S4). This suggests that some nonsynonymous polymorphisms present in the population were weakly deleterious, and would have eventually been purged. We also compared the ratio of nonsynonymous to synonymous polymorphisms at low frequency to the ratio for polymorphisms at high frequency. We found evidence of an elevation in the ratio of nonsynonymous to synonymous polymorphisms at low frequency compared to those at high frequency (Fisher's Exact Test, two-tailed: $p = 0.005$; table S5). This is again consistent with the expected impact of weak purifying selection.

Additional Bayesian skyline plot analyses

The Bayesian skyline plot analysis reported in the main text uses a lineage-specific long-term rate estimate for passenger pigeon mitochondria, estimated for 3rd codon positions (1.25×10^{-8} substitutions/site/year) (*14*). A long-term rate estimate for passenger pigeon mitochondria estimated for all sites is also available (3.0×10^{-9} substitutions/site/year) (*14*). These rate estimates are based on divergence between species. While we included all mitochondrial sites in our skyline plot, neither rate of the two published rate estimates are likely to be accurate for our data since evolution rates are likely to be faster within a species than between species. A difference between evolutionary rates within a species and evolutionary rates between species can result from weak purifying selection, which gradually removes deleterious variants from a population; weakly deleterious variants may contribute to short-term diversity, but are less likely to contribute to divergence between species. Since mitochondrial rates vary a great deal among bird species, a rate estimate cannot be easily be obtained from another species. We therefore used a rate estimate that is likely to be faster than the true rate of the variants present in our passenger pigeon mitochondrial genomes: the lineage-specific long-term rate estimate for 3rd codon positions in passenger pigeons. Since most sites in the mitochondria evolve slower than 3rd codon positions, this rate, despite it being a long-term rate estimate, is unlikely to be slower than the true rate. This means that our estimates of N_e using this rate are likely to be conservative, in the sense that our estimates of N_e will be lower than the true values and our estimates of the dates of demographic changes will be younger than the true dates.

For comparison, we also ran the Bayesian skyline plot analysis using the long-term rate estimate for all sites (fig. S10). This is likely to represent the slowest possible rate at which our variants are evolving. We note that other factors, such as population expansion or changing generation time, might result in a faster rate of evolution than that assumed here.

However without discovering very ancient passenger pigeon remains, it is not possible to test these hypotheses further.

2.2. Assessing the impact of ancient DNA damage

Ancient DNA data can contain artefacts due to the process of DNA decay (*73*, *74*). Common damage observed in ancient DNA data includes strand breakage and miscoding lesions, with the latter primarily occurring through the hydrolytic deamination of cytosine to uracil, particularly at the ends of sequences (*75*). If present, this type of damage can affect genotype calling. Since our analyses involve the comparison of genomes assembled from both ancient and modern DNA, we took precautions to ensure that our results were not biased by damage artefacts. These precautions included measures to ensure that any variants that could have resulted from DNA damage were excluded from our passenger pigeon genome assemblies, and additional checks to determine whether these measures were sufficient and whether our results are robust to any possible artefacts arising from DNA damage. Combined, these measures ensure that DNA damage is unlikely to have biased our results.

A list of the precautions and checks we undertook to minimise any possible impact of ancient DNA damage on our passenger pigeon genome assemblies:

1. We estimated the amount of damage in the all specimens used for genomic analysis. We did this by aligning the recovered reads using BWA-ALN(*59*). Unlike BWA-MEM, BWA-ALN does not soft-clip potentially damaged ends of reads. We then visualized damage patterns of the mapped reads in mapDamage v2.0.5 (*76*). The amount of damage was found to be low; 4% or fewer of cytosines were deaminated at read ends (fig. S11).

2. Although this amount of DNA damage is small, we implemented a conservative filtering strategy (as described in Methods) that should mitigate the influence of any damaged sites.

3. We generated our four passenger pigeons nuclear genomes and the majority of mitochondrial genomes to high coverage (all nuclear genomes and 40/41 mitochondrial genomes had at least coverage 13x coverage; tables S1 and S2), which means that any damaged bases are likely to be obscured by overlapping fragments that have non-damaged bases.

4. We estimated the error rate in both our passenger pigeon and band-tailed pigeon nuclear genome assemblies by identifying a region of the genome that is likely to be the z-

chromosome. While heterozygosity across most of the genome is similar among different individuals (within passenger pigeons and within band-tailed pigeons), the two female passenger pigeon samples and the one female band-tailed pigeon sample show much less heterozygosity than the male samples in a region that mostly maps to the chicken z-chromosome. As this region is likely to be the pigeon z-chromosome, the observed heterozygosity in females is likely to represent error in our calling of variants, either from DNA damage or from misassembly. In this region, we observe extremely low average heterozygosity: 3.8×10^{-4} differences/site for female band-tailed pigeons and 8.0×10^{-4} differences/site for female passenger pigeons (fig. S12). In addition to revealing a low error rate, this also indicates that we cannot exclude the possibility that diversity is close to zero in some regions of the passenger genome.

5. As the most common form of ancient DNA damage is an elevation in the ratio of transitions to transversions (Ts/Tv), we calculated Ts/Tv for passenger pigeons and band-tailed pigeons across their genomes (fig. S13). We found that Ts/Tv is strongly correlated, and consistently lower in passenger pigeons (mean = 2.05) than in band-tailed pigeons (mean = 2.22). This is the opposite pattern to what is expected to result from ancient DNA damage, and could be due to the greater impact of GC-biased gene conversion on passenger pigeon diversity. This hypothesis is consistent with our observation that there is a greater difference in Ts/Tv between passenger pigeons and band-tailed pigeons in high-diversity regions. Nevertheless, we would not expect that differences in Ts/Tv resulting from GC-biased gene conversion would be so great as to reverse an elevation resulting from ancient DNA damage.

In addition to these precautions, we undertook analyses to specifically test for the impact of DNA damage or any other source of sequencing error on one of our central results: the variation in genetic diversity across the genome. We re-estimated diversity omitting variants whose presence is more likely to be the result of error: transitions, singletons and mutations that change GC-content. We observed a similar pattern in diversity in all cases (fig. S2), demonstrating that this is a robust result.

2.3. Comparisons of codon usage bias

Codon usage bias is likely to reflect both mutational biases and natural selection for translational optimization (*17*). Since it is likely that the strength of selection on codon usage is weak and the impact on molecular evolution is widespread, the influence of codon usage bias on molecular evolution may be a particularly useful signal of variation in the efficacy of selection across populations. As a result of the larger population size of passenger pigeons,

and the recombination landscape of the bird genome, we expect that codon usage bias was stronger in the genomes of passenger pigeons than it is in those of band-tailed pigeons, and also was stronger in high-diversity regions of the passenger pigeon genome than in low-diversity regions. To test this we quantified the extent of codon usage bias across the passenger pigeon and band-tailed pigeon genomes using several summary statistics, implemented in a software package described in Novembre (*87*). In particular, we estimated the effective number of codons after accounting for background nucleotide composition (*87*, *88*), which is inversely proportional to the extent of nonuniform codon usage. We also estimated Akashi's scaled χ^2 (*89*) and the *B**(a) measure of Karlin and Mrazek (*90*), both of which also account for background nucleotide composition but are strongly affected by sequence length. We estimated each of these statistics for individual genes in one of our passenger pigeon haplotypes and one of our band-tailed pigeon haplotypes. We used estimates of base composition in non-coding regions for non-overlapping 5 Mb windows of these two haplotypes as a measure of background base composition for each gene.

We found that codon usage bias was higher in the passenger pigeon than in the band-tailed pigeon (fig. S14). We also found that codon usage bias was higher for genes in regions of the genome that have high-diversity in passenger pigeons than those in regions that have low-diversity (defined as 5 Mb windows that have greater than or less than 0.004 nucleotide diversity in passenger pigeons) (fig. S15). While we observed this in both passenger pigeons and band-tailed pigeons, we found that the difference was greater in passenger pigeons (fig. S15).

2.4. Estimation of the population-scaled recombination rate

Bird genomes have an unusually stable chromosome structure, and are highly syntenic (*16*). It has been observed in several bird species that recombination rates are higher at the edges of their large macrochromosomes and in their microchromosomes (*77*–*79*). This pattern has been found to be most extreme in zebra finches, but is present to some degree in all birds for which recombination rates have been estimated (*77*). The increase at the edges of chromosomes has also been observed in other taxa (*77*).

Although we cannot estimate the recombination rate directly in passenger pigeons, we generated a genome-wide recombination map for band-tailed pigeons using LDHat (*80*). We chose to produce the map for band-tailed rather than passenger pigeons because the latter's extraordinary population size posed problems for precomputing a population-scaled recombination rate (ρ) likelihood lookup table.

Starting with our filtered variant call set, we used VCFTools to limit analysis to a smaller subset of variants. We selected only biallelic variants in band-tailed pigeons that were a minimum of 500 bases apart, for which there were no missing genotypes and for which the probability of being in Hardy-Weinberg equilibrium was at least 0.05. We then partitioned variants into 500 SNP windows with a 250 SNP overlap between windows.

We ran LDHat's interval program on each window, running with a block penalty of 5 and 10,000,000 iterations per window as in Brunschwig *et al*. (*81*), and sampling every 40,000 iterations. We used a likelihood lookup table that assumed theta = 0.001. We then used LDHat's stat program to compute the mean population-scaled recombination rate value in each window, discarding the first 80 samples as burn-in. We further smoothed ρ values by averaging them across 5 Mb windows and discarding outliers that fell above the 99th percentile of the genome-wide distribution. We then compared the mean ρ value to passenger pigeon nucleotide diversity in each window, performed linear least-squares regression, and computed Spearman's r^2 and ρ.

We found that mean ρ and nucleotide diversity in the passenger pigeon genome were strongly correlated (fig. S16; Spearman's $r^2 = 0.68$, Spearman's rank correlation test $p < 2.2\text{x}10^{-16}$). Since ρ is the product of N_e and recombination rate, the correlation between mean ρ and diversity in the passenger pigeon genome could be driven by either differences in N_e or in recombination rate. However, we find little evidence of substantial variation in N_e across the band-tailed pigeon genome, which suggests that the correlation between ρ and diversity in the passenger pigeon genome is more likely to have been driven by variation in the recombination rate across the band-tailed pigeon genome.

2.5. Hung *et al.* (2014) and PSMC analyses

A previous study by Hung *et al*. (*4*) used a Pairwise Sequentially Markovian Coalescent (PSMC) model (*33*) of the passenger pigeon nuclear genome to infer the demographic history of the species. In this study we demonstrate that natural selection had a strong and widespread impact on diversity in the passenger pigeon nuclear genome. One implication of this result is that diversity in the nuclear genome cannot be used to reliably infer the demographic history of the species (*34*). Here we discuss the results reported by Hung *et al*. (*4*) and explore the effect that selection on linked sites has on a PSMC analysis of passenger pigeon genomes.

We created whole-genome diploid consensus sequences for the four passenger pigeon genomes and the two band-tailed pigeons genomes in our data set. In addition to performing

PSMC using all available data from these genomes, we performed an exploratory analysis in which we inferred demographic history using a PSMC analysis applied to high and low-diversity regions separately. To achieve this, we binned the passenger pigeon genome into high- and low-diversity regions, based on whether each 50 kb window had higher or lower nucleotide diversity in passenger than in band-tailed pigeons. We then created different PSMC input files for these high- and low-diversity genomic regions, using PSMC's fq2psmcfa. For all PSMC analyses, we used a window size of 10 bp. We then plotted the results using psmc_plot.pl, which is part of the PSMC package. In order to scale the population parameters, we used a generation time of 4 years (*82*) and a mutation rate of 5.68×10^{-9} substitutions/site/generation (equivalent to 1.42×10^{-9} substitutions/site/year (*56*)). We ran 100 bootstrap replicates for each analysis.

Our PSMC analyses of the whole genome suggest that the passenger pigeon population fluctuated in size several times between 20 million and 100,000 years ago (figs. S17A,B and S18A,B). These results differ from previously reported PSMC plots for passenger pigeons (*4*), both in the timing and the scale of population size fluctuations, with the previous results (fig. S19). These differences are due to the parameters selected for population size reconstruction. Specifically, when performing PSMC modeling with extinct taxa, it is necessary to adjust parameters for 1) the divergence between the reference genome and input data, and 2) the possibilities of damage and contamination, which are common in ancient DNA data sets. In Hung *et al. (4)*, the authors use the "-C50" flag as part of the consensus calling step in PSMC. This option lowers the mapping quality for reads containing excessive mismatches, and is intended to increase stringency when mapping reads to a reference genome of the same species. However, when mapping reads to a distantly related reference genome as in Hung *et al. (4)*, this option will exclude distantly related reads, resulting in underestimates of population size. This effect becomes more pronounced with increasing evolutionary distance between the reference genome and query sequencing library, and may therefore explain the different results observed by Hung *et al. (4)*.

PSMC analyses are only valid if most sites in the genome have evolved neutrally; otherwise they may return results that reflect the impact of selection rather than demography on genomic diversity (*34*). Our estimates of diversity across the passenger pigeon genome (Fig. 2) suggest a substantial impact of selection on sites along the passenger pigeon lineage, and that the extent of this impact varied across different regions of the genome. To explore this, we compared the results of PSMC analyses on the high- and low-diversity regions separately. These regions are of approximately equal size, and although the substantial variation in diversity across the 'high-diversity' regions means that we cannot assume

neutrality across any region of the passenger pigeon genome, we would expect that PSMC analyses of low-diversity regions will reflect selective forces more than analyses of high-diversity regions, and that analyses of the high-diversity regions will reflect demographic change more than low-diversity regions. Our results were consistent with this expectation: we observe a clear and consistent difference between the results of PSMC analyses on the high- and low-diversity regions of each passenger pigeon genome (figs. S17 and S18). Moreover, these results indicate an historic expansion in N_e for the high-diversity regions (figs. S17C,D and S18C,D), and an historic contraction in N_e for low-diversity regions (figs. S17E,F and S18E,F). While interpretation of these results is not straightforward, they are consistent with an expansion in the passenger pigeon population driving an increase in the N_e of high-diversity regions, and a decrease in the N_e of low-diversity regions due to an increased impact of selection.

In contrast, our PSMC analyses of the homologous regions of one of our band-tailed pigeons (AMNH DOT 14025) are very similar (fig. S20C,E). This is consistent with selection having had much less of an impact on neutral variation along the band-tailed pigeon lineage. While we do observe differences in the results of our PSMC analyses of the two regions of our other band-tailed pigeon (BTP2013) (fig. S20D,F), this individual was bred in captivity while AMNH DOT 14025 was captured from the wild. Captive breeding programs can complicate reconstructions of ancestral population size if they involve matings between individuals from distinct populations within a species. Also, a previous study found evidence that population structure within a bird species had a greater impact on variation in the low-recombination regions of the genome (*83*). Differences in the impact of population structure across the genome might drive differences in our PSMC reconstructions following recent admixture.

2.6. Tests for adaptive evolution in two different functional classes of genes

To determine whether the patterns we observe across genes in high- and low-diversity regions of the passenger pigeon genome (Fig. 3 and fig. S4) are the result of differences in the types of genes located in high- and low-diversity regions, we identified genes involved in two functions that are likely to be under positive selection: spermatogenesis and immunity. We compared patterns of polymorphism and divergence across high- and low-diversity regions within these two sets of genes, in both passenger pigeons and band-tailed pigeons, and compared these patterns to those we observed across all genes. An observation of similar patterns in selective constraint and rates of adaptive evolution within sets of genes with similar functions would support our hypothesis that this variation is driven by a factor specific to these regions of the genome, rather than factors specific to particular types of genes.

We identified 69 genes involved in spermatogenesis from a list of genes identified in other bird species (*84*): 48 are in high-diversity regions and 21 are in low-diversity regions. We also identified 99 genes in immunity pathways that are annotated for *Columba livia (85)*: 59 in high-diversity regions and 40 in low-diversity regions. Genes were identified both using our annotation, and a protein-level BLAST of representatives of the genes we identified from other bird species, obtained from GenBank, against our annotation. Then, for each set of genes, we compared counts of nonsynonymous and synonymous polymorphisms within passenger pigeons and within band-tailed pigeons, and fixed differences between the two species. In particular, we compared the ratio of nonsynonymous to synonymous polymorphism counts, and the ratio of nonsynonymous to synonymous fixed difference counts. We also tested for evidence of adaptive substitutions using a McDonald-Kreitman test, implemented in *R* using a Fisher's Exact Test (two-sided). We categorised genes according to whether they fall in a high- or low-diversity regions of the passenger pigeon genome (defined as regions that have higher or lower diversity than the homologous regions of the band-tailed pigeon genome), and summed counts from genes within either region to increase power. Due to the small numbers of polymorphic sites we did not differentiate between low and moderate frequency polymorphic sites, which can make the test conservative (*86*).

For both sets of genes we observe similar patterns within the set of genes as we do across all genes. In particular, we observe a lower ratio of nonsynonymous to synonymous polymorphism and a higher ratio of nonsynonymous to synonymous fixed difference in high-diversity regions of the passenger pigeon genome than in low-diversity regions of the passenger pigeon genome (tables S8 and S9). This suggests that genes in high-diversity regions have both more efficient selective constraint and a faster rate of adaptive substitution. Moreover, while broadly similar differences are observed across the band-tailed pigeon genome, the size of the differences are much less. McDonald-Kreitman tests on the set of spermatogenesis genes present in low-diversity regions yielded no evidence of adaptive change in passenger pigeons or in band-tailed pigeons (directions of selection were negative, and a Fisher's exact test, two-sided, gave $p = 0.88$ for passenger pigeons and $p = 0.73$ for band-tailed pigeons). The same test for genes present in high-diversity regions yielded evidence for adaptive substitutions in passenger pigeons (direction of selection was positive, $p = 3.4 \times 10^{-5}$), but not in band-tailed pigeons (direction of selection was negative, $p = 0.89$). McDonald-Kreitman tests on immunity-related genes in low-diversity regions also yielded no evidence of adaptive change in passenger pigeons or in band-tailed pigeons (directions of selection were negative, and a Fisher's exact test, two-sided, gave $p = 1.0$ for

passenger pigeons and $p = 0.73$ for band-tailed pigeons). The same test for genes present in high-diversity regions yielded evidence for adaptive substitutions in passenger pigeons (direction of selection was positive, $p = 3.3 \times 10^{-5}$), but not in band-tailed pigeons (direction of selection was negative, $p = 0.89$). These results further support our conclusion that adaptive evolution is more efficient in high-diversity regions of the passenger pigeon genome.

2.7. Gene density across the genome

The impact of selection on diversity at linked neutral sites is both expected to be negatively correlated with recombination rate, and positively correlated with the density of sites that may be targets of selection. We therefore estimated the density of genes across the genome, to see whether that could explain the patterns we observe in genetic diversity across the passenger pigeon genome. We found that while the number of genes per window was greater in higher diversity windows, these genes tended to be shorter (fig. S21). Nevertheless, overall, the density of genes (i.e. the proportion of sites within a window that are in protein-coding regions) tended to be higher in higher diversity windows (fig. S21). If differences in gene density were driving the patterns we observe in diversity and the density of selection, we would expect the reverse, i.e. a greater density of genes in low diversity windows.

2.8. Tajima's D and H-statistics for the nuclear genomes

Tajima's D and H-statistics can be used to distinguish the impact of selective sweeps (positive selection) and background selection (negative selection) on neutral diversity at sites linked to those under selection. Both selective sweeps and background selection are predicted to result in an excess of low-frequency variants (a negative Tajima's D), however only selective sweeps are expected to result in an excess of high frequency derived variants (a positive H-statistic). A negative Tajima's D can also be the result of population expansion, whereas demographic change is not expected to impact the H-statistic (*91*).

We estimated these statistics for 5 Mb windows across the nuclear genomes of both passenger and band-tailed pigeons. We calculated Tajima's D is as described in Tajima (*92*) and the H-statistic as described in Fay and Wu (*91*). While we observed a negative Tajima's D in both passenger and band-tailed pigeons (fig. S22), a test implemented in the DH software package to determine the significance of this deviation from neutrality yielded no evidence in either population (*93*, *94*). Nor did statistical tests implemented in this package yield any evidence that the H-statistics differed from zero.

2.9. Admixture analysis

While there are no reasons to expect admixture events to affect the genetic diversity at the edges of the macrochromosomes and the microchromosomes differently to the other chromosomal regions, we explored whether introgression from a different species could have influenced diversity across the passenger pigeon genome (Fig. 2). The geographic ranges of the passenger pigeon and band-tailed pigeon did not overlap or border one another (Fig. 1), and there are no historical records of passenger pigeons breeding with other pigeon species. Despite that, we tested whether there was any evidence that the closest living relative of passenger pigeons, the band-tailed pigeon, introgressed into passenger pigeons using the D-statistic (ABBA-BABA test), and quantified potential admixture with the related $\hat{f}$ statistic (*95*, *96*). We generated representative haploid sequences for each individual by randomly selecting a single high quality base call mapped to each position in the reference genome (BaseQ≥30 and MapQ≥30), and carried out all possible D-statistic and $\hat{f}$ statistic tests, consistent with the species tree, between the four passenger pigeons and two band-tailed pigeons. To determine the ancestral state of alleles we used the rock dove as an outgroup for all comparisons. To avoid post-mortem damage introduced biases we restricted D and $\hat{f}$ analyses to transversion sites that are not affected by cytosine deamination damage, characteristic of ancient DNA (*73*).

We found potential evidence of introgression from band-tailed pigeons into passenger pigeons (D up to 0.17, Z score > 3.0) (table S6). However, the total amount of band-tailed ancestry in passenger pigeons, as measured by $\hat{f}$, was consistent with only a difference of 0.6% in band-tailed pigeon ancestry among passenger pigeons ($\hat{f}$=0.006) (table S7). In the case of band-tailed introgression into passenger pigeons, we observe D-statistic values much greater than $\hat{f}$, which can happen when the number of informative sites is low and most of the genome is congruent to the species tree. While our D and $\hat{f}$ statistic results are consistent with a very small amount of band-tailed pigeon introgression into passenger pigeons, there are not enough species tree incongruent sites to explain the varying amount of diversity along the passenger pigeon chromosomes.

3. Supplementary Figures and Tables

Figure S1. Relationship between nucleotide diversity across the passenger pigeon and band-tailed pigeon genomes. Diversity within passenger pigeons plotted against diversity within band-tailed pigeons, estimated for 5 Mb windows across the genome. Diversity is calculated as the mean proportions of sites that are heterozygous (similar results are obtained for comparisons between samples). Diversity in the two species is positively correlated (Spearman's rank correlation test, $p < 2x10^{-16}$). The green dashed line represents equality.

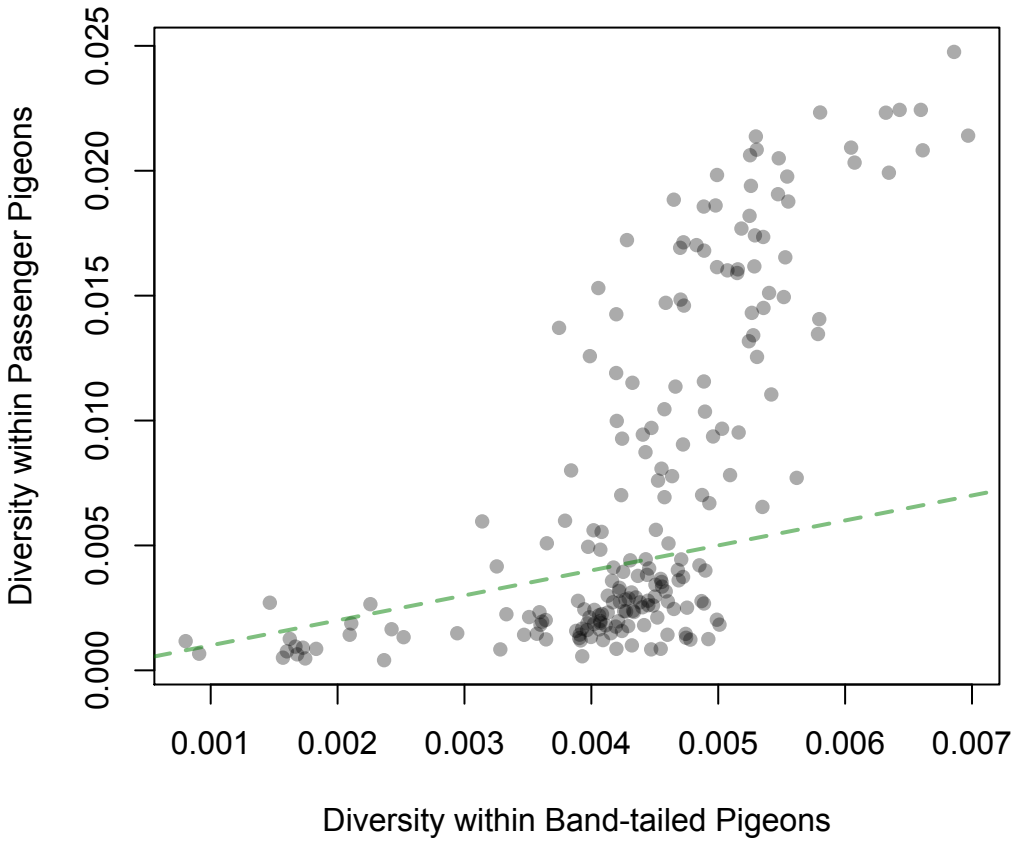


Figure S2. Estimates of diversity across the passenger pigeon and band-tailed pigeon genomes omitting certain types of variant. The proportion of synonymous sites that are bi-allelic polymorphisms, that are (**a**) the result of transitions (dashed lines) or transversions (solid lines), (**b**) at a frequency of 2/4 in either species sample (passenger pigeons subsampled for comparison), and (**c**) the result of G/C to G/C mutations or A/T to A/T mutations, estimated for 5 Mb windows along our scaffolds, ordered according to our mapping to the chicken genome. Estimates for passenger pigeons are red lines and estimates for band-tailed pigeons are blue lines. In (a) only sites outside of genes are included, in (b) and (c) only synonymous sites within genes are counted. Vertical dashed lines represent chromosome boundaries.

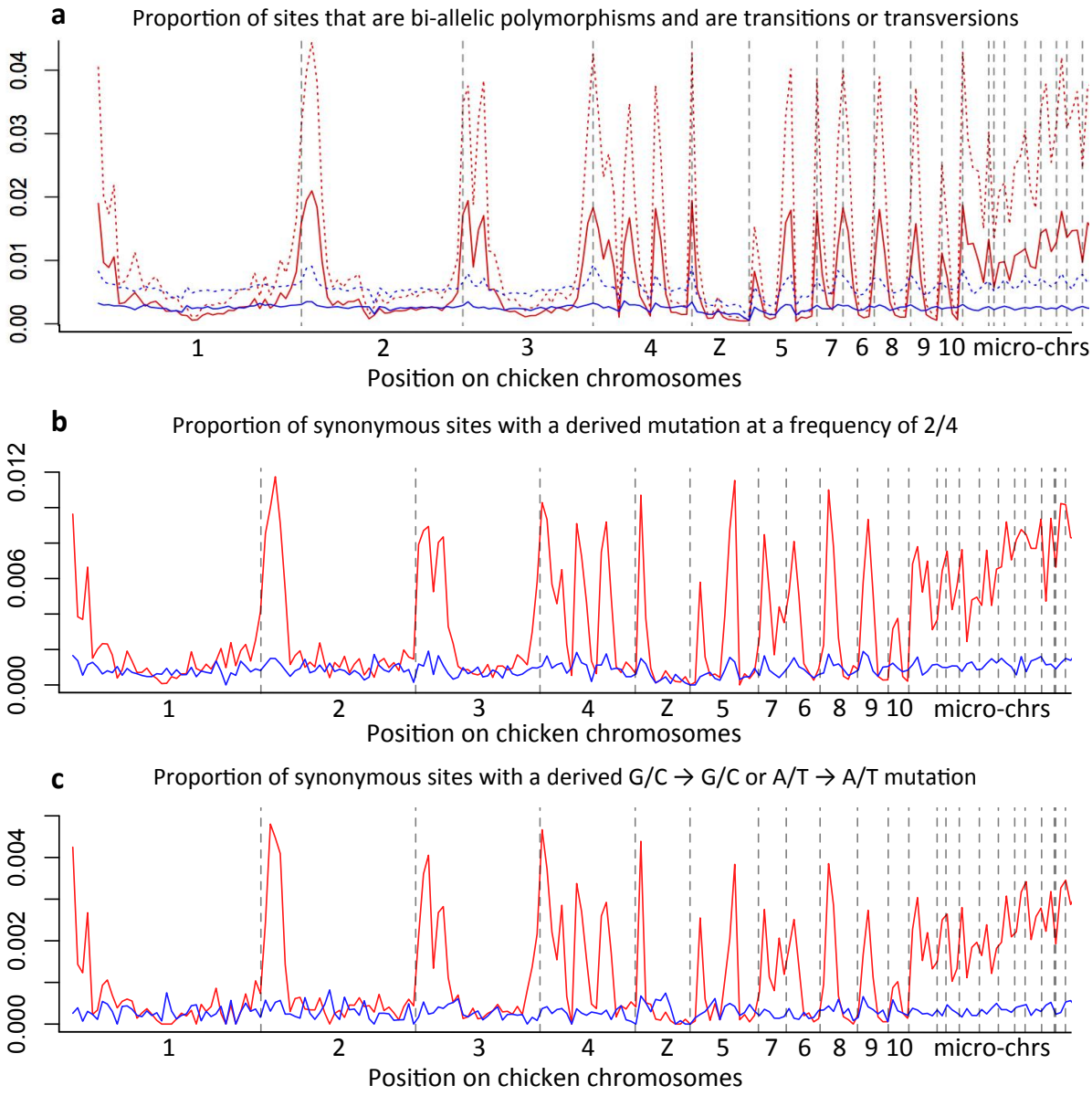


Figure S3. Estimates of ω_a and α, and *pN/pS* for different frequencies of derived mutations in passenger pigeons. (a) and (b) show estimates of the rate of adaptive evolution (ω_a). (c) and (d) show estimates of the proportion of amino-acid substitutions that were driven to fixation by positive selection (α). (e) and (f) show the ratio of nonsynonymous to synonymous rates of polymorphism (*pN*/*pS*). In each plot, estimates are based on derived mutations in passenger pigeons at particular frequencies in our sample. In (a), (c) and (e) estimates are for 5Mb windows in either high-diversity (left, filled boxes) or low-diversity (right, empty boxes). In (b), (d) and (f) estimates are obtained by summing over all genes in high-diversity regions (filled circles) and low-diversity regions (empty squares) and 95% confidence intervals obtained by bootstrapping across genes. This method of estimating ω_a and α follows an approach described in Messer and Petrov(*97*). Dashed lines (in black in a, c and e, and in red in b, d and f) represent the median estimate from mutations at frequencies of 3/8 - 6/8 for high-diversity regions. Dotted lines (in black in a, c and e, and in red in b, d and f) represent the median estimate from mutations at frequencies of 3/8 - 6/8 for low-diversity regions. The values of these estimates are reported in each plot.

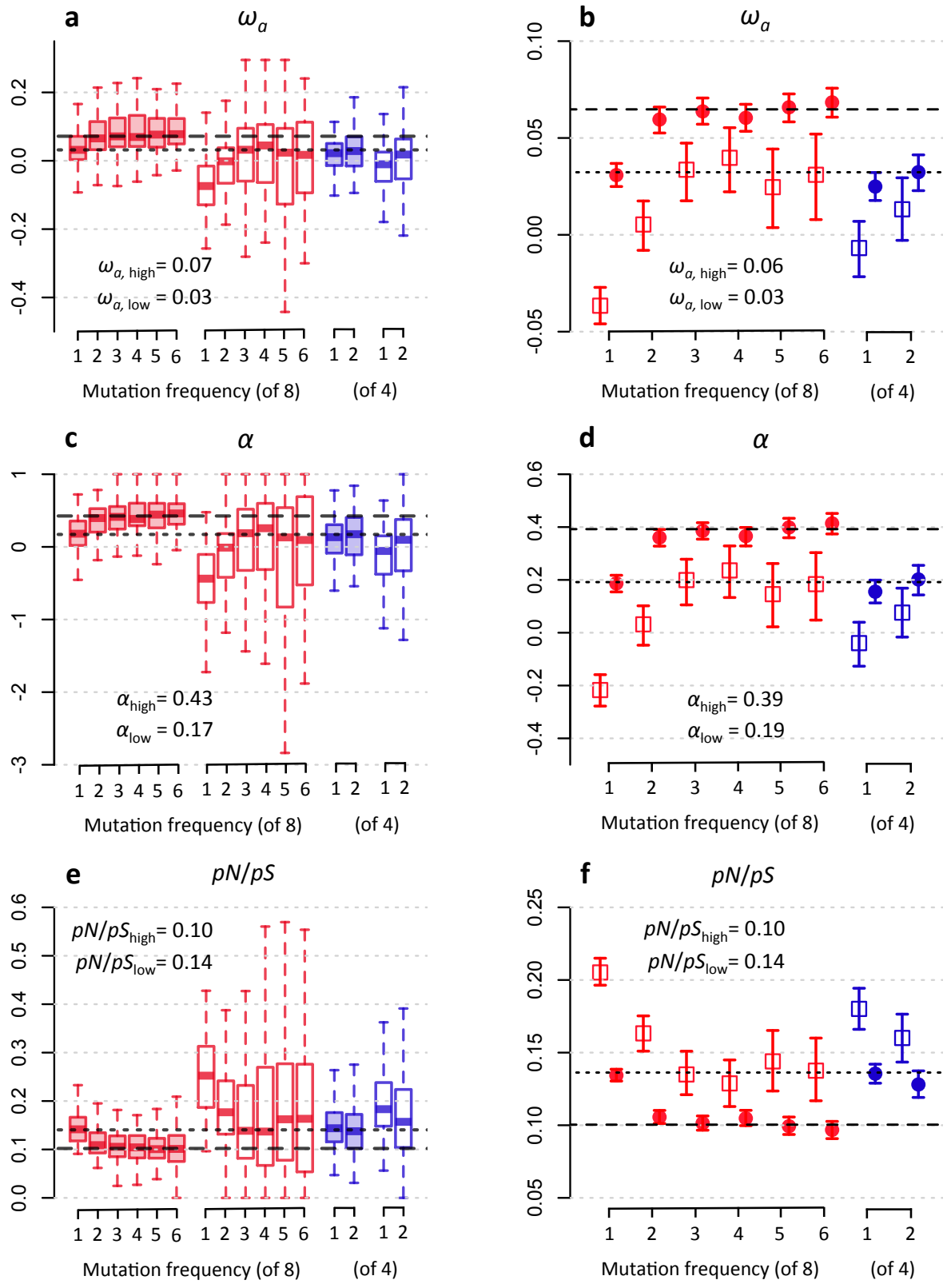


Figure S4. Estimates of the direction of selection (DoS) for individual genes in different regions of the genome, in passenger pigeons and band-tailed pigeons. (a) and (b) show estimates of DoS for individual genes, plotted against the relative diversity of the 5 Mb window in which they are located. Only genes with 5 or more synonymous polymorphisms within both the passenger pigeon and band-tailed pigeon samples, and 5 or more synonymous substitutions along both the passenger pigeon and band-tailed pigeon lineages (other genes were excluded due to the likelihood of inaccurate estimates). Horizontal dashed lines indicate neutrality, and vertical dashed lines equality of diversity between the two species. Grey solid lines show a best-fit linear regression.

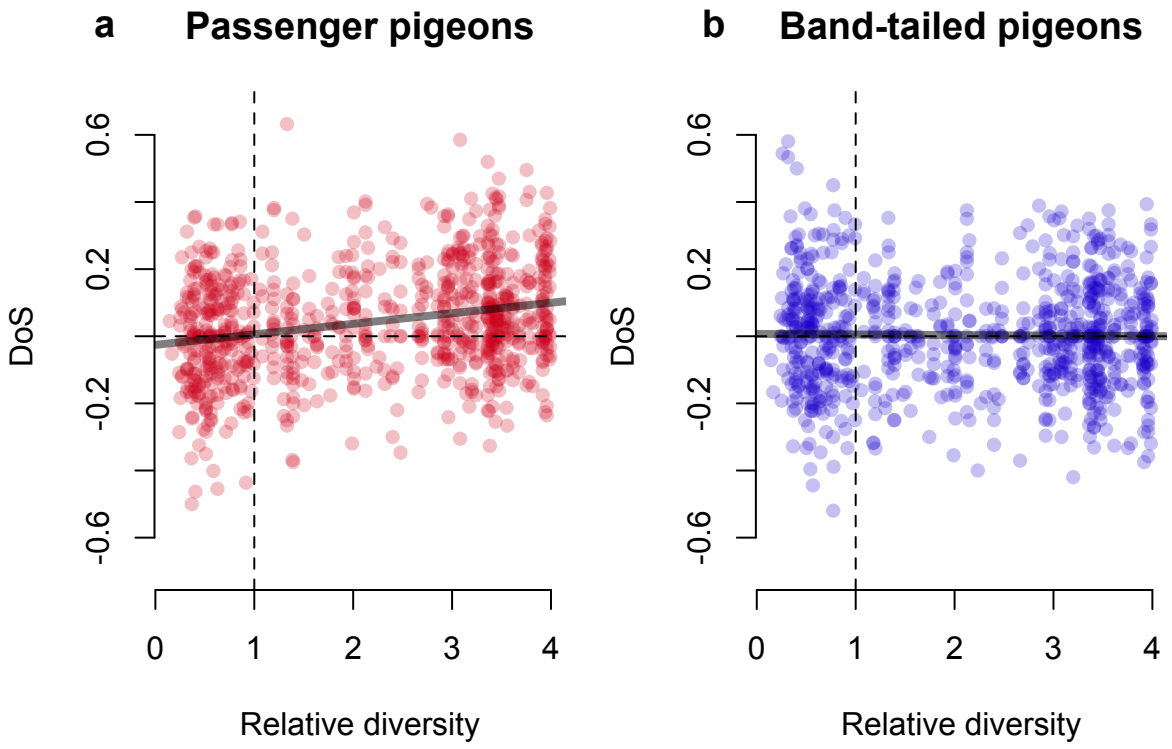


Figure S5. GC-content and neutral substitution biases across the passenger pigeon and band-tailed pigeon genomes. (a) and (b) show the GC-content of the passenger pigeon (a) and band-tailed pigeon (b) genomes against relative diversity across the two species, estimated for 5Mb windows across the genome. (c) shows the GC-content in the passenger pigeon genome relative to the GC-content in the band-tailed pigeon genome. (d) to (g) show substitution biases for non-coding regions of the genome, against relative diversity across the two species, estimated for 5Mb windows across the genome. (d) and (e) show the rate of A/T to G/C substitution relative to the rate of A/T to A/T substitution along the passenger pigeon (d) and band-tailed pigeon (e) lineages. (f) and (g) show the rate of G/C to A/T substitution relative to the rate of G/C to G/C substitution along the passenger pigeon (f) and band-tailed pigeon (g) lineages.

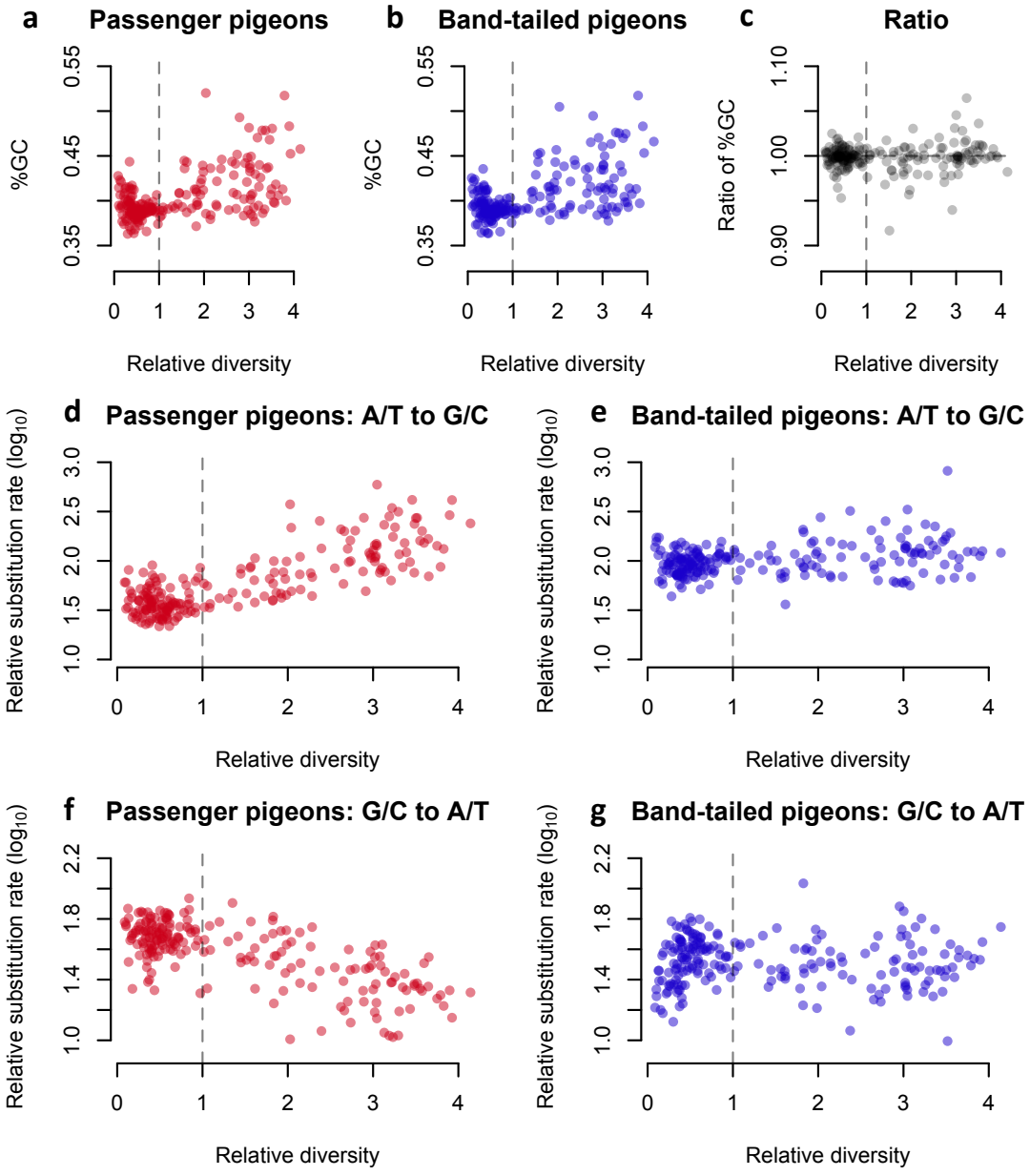


Figure S6. The proportion of substitutions that are nonsynonymous along the passenger pigeon and band-tailed pigeon lineages against relative diversity across their genomes. (a) and (b) show $dN/(dN+dS)$ for all types of substitution. (c) and (d) show $dN/(dN+dS)$ for just substitutions that are unaffected by gBGC. (e) and (f) show $dN/(dN+dS)$ for substitutions that are impeded by gBGC (G/C to A/T) relative to those that are promoted by gBGC (A/T to G/C). (a), (c) and (e) are for the passenger pigeon lineage, and (b), (d) and (f) are for the band-tailed pigeon lineage. Relative diversity is π in passenger pigeons relative to π in band-tailed pigeons. All estimates are for 5 Mb windows across the genome. Vertical dashed lines represent equal diversity across the two species, solid lines represent best-fit linear regressions.

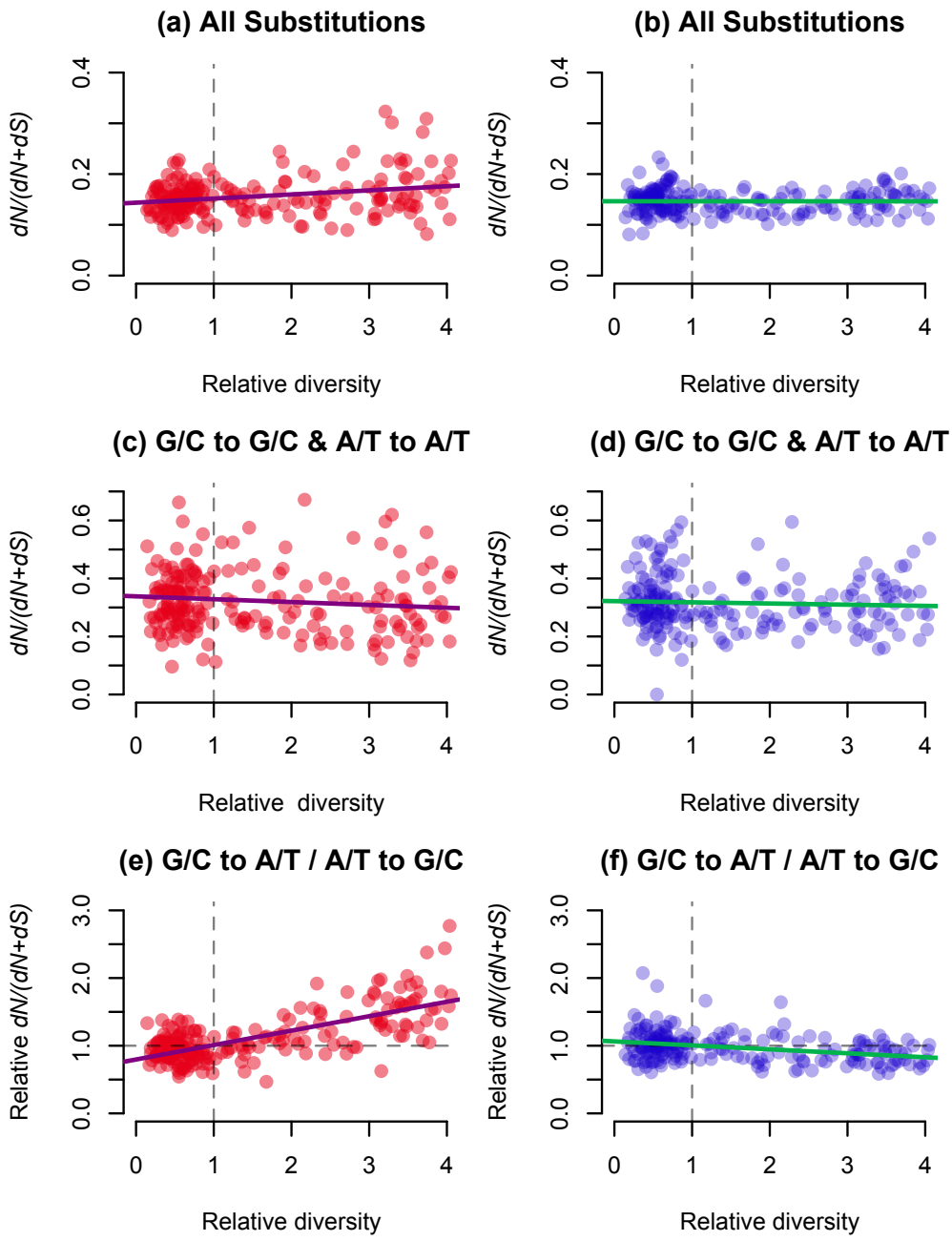


Figure S7. Uncorrected estimates of ω_a, α and *pN/pS* for different types of nucleotide base change. These estimates are not corrected for differences in base composition across nonsynonymous and synonymous sites, and across different regions of the genome, and so, with the exception of G/C to G/C and A/T to A/T mutations (which should be independent of base composition), they should be considered only for comparative purposes. Otherwise, these plots and the methods used to obtain the estimates are the same as in figure S3.

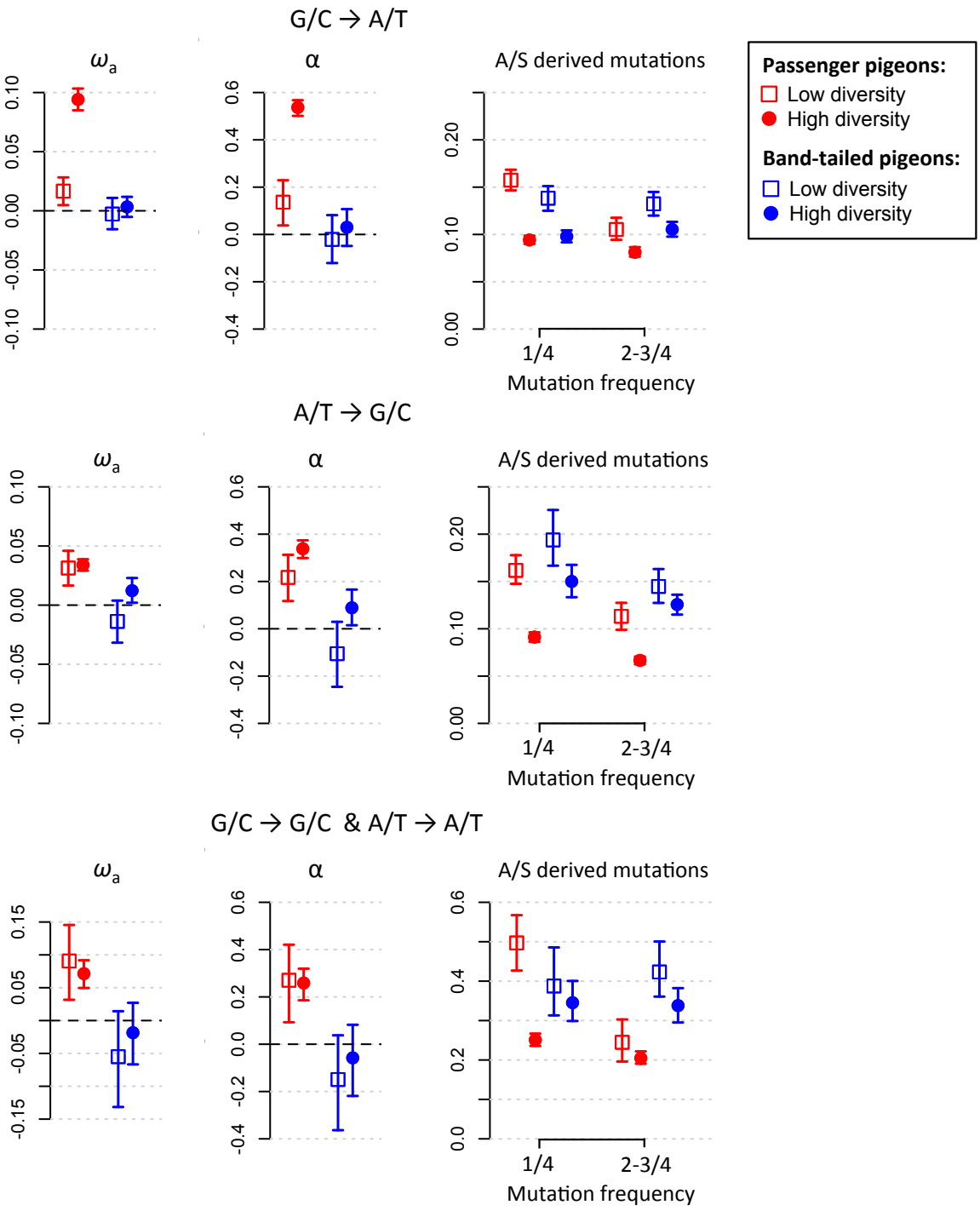


Figure S8. Comparisons of estimates of the ratio of nonsynonymous to synonymous counts of different types of derived nucleotide base change, at different frequencies in our sample, using all 8 passenger pigeon alleles. The estimates are calculated as described in figures S3 and S7.

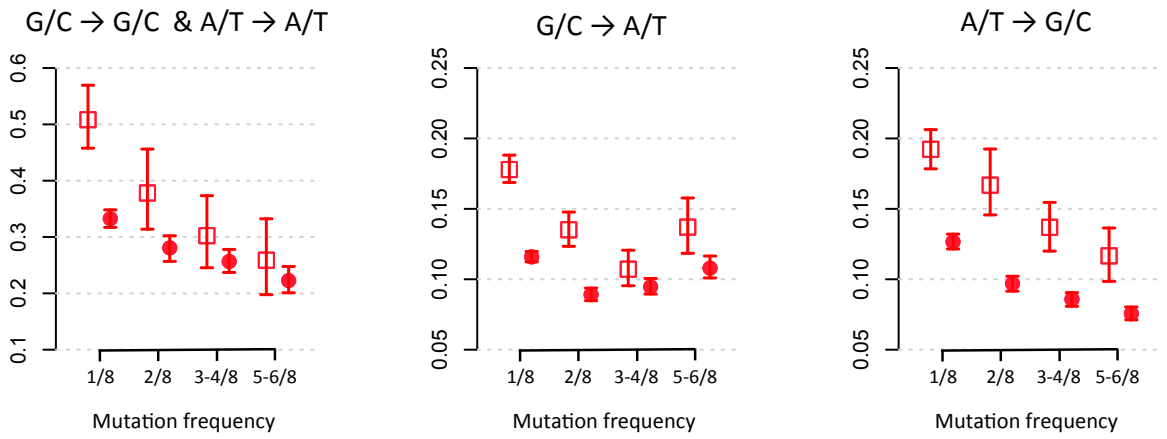


Figure S9. A minimum spanning network of the 41 passenger pigeon mitochondrial genomes. Large circles represent observed haplotypes and small circles represent inferred intermediate haplotypes. Each step between circles represents a single substitution. Colors follow those depicted in the mitochondrial phylogeny in Fig. 1B.

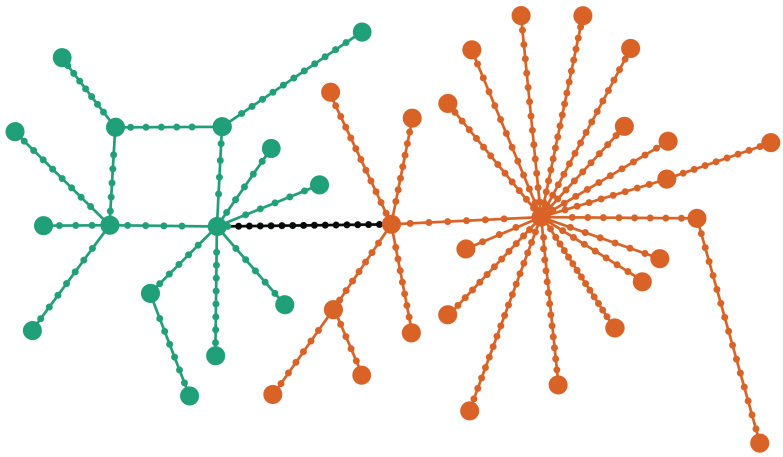


Figure S10. Inferred N_e estimated using using two different calibration rates. Both rates are lineage-specific estimates for passenger pigeons in (*14*): 3.00×10^{-9} substitutions/site/year rate, which was estimated for all sites, for the passenger pigeon lineage, and 1.25×10^{-8} substitutions/site/year estimated for the third codon position of cytochrome *b* for the passenger pigeon lineage. Dashed lines represent the Last Glacial Maximum (LGM) period. LGM estimates are from (*98*).

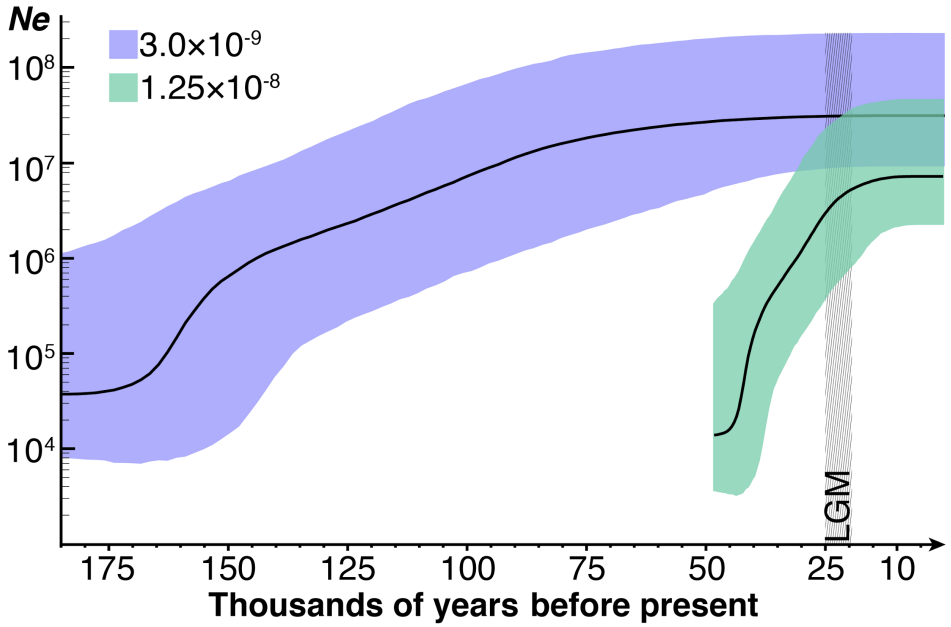
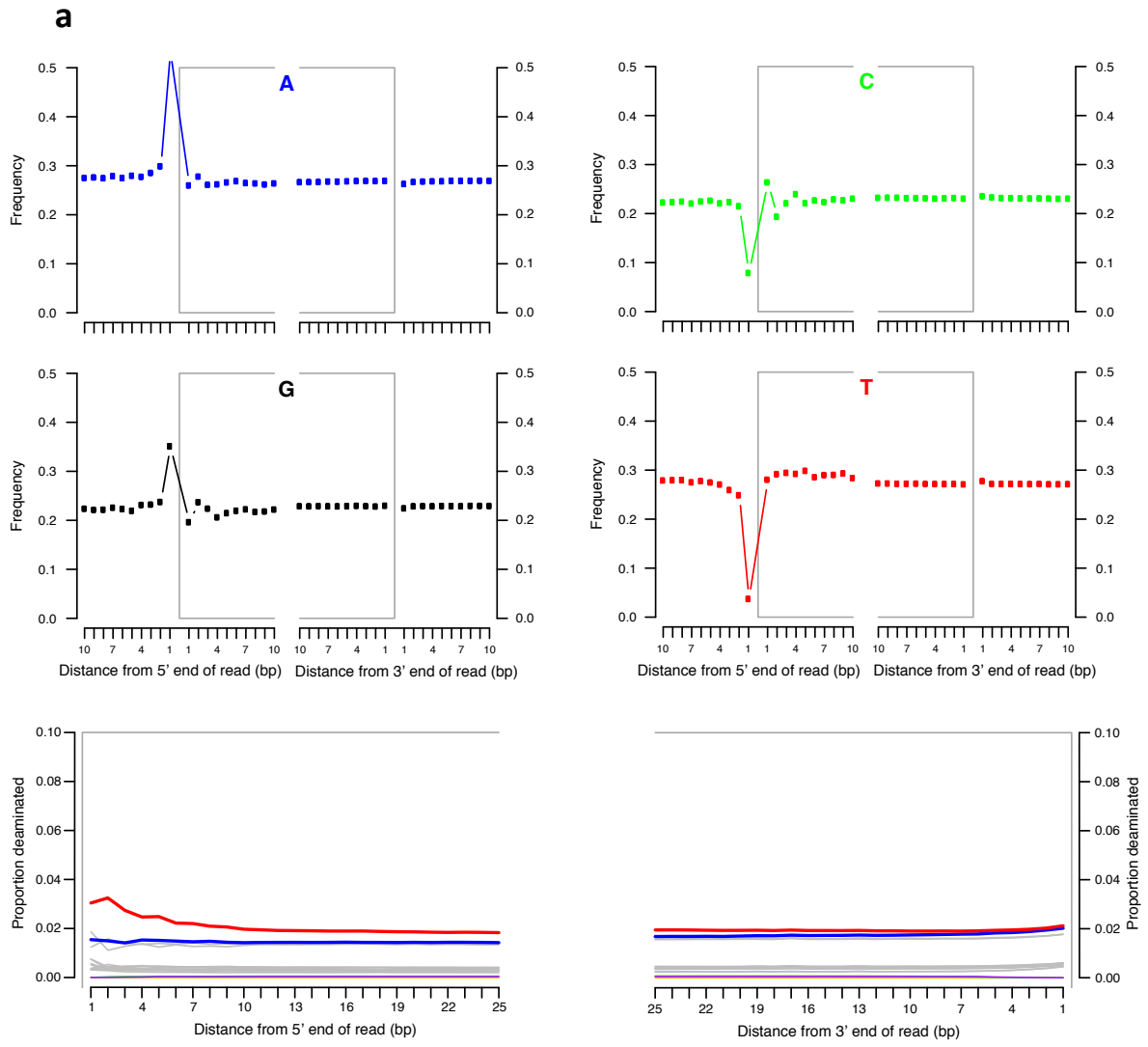
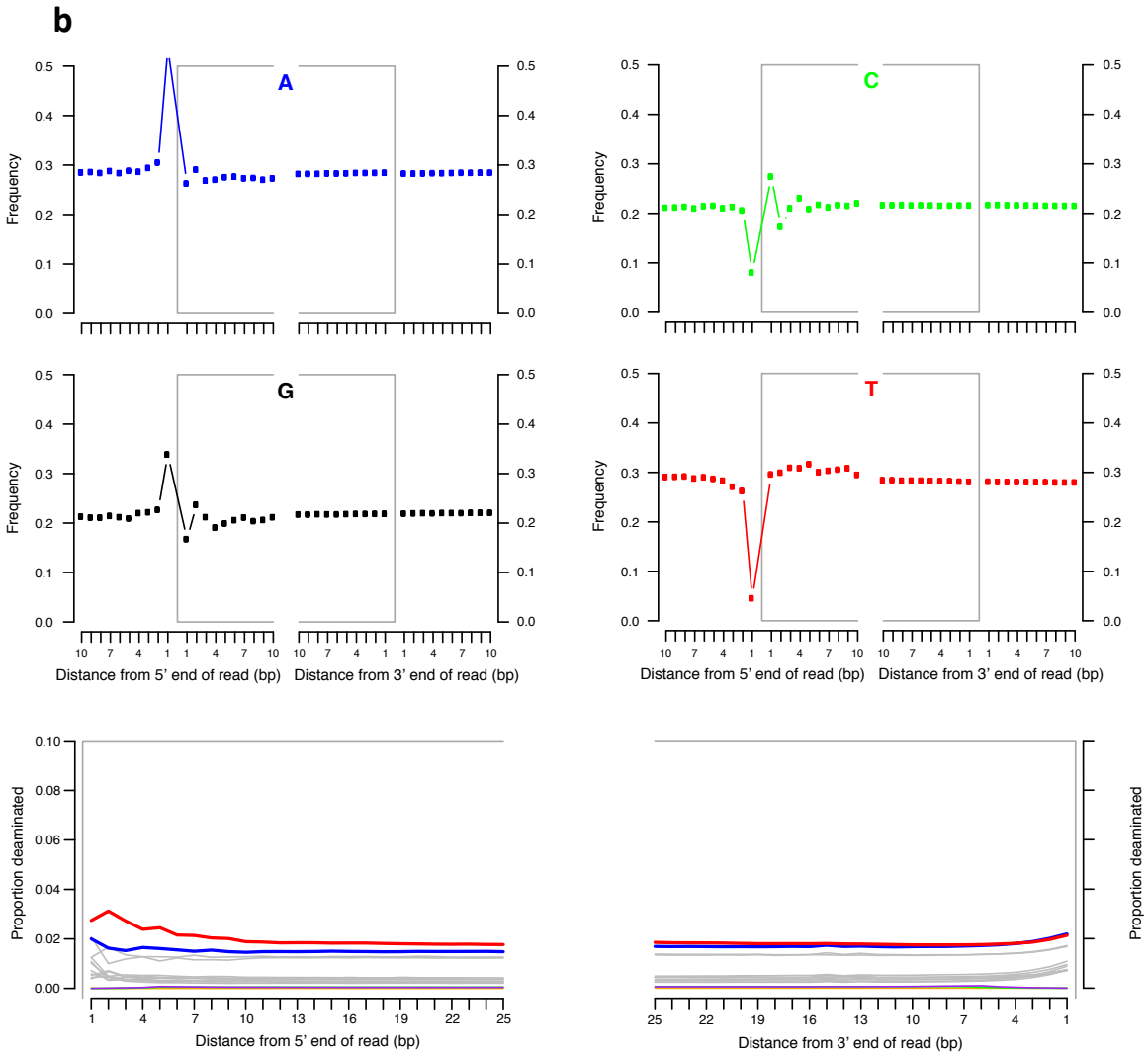
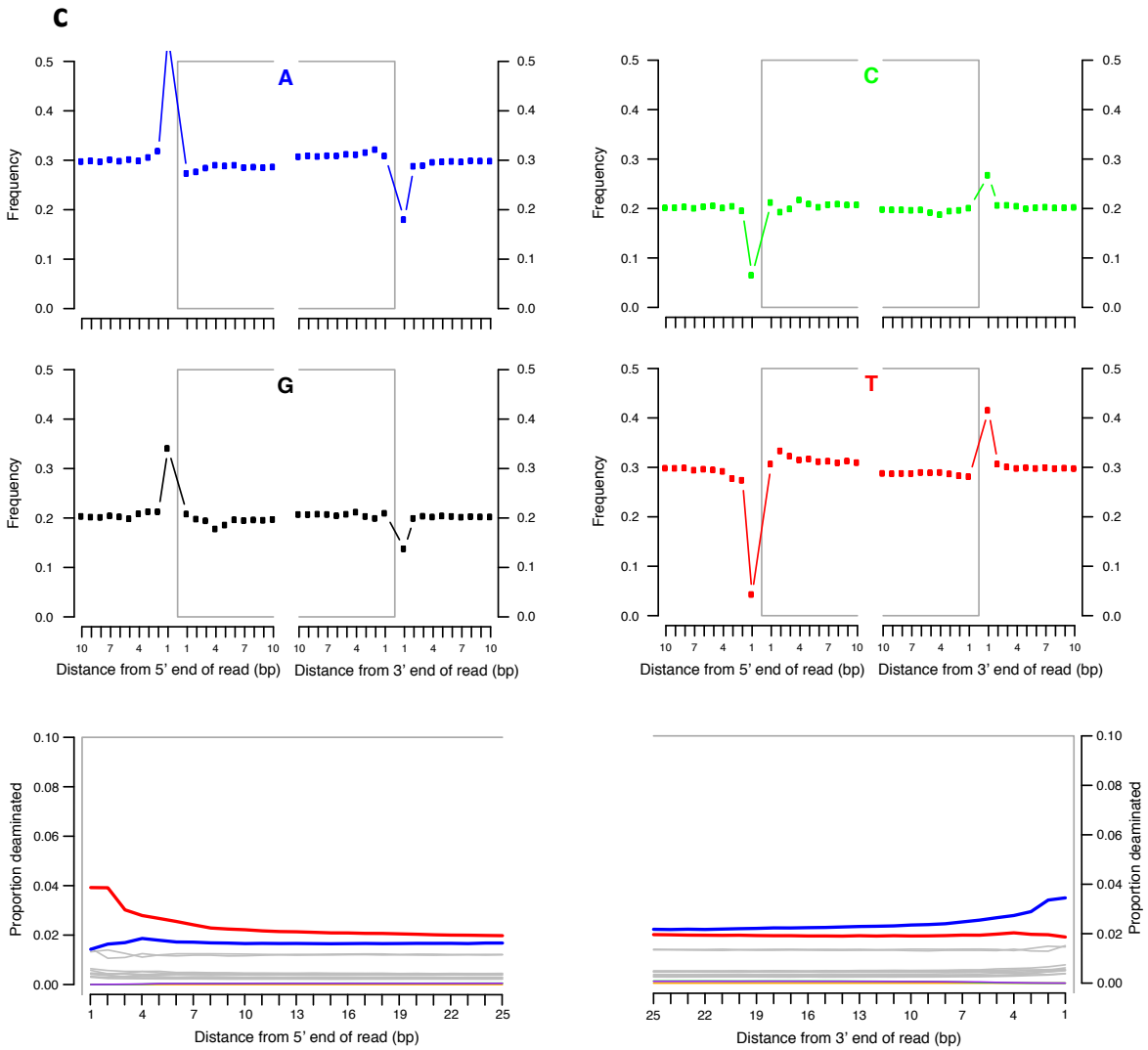


Figure S11. Characterization of damage patterns in genomic DNA from passenger pigeons based on mapping to the band-tailed pigeon genome: (A) BMNH794, (B) BMNH1149, (C) ROM 34.3.23.2, and (D) ROM 40360,. The increased frequency of purines (guanine, G; adenine, A) immediately, upstream of the 5', and a corresponding increase in pyramidines (cytosine, C; thymine, T) immediately downstream of the 3', ends of reads is consistent with depurination-induced fragmentation. The lower plots show an increasing proportion of cytosines that are deaminated toward the read ends (C→T: red, G→A: blue).







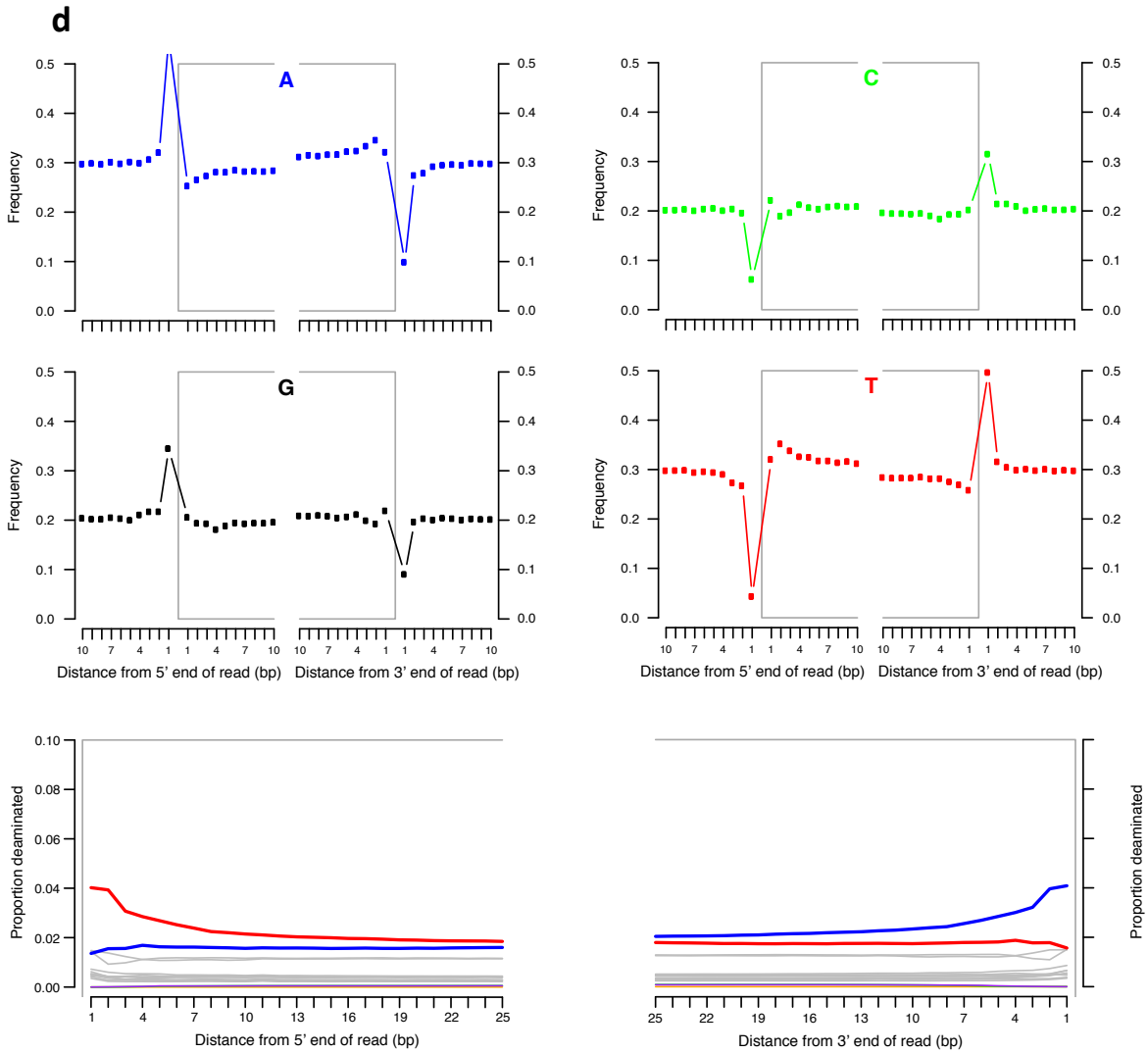


Figure S12. Estimation of error from heterozygosity on the z-chromosome of female samples. Plots of heterozygosity in the four passenger pigeon samples (red) and two band-tailed pigeon samples (blue), with females represented by solid lines and males by dashed lines. While across most of the genome heterozygosity is similar among different individuals, across a region that mostly maps to the chicken z-chromosomes, females show much less heterozygosity than males. This region is therefore likely to be the pigeon z-chromosome, and heterozygosity in the females is therefore likely to represent error in our calling of variants. In this region, the female band-tailed pigeon has an average heterozygosity of 3.8×10^{-4} differences/site and the female passenger pigeons have an average heterozygosity of 8.0×10^{-4} differences/site.

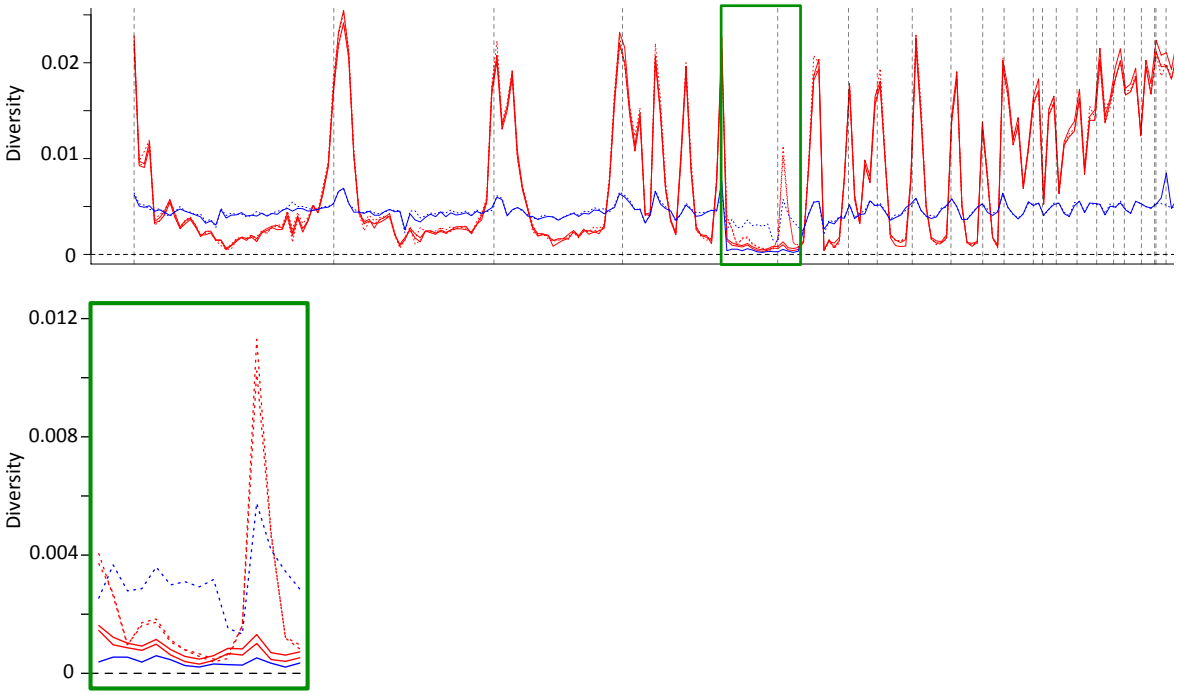


Figure S13. Transition to transversion rates (Ts/Tv): for passenger pigeons (red) and band-tailed pigeons (blue), for 5 Mb windows along our scaffolds, ordered according to our mapping to the chicken genome. Horizontal dashed lines represent chromosome boundaries in the chicken genome.

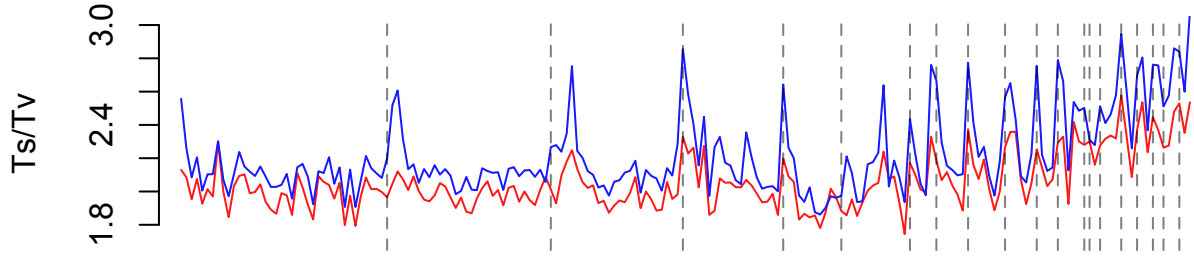


Figure S14. Comparisons of codon usage bias statistics across passenger pigeons and band-tailed pigeons. Histograms of summary statistics of codon usage bias for individual genes in passenger pigeons (red; A, C, E) and band-tailed pigeons (blue; A, C, E), and the difference between these statistics in the two species (B, D, F).

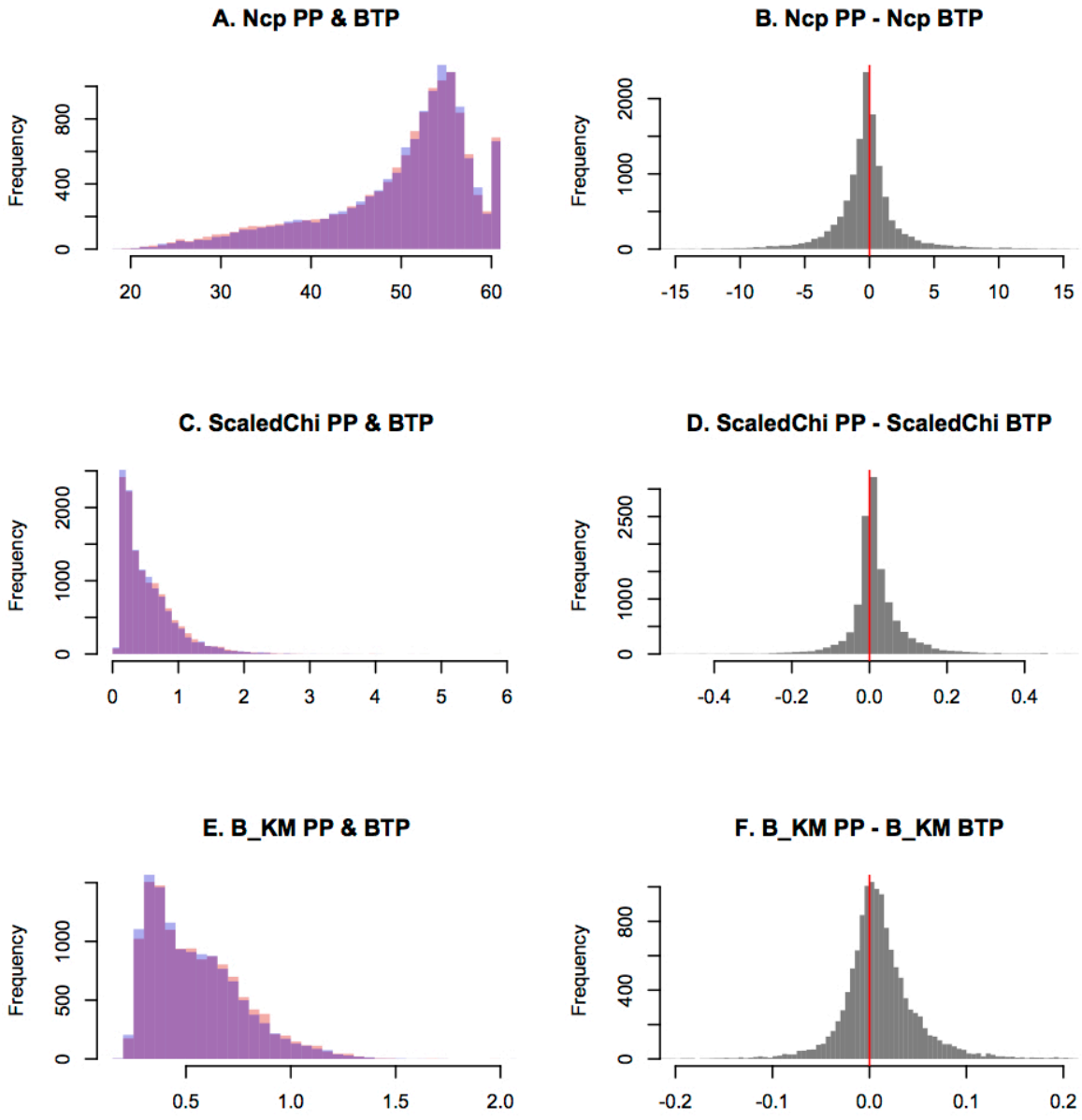


Figure S15. Comparisons of codon usage bias statistics across high- and low-diversity regions of the genome. Histograms of summary statistics of codon usage bias for individual genes in passenger pigeons (A, D, G) and band-tailed pigeons (B, E, H), and the difference between these statistics in the two species (C, F, I), for genes in high-diversity regions (green) and in low-diversity regions (blue) of the passenger pigeon genome.

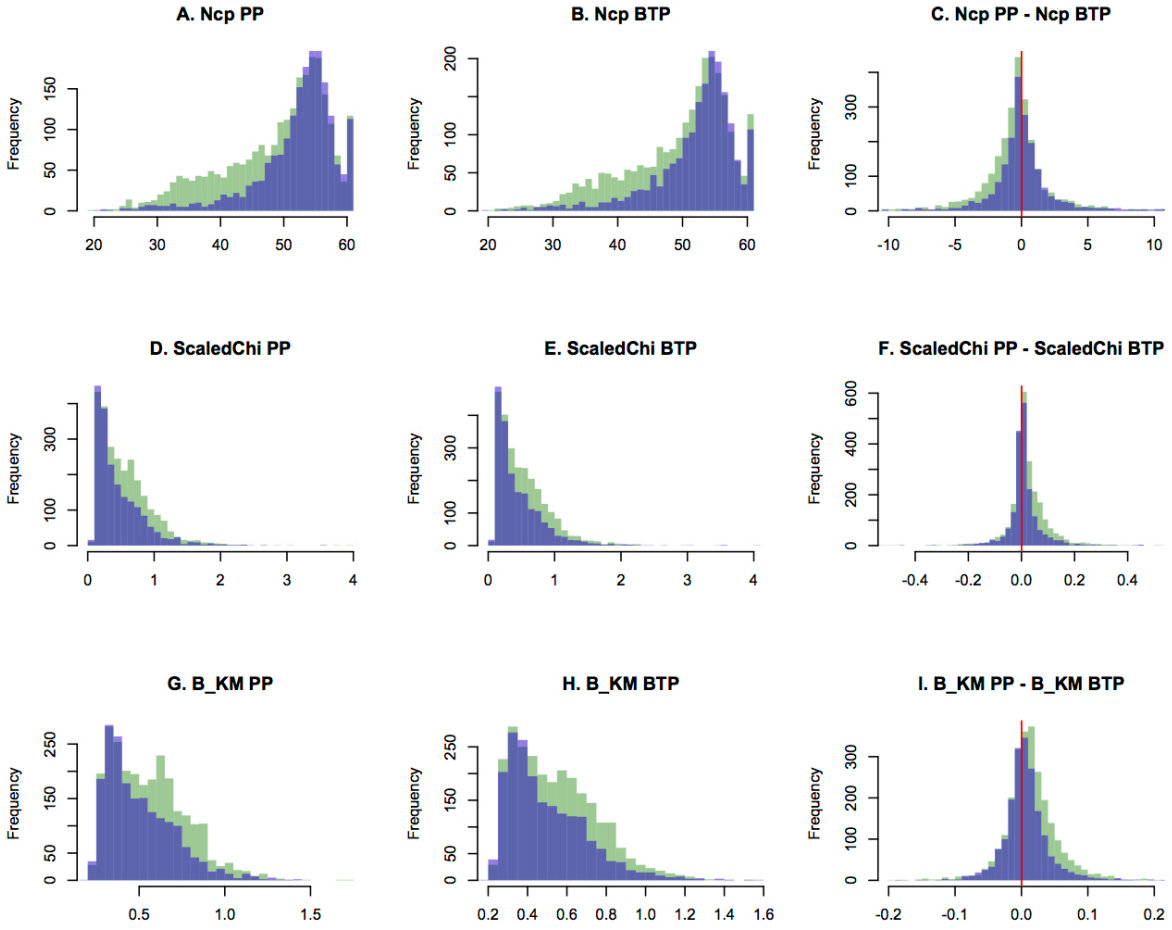
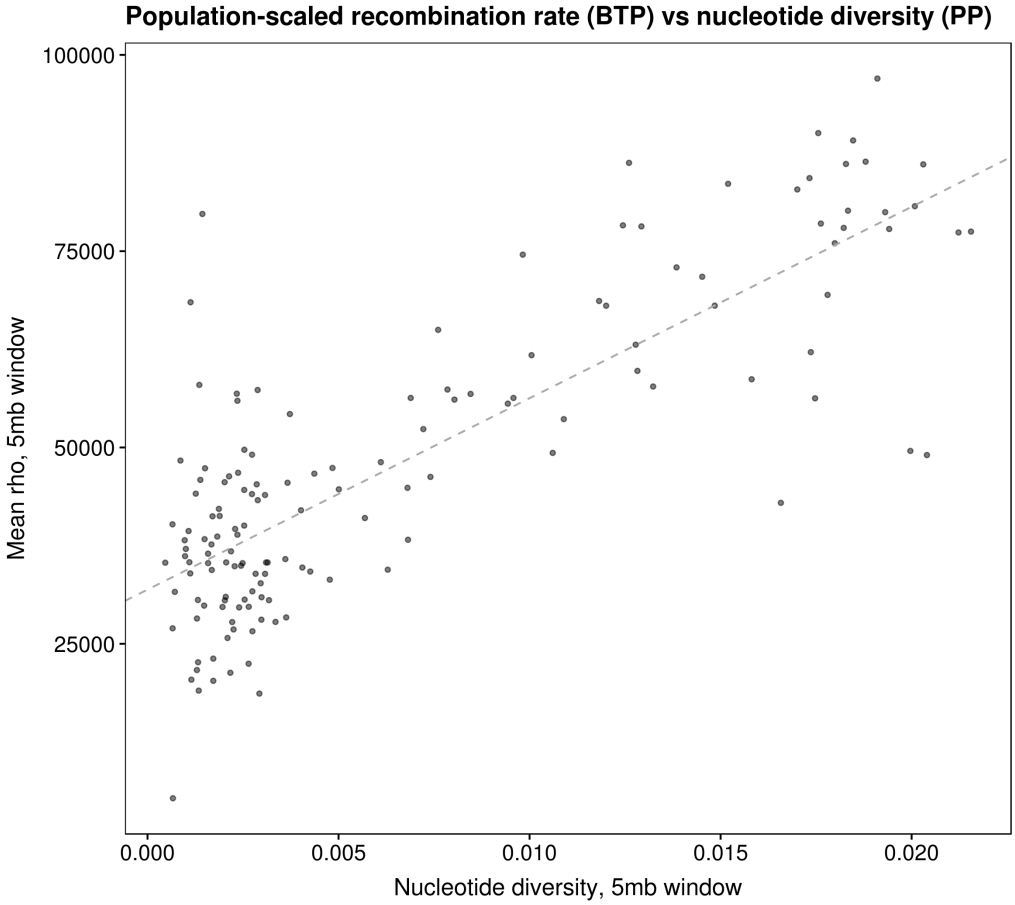


Figure S16. Population-scaled recombination rate (*ρ*) estimates from band-tailed pigeons using LDhat. (A) Mean *ρ* (in band-tailed pigeons) per 5 Mb window plotted against mean nucleotide diversity in passenger pigeons per 5 Mb window. (**B**) The same data shown mapped to the chicken genome, with *ρ* in blue and nucleotide diversity in red. Both recombination rate and nucleotide diversity tend to increase near chromosome boundaries, but *ρ* estimates are noisier, which may be due to the uncertainty inherent in quantifying linkage between variants observed in only a small number (4) of haplotypes.

A



B

Mean population–scaled recombination rate (BTP) and nucleotide diversity (PP)

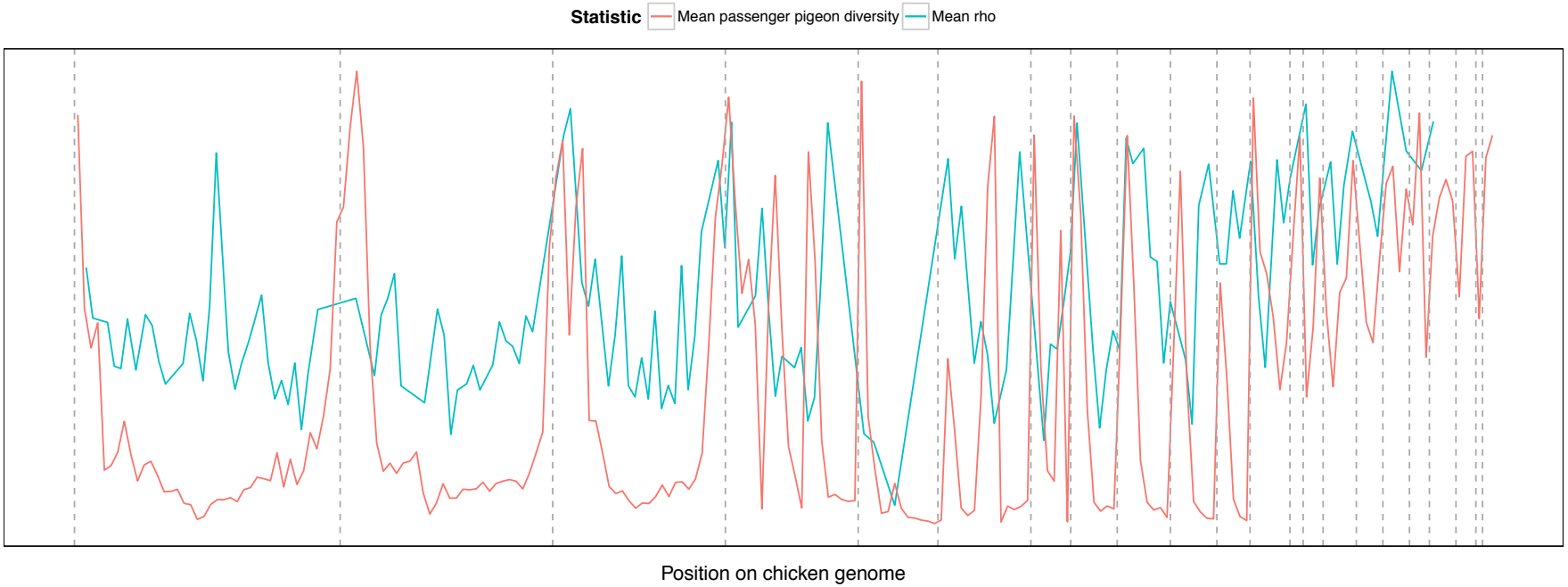


Figure S17. PSMC results for the passenger pigeons ROM 34.3.23.2 and ROM 40360. These plots show *Ne* over time, with the x axis indicating years before the present, scaled according to a generation time of 4 years(*82*) and a mutation rate of 5.68×10^{-9} substitutions/site/generation. Every PSMC plot includes 100 bootstrap replicates. A and B) PSMC plots showing the results of *Ne* over time according to the whole genome; C and D) PSMC plots for the high-diversity regions; E and F) PSMC plots for the low-diversity regions.

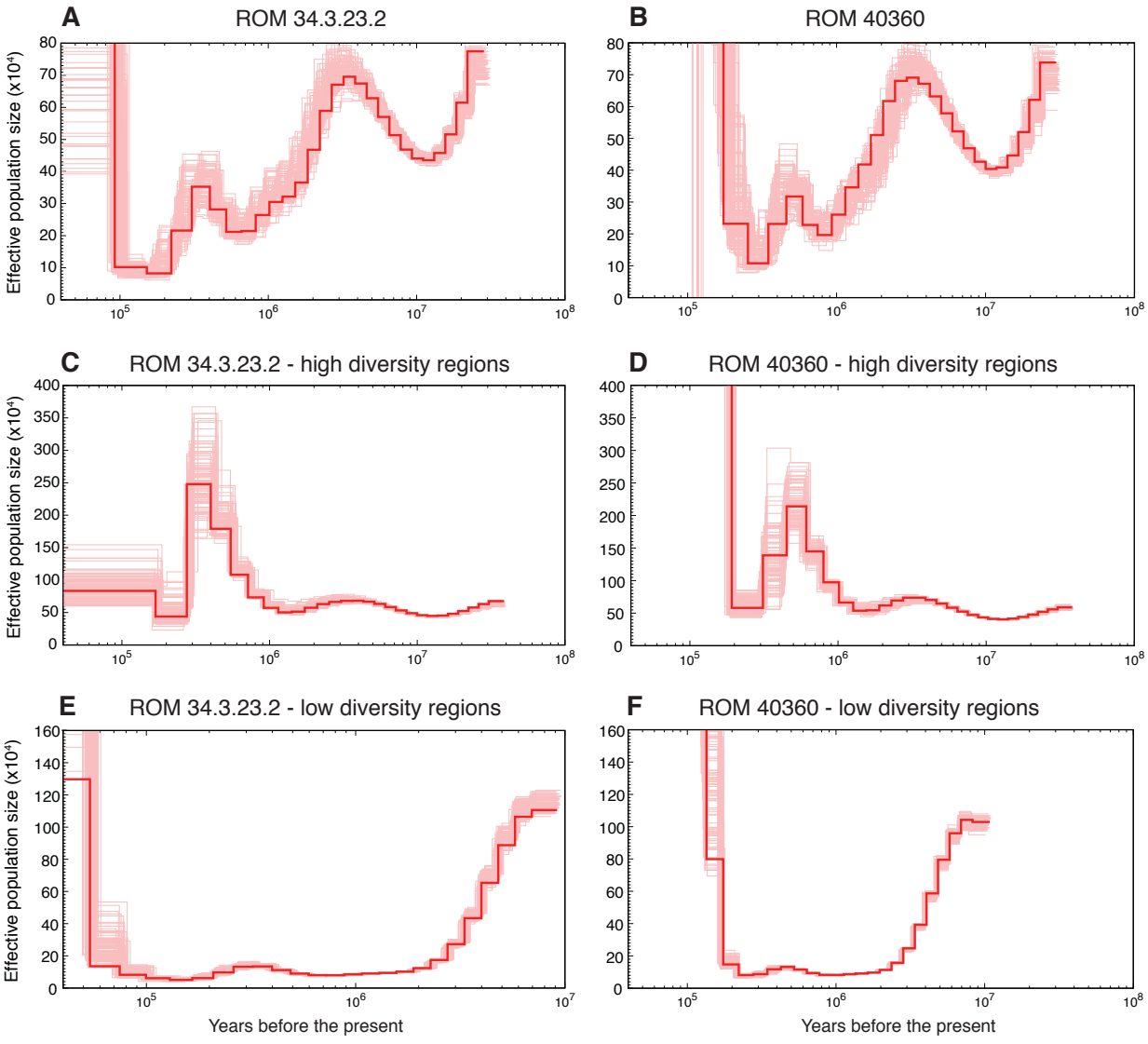


Figure S18. PSMC results for the passenger pigeons BMNH794 and BMNH1149. These plots show *Ne* over time, with the x axis indicating years before the present, scaled according to a generation time of 4 years(*82*) and a mutation rate of 5.68×10^{-9} substitutions/site/generation. Every PSMC plot includes 100 bootstrap replicates. A and B) PSMC plots showing the results of *Ne* over time according to the whole genome; C and D) PSMC plots for the high-diversity regions; E and F) PSMC plots for the low-diversity regions.

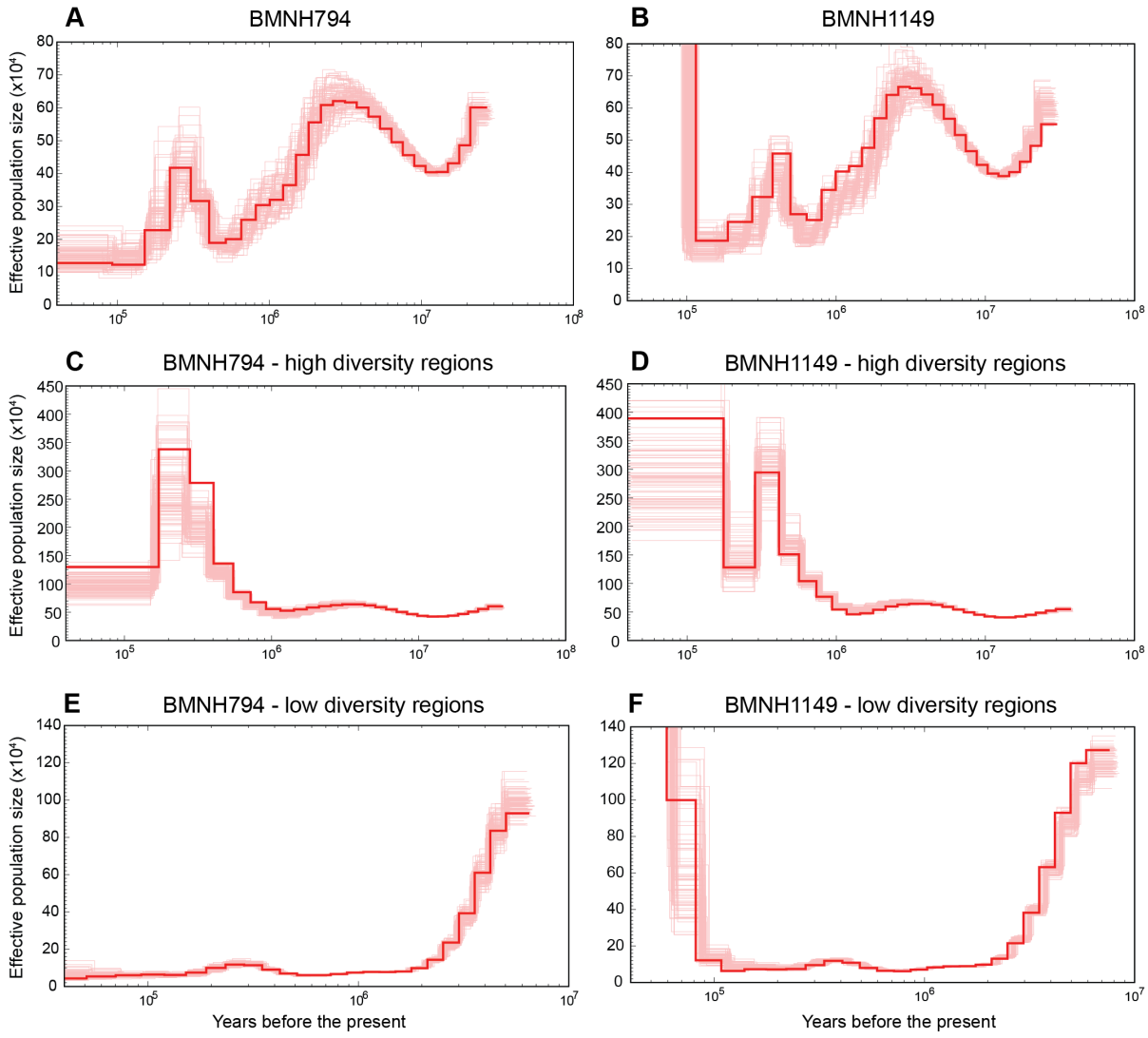


Figure S19. PSMC results for the whole-genome of passenger pigeon ROM 34.3.23.2 using two different parameter choices. The red line represents the result of the PSMC analysis when using the parameters described in Hung *et al. (4)* and the green line represents the result when using parameters that are more appropriate for this data (as discussed above), and as used in the analyses shown in figures S14-S16.

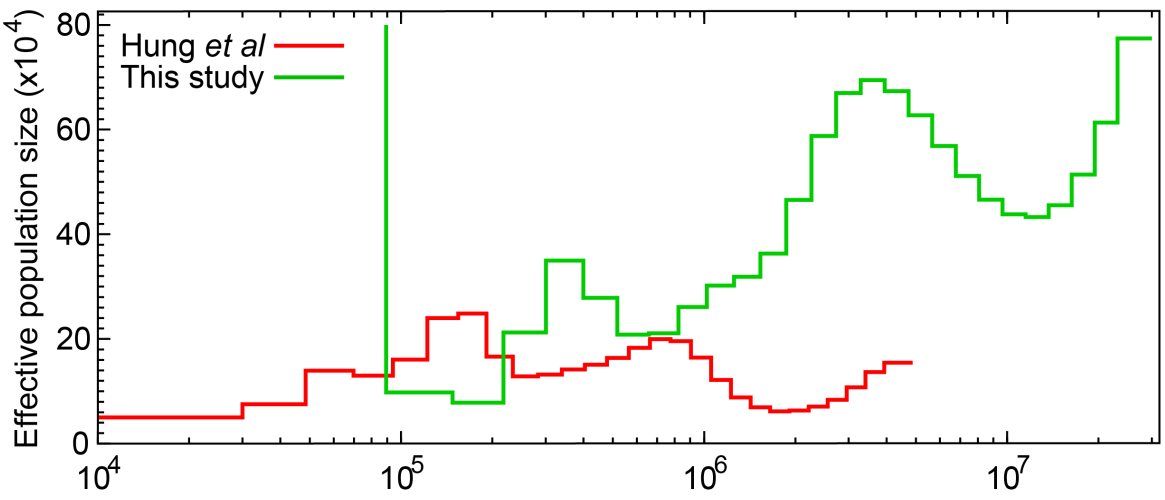


Figure S20. PSMC results for the band-tailed pigeons individuals AMNH DOT 14025 (the reference genome) and BTP2013. These plots show *Ne* over time, with the x axis indicating years before the present, scaled according to a generation time of 4 years(*82*) and a mutation rate of 5.68×10^{-9} substitutions/site/generation. Every PSMC plot includes 100 bootstrap replicates. A and B) PSMC plots showing the results of *Ne* over time according to the whole genome; C and D) PSMC plots for the high-diversity regions; E and F) PSMC plots for the low-diversity regions. For BP2013, which was bred in captivity, *Ne* may be overestimated in low-diversity regions (the majority of the genome) due to recent outbreeding with a distantly related individual.

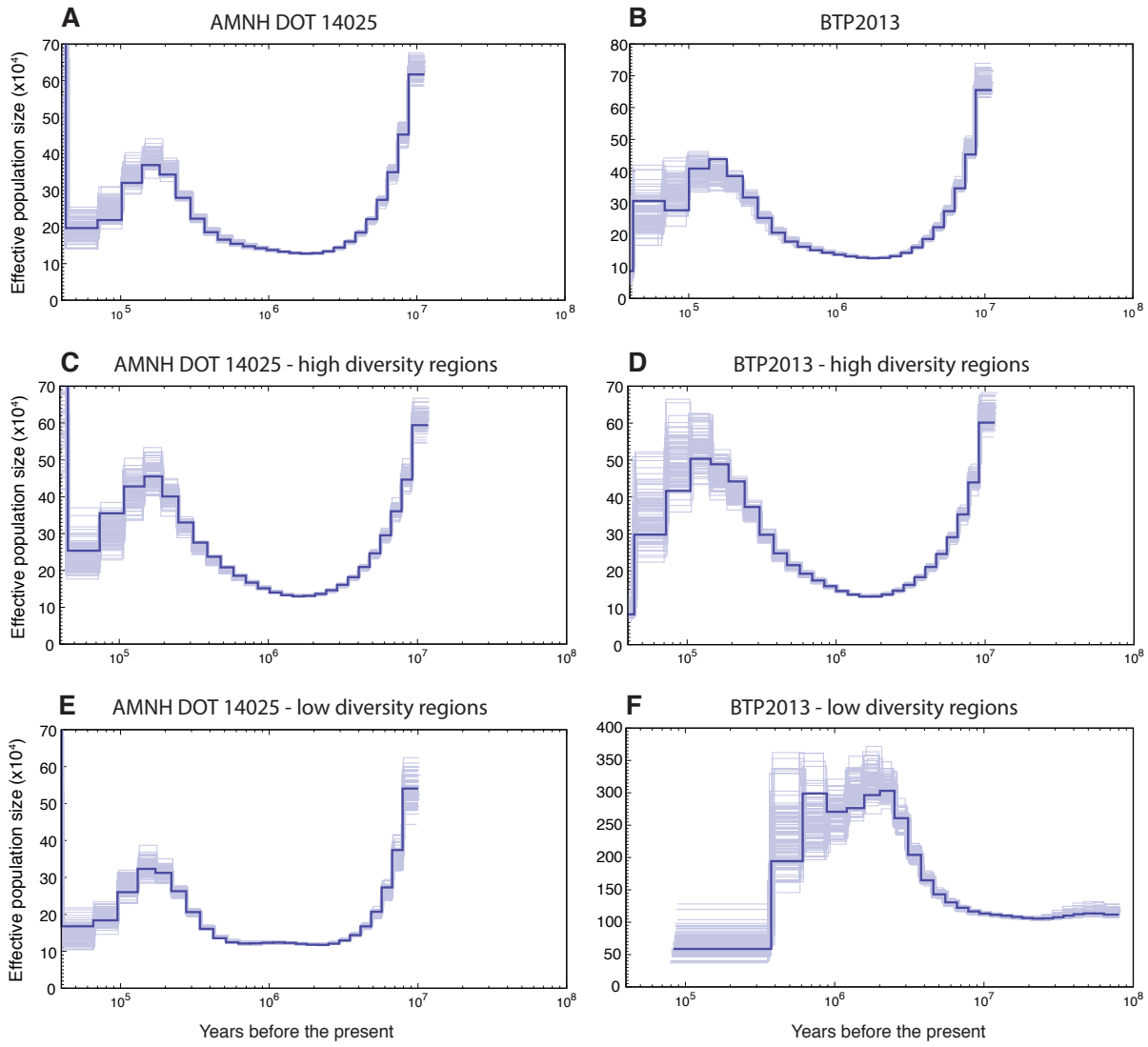


Figure S21. Comparisons of gene count, average gene length, and gene density and genetic diversity for 5 Mb windows across the passenger pigeon genome. Both gene count (Spearman's rank correlation; ρ = 0.51, $p = 1.2 \times 10^{-15}$) and gene density (ρ = 0.31, $p = 4.4 \times 10^{-6}$) are positively correlated with genetic diversity in passenger pigeons, whereas gene length is negatively correlated with genetic diversity in passenger pigeons (ρ = -0.47, $p < 2.2 \times 10^{-16}$).

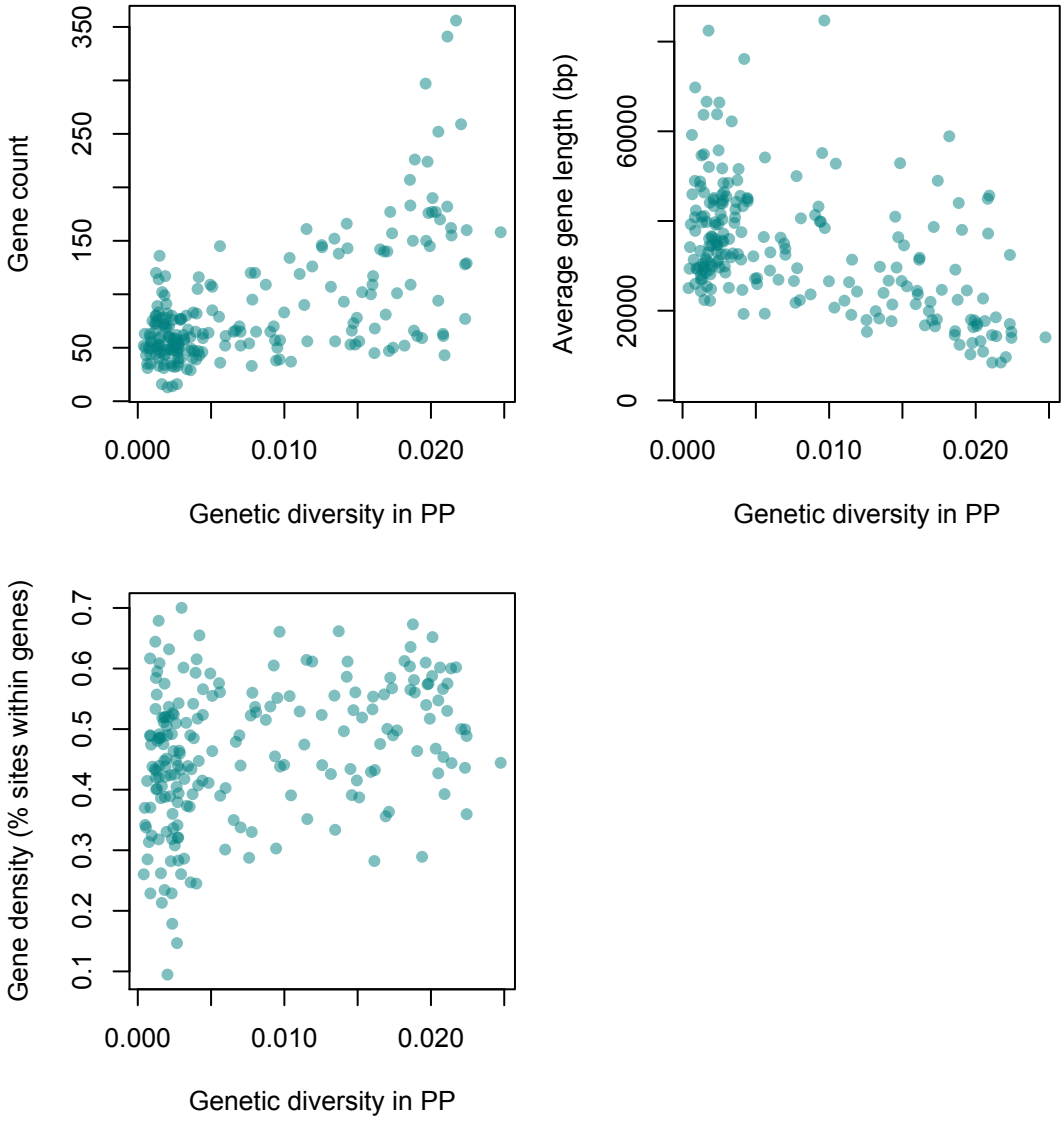


Figure S22. Estimates of Tajima's D and Fay and Wu's H-statistic: for synonymous (circles) and nonsynonymous (triangles) mutations plotted against estimates of relative neutral diversity (diversity in the passenger pigeon divided by diversity in the band-tailed pigeon) for 5 Mb windows across the genome.

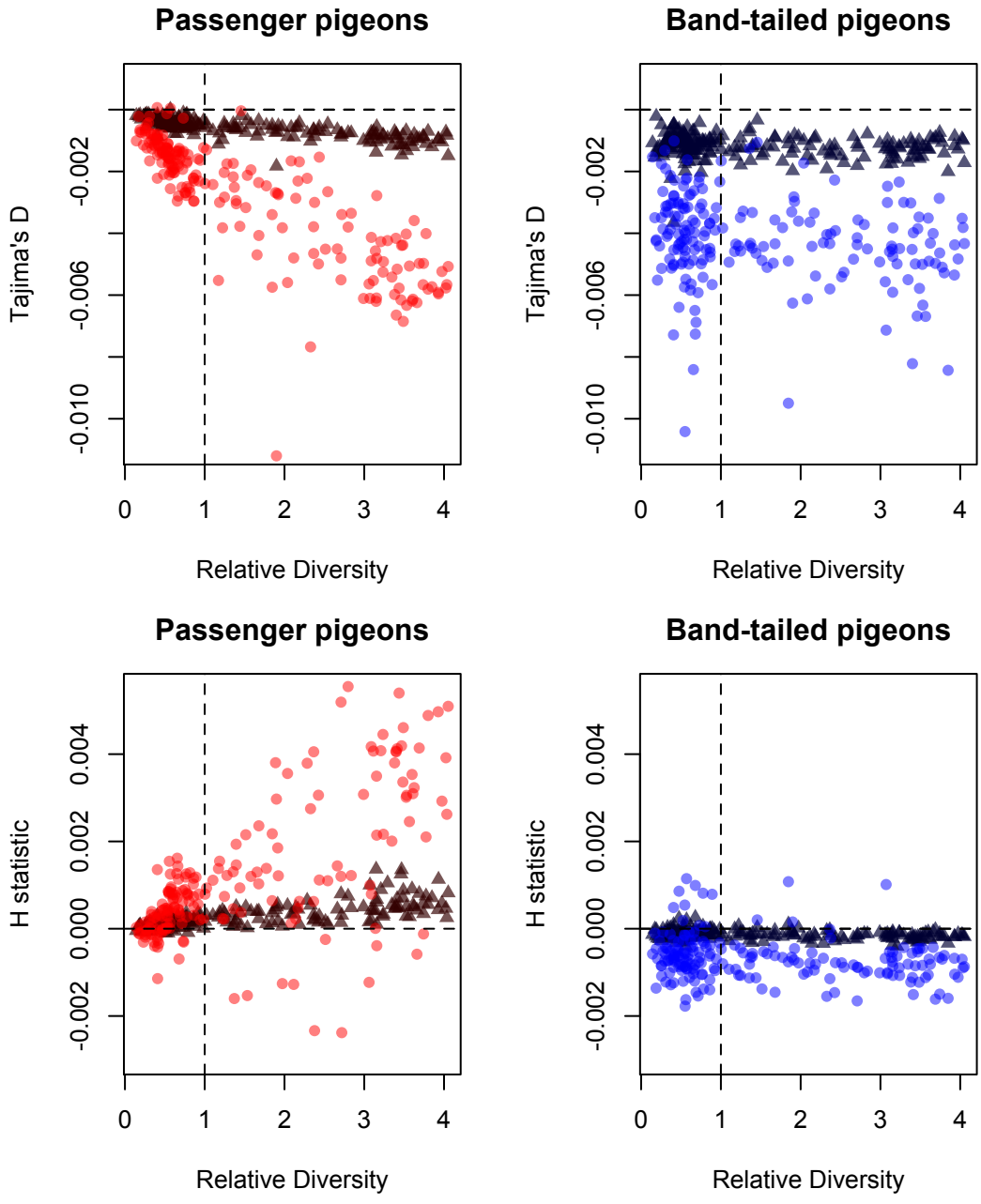


Figure S23. **The ratio of nonsynonymous to synonymous counts of fixed differences for the 32 genes identified as showing evidence of adaptive substitution.** Plots show the ratio of nonsynonymous to synonymous counts of fixed differences (A/S) along the passenger pigeon lineage (PP, red) and band-tailed pigeon lineage (BTP, blue) for the 32 genes identified as showing evidence of adaptive substitution ('M-K test genes', filled boxes) and for all other genes ('Other genes', empty boxes).

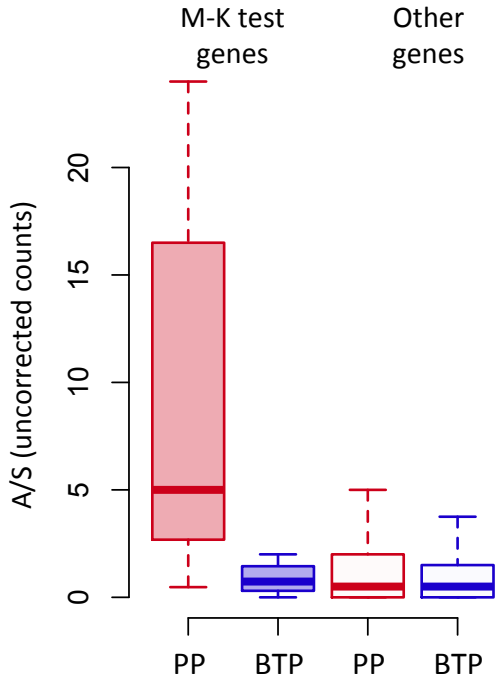


Table S2. Nuclear genomes. Number of reads, % duplicates, number and % of mapped reads all calculated using the samtools flagstat function. We calculated the number of called sites by accepting sites called within a 97.5% of depth coverage for each sample, plus a minimum variant quality (VQ) of 50 and minimum genotype quality (GQ) of 30.

Sample ID	No. reads	% duplicates	No. mapped reads*	% mapped	Number of called sites (bp)	Median coverage
ROM 34.3.23.2	790,186,760	14.95	645,129,546	93.80	956,718,155	51
ROM 40360	791,517,028	18.60	611,508,990	92.00	882,931,867	41
BMNH749	650,804,970	36.82	420,746,316	88.98	720,206,927	13
BMNH1149	598,360,026	55.76	351,344,034	90.73	952,051,353	34
BTP2013†	358,314,562	13.97	300,775,255	93.85	1,006,217,837	25

* = after duplicates removed, † = band-tailed pigeon

Table S3. The genes with evidence of adaptive evolution in passenger pigeons. McDonald-Kreitman tests on individual genes (see Materials and Methods) yielded significant evidence of adaptive substitution or recent increased constraint (i.e. $dN/dS > pN/pS$) for 11 genes after correcting for multiple testing using a conservative Bonferroni correction ($p < 0.05$) and for 32 genes after correcting for a false discovery rate of 5% ($p < 0.05$). These 32 genes are described here. Gene name, function and the biological process GO function were identified through BLAST searches on GenBank, literature searches and online gene databases. The reported *p*-value is uncorrected (18,708 genes were tested). The numbers of nonsynonymous (N) and synonymous (S) derived substitutions (Subs) and polymorphisms (Poly) identified in the passenger pigeon (PP) and band-tailed pigeon (BTP) samples are also described. A comparison of the ratio of nonsynonymous to synonymous counts of fixed differences for the 32 genes identified as showing evidence of adaptive substitution, and other genes, is shown in figure S23.

Gene name	**Function**	**GO function: *Biological Process***	***p*-value**		**PP**		**BTP**	
					N	**S**	**N**	**S**
STAB1 Stabilin-1	Involved in determining immune cell trafficking in physiological and inflammatory conditions (*99*).	cell-matrix adhesion; signal transduction	6.2×10^{-12}	**Subs**	32	10	23	12
				Poly	10	67	7	16
CPD Carboxypeptidase D	A receptor molecule for the uptake of duck hepatitis B virus (*18*).	cellular process; nitrogen compound metabolic process; protein metabolic process	7.5×10^{-11}	**Subs**	11	2	5	13
				Poly	2	80	7	9
SI Sucrase-isomaltase, intestinal	Plays an important role in the final stage of carbohydrate	carbohydrate metabolic process, polysaccharide digestion	8.2×10^{-11}	**Subs**	33	11	17	10
				Poly	14	71	6	9

	digestion and associated with different diets in birds (*19*).							
HELZ2 Helicase with zinc finger domain 2	A nuclear transcriptional coactivator for several nuclear receptors, including ones involved in metabolism in the liver and intestine.	DNA replication, RNA catabolic process, mRNA splicing, regulation of transcription from RNA polymerase II, tRNA metabolic process	$3.5x10^{-8}$	**Subs**	26	11	25	21
				Poly	28	107	7	18
ALG13 Putative bifunctional UDP-N-acetylglucosamine transferase and deubiquitinase		dolichol-linked oligosaccharide biosynthetic process	$3.\ 5x10^{-8}$	**Subs**	22	2	2	5
				Poly	7	28	6	2
FAAH Fatty acid amide hydrolase	Breaks down anandamide ('bliss molecule'), which may reduce anxiety and influences feeding behavior (*20*, *100*).	fatty acid metabolic process, tRNA aminoacylation for protein translation	$1.8x10^{-7}$	**Subs**	15	1	2	3
				Poly	3	22	1	3
PRCP Lysosomal Pro-X carboxypeptidase	Associated with blood clotting strength in birds (*101*).	proteolysis	$2.0x10^{-7}$	**Subs**	9	0	1	4
				Poly	3	34	1	2
DNAH1	Force	cellular	$4.9x10^{-7}$	**Subs**	17	4	13	16

Dynein heavy chain 1, axonemal	generating protein of respiratory cilia. Involved in sperm motility and implicated in sperm flagellar assembly (*102*, *103*).	component morphogenesis, cellular component movement, chromosome segregation, fertilization, intracellular protein transport, mitosis, spermatogenesis, vesicle-mediated transport		**Poly**	22	79	10	9
NEB Nebulin	A giant muscle protein that may be involved in maintaining the structural integrity of sarcomeres and the membrane system associated with the myofibrils (*104*).	muscle filament sliding, muscle organ development, regulation of actin filament length, somatic muscle development	5.2×10^{-7}	**Subs**	18	38	5	26
				Poly	16	247	6	19
CDHR2 Cadherin-related family member 2	Intermicrovillar adhesion molecule. Plays a central role in microvilli and epithelial brush border differentiation.	cellular process, heart development, muscle organ development, nervous system development, sensory perception of sound, visual perception	7.6×10^{-7}	**Subs**	21	3	12	12
				Poly	20	48	2	11

DNAH1 Dynein heavy chain 1, axonemal *(this is a distinct gene in our assembly, but its closest match in other annotations is identical to the above gene)*			$1.0x10^{-6}$	**Subs**	15	4	12	28
				Poly	49	172	16	24
TDRD6 Tudor domain-containing protein 6	Involved in spermiogenesis, chromatoid body formation and mature miRNA expression (*84*).	nucleobase-containing compound metabolic process, multicellular organism development, spermatogenesis, cell differentiation	$3.5x10^{-6}$	**Subs**	31	5	16	18
				Poly	109	132	21	14
SOAT1 Sterol O-acyltransferase 1	Catalyzes the formation of fatty acid-cholesterol esters. Plays a role in lipoprotein assembly and dietary cholesterol absorption (*105*).	Cholesterol metabolic process	$3.8x10^{-6}$	**Subs**	9	2	3	0
				Poly	0	19	1	5
APOB Apolipoprotein B-100	Primary lipoprotein component of low-density lipoprotein.	Cholesterol transporter activity, lipid binding	$4.0 x10^{-6}$	**Subs**	24	12	24	30
				Poly	48	139	5	13

	Expression associated with feed restriction in chickens (*106*). It enables the transport of fat molecules in blood plasma and lymph and facilitates the movement of molecules such as cholesterol into cells (*107*). Has been identified as a target of selection in polar bears (*108*).							
DENND5A DENN domain-containing protein 5A	Guanine nucleotide exchange factor.	Cellular process, response to abiotic stimulus, response to external stimulus	4.0×10^{-6}	**Subs**	4	0	1	9
				Poly	0	47	0	8
PRKAA1 5'-AMP-activated protein kinase catalytic subunit alpha-1	Catalytic subunit of AMP-activated protein kinase (AMPK), an energy sensor protein kinase that plays a key role in regulating cellular energy metabolism (*109*). Also	Intracellular signal transduction, phosphate-containing compound metabolic process biological regulation, response to stimulus	4.2×10^{-6}	**Subs**	8	6	3	11
				Poly	6	110	5	19

	implicated in sexual differentiation in chickens (*110*).							
HNF4A Hepatocyte nuclear factor 4-alpha	Transcriptionally controlled transcription factor. May be essential for development of the liver, kidney and intestine. Associated with type 2 diabetes in humans (*111*).	cell cycle, lipid metabolic process, sensory perception, visual perception, vitamin metabolic process, sex differentiation	9.5×10^{-6}	**Subs**	10	0	3	5
				Poly	6	24	2	6
NUP160 Nuclear pore complex protein Nup160	Involved in mRNA transport (*112*). Associated with reproductive isolation in *Drosophila*, where there is evidence of positive selection (*113*).	cellular process, nuclear transport	1.1×10^{-5}	**Subs**	6	3	1	14
				Poly	1	50	2	5
FAT4 Protocadherin Fat 4	Cadherins are calcium-dependent cell adhesion proteins. Plays a role in neurogenesis and in planar cell polarity and Hippo signalling recapitulating	Nervous system development	1.4×10^{-5}	**Subs**	29	9	22	17
				Poly	93	151	12	10

	(*114*, *115*).							
Mep1a Meprin A subunit alpha	Hydrolysis of protein and peptide substrates preferentially on carboxyl side of hydrophobic residues. Associated with insulin metabolism and diabetic nephropathy in humans (*116*, *117*).	angiogenesis, blood coagulation, ectoderm development localization, endocytosis, heart development, intracellular protein transport, nervous system development, proteolysis, sensory perception, skeletal system development, synaptic transmission, visual perception, vitamin transport	$1.4x10^{-5}$	**Subs**	24	1	7	7
				Poly	18	22	6	6
CYP2U1 Cytochrome P450 2U1	Catalyzes the hydroxylation of fatty acids, including arachidonic acid. May modulate the arachidonic acid signaling pathway and play a role in other fatty acid signaling processes.	omega-hydroxylase P450 pathway	$1.5x10^{-5}$	**Subs**	5	0	0	6
				Poly	0	21	0	0
ABCA5	May play a role	catabolic	$1.5x10^{-5}$	**Subs**	10	16	6	12

ATP-binding cassette sub-family A member 5	in lysomoal trafficking and is thought to be expressed in skeletal muscle, kidney, liver and placenta (*118*).	process, cellular process, lipid transport, nitrogen compound metabolic process, nucleobase-containing compound metabolic process, phosphate-containing compound metabolic process, cholesterol efflux, reverse cholesterol transport		**Poly**	7	133	3	11
ABCC3 ATP binding cassette subfamily C member 3	May act as an inducible transporter in the biliary and intestinal excretion of organic anions. Acts as an alternative route for the export of bile acids and glucuronides from cholestatic hepatocytes (By similarity). Has been associated with	extracellular transport, immune system process, response to toxic substance	$1.7x10^{-5}$	**Subs**	13	19	5	11
				Poly	5	81	2	5

	response to avian influenza infection in humans(*119*)							
COL22A1 Collagen alpha-1 (XXII) chain	A structural protein. Associated with serum creatinine level (*120*), a biomarker for kidney function.		$1.8x10^{-5}$	**Subs**	6	0	2	9
				Poly	12	80	5	9
MSLN Mesothelin	Membrane-anchored forms may play a role in cellular adhesion. Associated with host-response to avian influenza virus in the chicken lung (*121*).		$2.7x10^{-5}$	**Subs**	18	1	12	6
				Poly	46	58	14	9
SLC13A2 Solute Carrier Family 13 Member 2	A sodium-coupled citrate transporter. The encoded protein may play a role in the formation of kidney stones.		$3.6x10^{-5}$	**Subs**	8	0	4	1
				Poly	4	21	4	1
PABPN1L Poly(A) Binding Protein Nuclear 1 Like, Cytoplasmic	Binds the poly(A) tail of mRNA.		$3.7x10^{-5}$	**Subs**	11	2	7	4
				Poly	3	20	1	2
VTN	Vitronectin is a	cell-matrix	$3.8x10^{-5}$	**Subs**	7	3	1	3

Vitronectin	cell adhesion and spreading factor found in serum and tissues. Vitronectin interact with glycosaminoglycans and proteoglycans. Is recognized by certain members of the integrin family and serves as a cell-to-substrate adhesion molecule.	adhesion, extracellular matrix organization, immune response, liver regeneration, oligodendrocyte differentiation, positive regulation of cell-substrate adhesion, protein polymerization		**Poly**	4	53	1	3
INPP5B Type II inositol 1,4,5-trisphosphate 5-phosphatase	Encodes the type II 5-phosphatase. The protein is localized to the cytosol and mitochondria. Associated with reduced ability of sperm to fertilise eggs in mice (*122*).	flagellated sperm motility, in utero embryonic development, regulation of protein processing, spermatogenesis	$3.8x10^{-5}$	**Subs**	6	2	0	2
				Poly	0	23	0	2
INPP5B Type II inositol 1,4,5-trisphosphate 5-phosphatase *(this is a distinct gene in our assembly, but its*			$4.1x10^{-5}$	**Subs**	28	4	12	3
				Poly	37	45	9	9

closest match in other annotations is identical to the above gene)								
CFAP57 Cilia- and flagella-associated protein 57			$8.0x10^{-5}$	**Subs**	14	4	8	5
				Poly	24	65	9	4
NCAPH Condensin complex subunit 2	Regulatory subunit of the condensin complex, a complex required for conversion of interphase chromatin into mitotic-like condensed chromosomes. Associated with susceptibility to bacterial infection (*123*).		$8.2x10^{-5}$	**Subs**	13	2	7	6
				Poly	16	39	1	4

Table S4. McDonald-Kreitman test for neutral evolution of variants present in the passenger pigeon mitochondrial protein-coding genes. Counts of sites that are non-synonymous (NS) and synonymous (S) polymorphisms within passenger pigeons and sites that are non-synonymous and synonymous fixed differences between passenger and band-tailed pigeons, and the corresponding ratios (accounting for differences in mutational opportunities for nonsynonymous and synonymous change).

	NS	**S**	**Ratio**
Polymorphism	32	131	0.07
Divergence	153	979	0.04

Table S5. Comparison of variants at high and low frequency in the passenger pigeon mitochondrial protein-coding genes. Counts of non-synonymous (NS) and synonymous (S) polymorphic sites at low (1 or 2 individuals) and high (40 or 41 individuals) frequencies in passenger pigeons, and the corresponding ratios (accounting for differences in mutational opportunities for nonsynonymous and synonymous change).

	NS	**S**	**Ratio (NS/S)**
Low frequency	26	73	0.10
High frequency	4	49	0.02

Table S6. D-statistic Tests for variation in shared derived alleles between passenger pigeons and band-tailed pigeons. Positive D values indicate that the P2 and P3 individuals share an excess of shared derived alleles (ABBA sites) compared to the P1 and P3 individuals (BABA sites). Weighted block jackknife standard errors are calculated from 5Mb non-overlapping blocks and Z scores (D/standard error) greater than 3 are considered significant. Because the passenger pigeon samples are impacted by cytosine deamination damage, we excluded transition differences from our counts of ABBA and BABA sites.

D(Passenger, Passenger, Band Tailed, Rock Dove)								
P1	**P2**	**P3**	**Outgroup**	**D**	**Weighted block jackknife standard error**	**Z score**	**ABBA sites**	**BABA sites**
BMNH794	ROM 40360	BTP2013	*C. livia*	0.166277	0.015543	10.698.148	72350	51720
BMNH794	ROM 40360	AMNH DOT 14025	*C. livia*	0.147093	0.014584	10.085.934	68977	51287
BMNH1149	ROM 40360	BTP2013	*C. livia*	0.099265	0.008906	11.145.796	72895	59730
ROM 34.3.23.2	ROM 40360	BTP2013	*C. livia*	0.095895	0.009461	10.136.086	71300	58822
ROM 34.3.23.2	ROM 40360	AMNH DOT 14025	*C. livia*	0.088930	0.00911	9.761.491	68265	57115
BMNH1149	ROM 40360	AMNH DOT 14025	*C. livia*	0.087512	0.008385	10.436.869	69815	58579
BMNH794	ROM 34.3.23.2	BTP2013	*C. livia*	0.077501	0.009257	8.372.189	61778	52891
BMNH794	BMNH1149	BTP2013	*C. livia*	0.066861	0.00841	7.950.513	61416	53718
BMNH794	ROM 34.3.23.2	AMNH DOT 14025	*C. livia*	0.064494	0.008612	7.488.762	59386	52190
BMNH794	BMNH1149	AMNH DOT 14025	*C. livia*	0.058168	0.00826	7.042.337	59705	53141
BMNH1149	ROM 34.3.23.2	BTP2013	*C. livia*	0.009859	0.004666	2.113.111	62377	61159
BMNH1149	ROM 34.3.23.2	AMNH DOT 14025	*C. livia*	0.004961	0.00472	105.122	60260	59665
AMNH DOT 14025	BTP2013	BMNH794	*C. livia*	-0.058011	0.009662	-6.004.283	12211	13715

AMNH DOT 14025	BTP2013	BMNH1149	*C. livia*	-0.016825	0.010368	-1.622.813	13645	14112
AMNH DOT 14025	BTP2013	ROM 34.3.23.2	*C. livia*	-0.005319	0.010259	-0.518484	14025	14175
AMNH DOT 14025	BTP2013	ROM 40360	*C. livia*	0.029907	0.011372	2.629.996	15066	14191

Table S7. $\hat{f}$ estimates of band-tailed pigeon ancestry in passenger pigeons. All comparisons use the BMNH794 individual as a baseline unadmixed passenger pigeon (P1). While the other three passenger pigeons exhibit significantly more band-tailed pigeon derived allele sharing (Z=12.4 to Z=21.9) the total excess passenger pigeon ancestry accounts for only 0.23% to 0.61% of the more admixed individual's genome. Thus, while some admixture is consistent with these results it accounts for a small portion of the passenger pigeon's total diversity.

P1 (unadmixed)	P2 (admixed)	P3 (introgressor)	P4 (introgressor)	Outgroup	$\hat{f}$	Weighted block jackknife error	Z score
BMNH794	BMNH1149	AMNH DOT 14025	BTP2013	*C. livia*	0.23%	0.000184	12.388118
BMNH794	ROM 34.3.23.2	AMNH DOT 14025	BTP2013	*C. livia*	0.25%	0.000182	13.903729
BMNH794	ROM 40360	AMNH DOT 14025	BTP2013	*C. livia*	0.61%	0.000277	21.873552
BTP2013	AMNH DOT 14025	BMNH794	ROM 34.3.23.2	*C. livia*	0.02%	0.000087	1.949575

Table S8. Counts of nonsynonymous and synonymous polymorphisms and substitutions in passenger and band-tailed pigeons for genes involved in spermatogenesis.

		Fixed			Polymorphic		
	Counts:	**N**	**S**	**Ratio**	**N**	**S**	**Ratio**
Passenger pigeons	*High-diversity*	57	90	0.63	103	383	0.27
	Low-diversity	43	113	0.38	22	53	0.42
Band-tailed pigeons	*High-diversity*	49	87	0.56	36	60	0.60
	Low-diversity	51	92	0.55	19	30	0.63

Table S9. Counts of nonsynonymous and synonymous polymorphisms and substitutions in passenger and band-tailed pigeons for genes in immunity pathways.

		Fixed			Polymorphic		
	Counts:	**N**	**S**	**Ratio**	**N**	**S**	**Ratio**
Passenger pigeons	*High-diversity*	93	155	0.60	299	814	0.37
	Low-diversity	65	125	0.52	42	80	0.53
Band-tailed pigeons	*High-diversity*	103	152	0.68	74	122	0.61
	Low-diversity	51	135	0.38	35	73	0.48

Table S10. Counts of synonymous and nonsynonymous derived mutations at different frequencies in passenger and band-tailed pigeons. Each table describes counts for different types of nucleotide base change. Genes are divided according to whether they are found in a high-diversity or a low-diversity region of the passenger pigeon genome (i.e. a 5 Mb region with higher or lower diversity than the homologous region in the band-tailed pigeon genome). Polymorphisms are divided into those that are at a low frequency in the population (defined as singletons) and those that are at a high frequency (not singletons). To facilitate comparison with band-tailed pigeons, the counts presented are based on a subsample of two passenger pigeons.

	G/C to G/C or A/T to A/T	**Fixed differences**			**Polymorphism: High frequency**			**Polymorphism: Low frequency**		
	Counts:	**N**	**S**	**Ratio**	**N**	**S**	**Ratio**	**N**	**S**	**Ratio**
PP	*High-diversity*	3216	2519	1.28	876	882	0.99	2791	2470	1.13
	Low-diversity	2232	1362	1.64	112	97	1.15	665	307	2.17
BTP	*High-diversity*	2583	1839	1.40	187	140	1.34	441	286	1.54
	Low-diversity	2106	1295	1.63	140	89	1.57	281	181	1.55

	G/C to A/T	**Fixed differences**			**Polymorphism: High frequency**			**Polymorphism: Low frequency**		
	Counts:	**N**	**S**	**Ratio**	**N**	**S**	**Ratio**	**N**	**S**	**Ratio**
PP	*High-diversity*	5404	7696	0.70	1824	5267	0.35	9031	21706	0.42
	Low-diversity	6212	12072	0.51	393	860	0.46	1955	2881	0.68
BTP	*High-diversity*	6151	13085	0.47	550	1397	0.39	1509	3731	0.40
	Low-diversity	5098	9344	0.55	315	662	0.48	901	1622	0.56

	A/T to G/C	**Fixed differences**			**Polymorphism: High frequency**			**Polymorphism: Low frequency**		
	Counts:	**N**	**S**	**Ratio**	**N**	**S**	**Ratio**	**N**	**S**	**Ratio**
PP	*High-diversity*	6117	13520	0.45	1999	6399	0.31	5952	14841	0.40
	Low-diversity	3880	6382	0.61	231	471	0.49	1011	1555	0.65
BTP	*High-diversity*	5998	10357	0.58	325	664	0.49	564	928	0.61
	Low-diversity	4668	8508	0.55	177	319	0.55	418	618	0.68

4. References

1. A. W. Schorger, *The Passenger Pigeon: Its Natural History And Extinction* (Literary Licensing, LLC, 1955).
2. E. H. Bucher, "The causes of extinction of the Passenger Pigeon", in *Current Ornithology*, D. M. Power, Ed. (Springer US, 1992), pp. 1–36.
3. E. M. Leffler, K. Bulluaghey, D. R. Matute, W. K. Meyer, L. Ségurel, A. Venkat, P. Andolfatto, M. Przeworski, Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. **10**, e1001388 (2012).
4. C.-M. Hung, P.-J. L. Shaner, R. M. Zink, W.-C. Liu, T.-C. Chu, W.-S. Huang, S.-H. Li, Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10636–10641 (2014).
5. J. M. Smith, J. Haigh, The hitch-hiking effect of a favourable gene. *Genet. Res.* 23, 23–35 (1974).
6. D. J. Begun, C. F. Aquadro, Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*. **356**, 519–520 (1992).
7. A. D. Cutter, B. A. Payseur, Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat. Rev. Genet.* **14**, 262–274 (2013).
8. B. Charlesworth, The effects of deleterious mutations on evolution at linked sites. *Genetics*. **190**, 5–22 (2012).
9. A. Eyre-Walker, The genomic rate of adaptive evolution. *Trends Ecol. Evol.* **21**, 569–575 (2006).
10. R. C. Lewontin, *The Genetic Basis of Evolutionary* Change (Columbia University Press, 1974).
11. R. B. Corbett-Detig, D. L. Hartl, T. B. Sackton, Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol.* **13**, e1002112 (2015).
12. K. P. Johnson, D. H. Clayton, J. P. Dumbacher, R. C. Fleischer, The flight of the Passenger Pigeon: phylogenetics and biogeographic history of an extinct species. *Mol. Phylogenet. Evol.* **57**, 455–458 (2010).
13. T. A. Sanders, Band-tailed pigeon population status, 2014. *U.S. Department of the Interior, Fish and Wildlife Service, Division of Migratory Bird Management, Washington, D.C.* (2014).
14. B. Nabholz, R. Lanfear, J. Fuchs, Body mass-corrected molecular rate for bird mitochondrial DNA. *Mol. Ecol.* **25**, 4438–4449 (2016).

15. International Chicken Genome Sequencing Consortium, Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. **432**, 695–716 (2004).
16. H. Ellegren, Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol. Evol*. **25**, 283–291 (2010).
17. R. Hershberg, D. A. Petrov, Selection on codon bias. *Annu. Rev. Genet.* **42**, 287–299 (2008).
18. S. Tong, J. Li, J. R. Wands, Carboxypeptidase D Is an avian hepatitis B virus receptor. *J. Virol*. **73**, 8696–8702 (1999).
19. N. Ramírez-Otárola, P. Sabat, Are levels of digestive enzyme activity related to the natural diet in passerine birds? *Biol. Res*. **44**, 81–88 (2011).
20. F. A. Moreira, N. Kaiser, K. Monory, B. Lutz, Reduced anxiety-like behaviour induced by genetic and pharmacological inhibition of the endocannabinoid-degrading enzyme fatty acid amide hydrolase (FAAH) is mediated by CB1 receptors. *Neuropharmacology*. **54**, 141–150 (2008).
21. B. T. Grenfell, A. P. Dobson, *Ecology of Infectious Diseases in Natural Populations* (Cambridge University Press, 1995).
22. R. K. Murton, A. J. Isaacson, N. J. Westwood, The significance of gregarious feeding behaviour and adrenal stress in a population of Wood-pigeons *Columba palumbus*. *J. Zool*. **165**, 1, 53-84 (1971).
23. N. Galtier, Adaptive protein evolution in animals and the effective population size hypothesis. *PLoS Genet.* **12**, e1005774 (2016).
24. W. G. Hill, A. Robertson, The effect of linkage on limits to artificial selection. *Genet. Res*. **8**, 269–294 (1966).
25. B. M. Van Doren, L. Campagna, B. Helm, J. C. Illera, I. J. Lovette, M. Liedvogel, Correlated patterns of genetic diversity and differentiation across an avian family. *Mol. Ecol*. mec.14083 (2017).
26. L. Dutoit, N. Vijay, C. F. Mugal, C. M. Bossu, R. Burri, J. Wolf, H. Ellegren, Covariation in levels of nucleotide diversity in homologous regions of the avian genome long after completion of lineage sorting. *Proc. Biol. Sci.* **284**, 20162756 (2017).
27. A. Künstner, J. B. W. Wolf, N. Backström, O. Whitney, C. N. Balakrishnan, L. Day, S. V. Edwards, D. E. Janes, B. A. Schlinger, R. K. Wilson, E. D. Jarvis, W. C. Warren, H. Ellegren, Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Mol. Ecol.* **19** (suppl. 1), 266–276 (2010).

28. K. Nam, C. Mugal, B. Nabholz, H. Schielzeth, J. B. Wolf, N. Backström, A. Künstner, C. N. Balakrishnan, A. Heger, C. P. Ponting, D. F. Clayton, H. Ellegren, Molecular evolution of genes in avian genomes. *Genome Biol.* **11**, R68 (2010).
29. T. I. Gossmann, A. W. Santure, B. C. Sheldon, J. Slate, K. Zeng, Highly variable recombinational landscape modulates efficacy of natural selection in birds. *Genome Biol. Evol*. **6**, 2061–2075 (2014).
30. P. Bolívar, C. F. Mugal, A. Nater, H. Ellegren, Recombination rate variation modulates gene sequence evolution mainly via GC-biased gene conversion, not Hill–Robertson interference, in an avian system. *Mol. Biol. Evol*. **33**, 216–227 (2016).
31. L. Duret, A. Eyre-Walker, N. Galtier, A new perspective on isochore evolution. *Gene*. **385**, 71–74 (2006).
32. T. Nagylaki, Evolution of a finite population under gene conversion. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6278–6281 (1983).
33. H. Li, R. Durbin, Inference of human population history from whole genome sequence of a single individual. *Nature*. **475**, 493–496 (2011).
34. D. R. Schrider, A. G. Shanku, A. D. Kern, Effects of linked selective sweeps on demographic inference and model selection. *Genetics*. **204**, 3, 1207-1223 (2016).
35. C.-M. Hung, R.-C. Lin, J.-H. Chu, C.-F. Yeh, C.-J. Yao, S.-H. Li., The *de novo* assembly of mitochondrial genomes of the extinct Passenger Pigeon (*Ectopistes migratorius*) with next generation sequencing. *PLoS One*. **8**, e56301 (2013).
36. H. N. Poinar, A. Cooper, Ancient DNA: do it right or not at all. *Science*. **289**, 5482, 1139 (2000).
37. N. Rohland, H. Siedel, M. Hofreiter, A rapid column-based ancient DNA extraction method for increased sample throughput. *Mol. Ecol. Resour*. **10**, 677–683 (2010).
38. J. Dabney, M. Knapp, I. Glocke, M. T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. García, S. Pääbo, J.-L. Arsuaga, M. Meyer, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013).
39. T. L. Fulton, S. M. Wagner, B. Shapiro, Case study: recovery of ancient nuclear DNA from toe pads of the extinct passenger pigeon. *Methods Mol. Biol.* **840**, 29–35 (2012).
40. B. Shapiro, D. Sibthorpe, A. Rambaut, J. Austin, G. M. Wragg, O. R. P. Bininda-Emonds, Flight of the dodo. *Science*. **295**, 5560,1683 (2002).
41. T. L. Fulton, M. Stiller, "PCR Amplification, Cloning, and Sequencing of Ancient DNA", in *Ancient DNA: Methods in Molecular Biology*, B. Shapiro, M. Hofreiter, Eds. (Humana Press), chap. 15.
42. N. Rohland, D. Reich, Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939–946 (2012).

43. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 2010, db.prot5448 (2010).
44. A. E. R. Soares, B. J. Novak, J. Haile, T. H. Heupink, J. Fjeldså, M. T. P. Gilbert, H. Poinar, G. M. Church, B. Shapiro, Complete mitochondrial genomes of living and extinct pigeons revise the timing of the columbiform radiation. *BMC Evol. Biol.* **16**, 230 (2016).
45. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
46. M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).
47. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 8, 1969–1973 (2012).
48. A. J. Drummond, A. Rambaut, B. Shapiro, O. G. Pybus, Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
49. S. Y. W. Ho, B. Shapiro, Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol. Ecol. Resour.* **11**, 423–434 (2011).
50. A. G. Rambaut, A. J. Drummond, Tracer v1.6, Available from *http://beast.bio.ed.ac.uk/Tracer* (2014).
51. A. Rambaut, A. J. Drummond, TreeAnnotator v1.7.0. Available from *http://beast.bio.ed.ac.uk/treeannotator* (2013).
52. J. A. Chapman, I. Ho, S. Sunkara, S. Luo, G. P. Schroth, D. S. Rokhsar, Meraculous: *de novo* genome assembly with short paired-end reads. *PLoS One.* **6**, e23501 (2011).
53. N. H. Putnam, B. L. O'Connell, J. C. Stites, B. J. Rice, M. Blanchette, R. Calef, C. J. Troll, A. Fields, P. D. Hartley, C. W. Sugnet, D. Haussler, D. S. Rokhsar, R. E. Green, Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **26**, 342–350 (2016).
54. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
55. M. Stanke, O. Keller, I. Gunduz, S. Waack, B. Morgenstern, AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*. **34**, W435–9 (2006).
56. M. D. Shapiro, Z. Kronenberg, C. Li, E. T. Domyan, H. Pan, M. Campbell, H. Tan, C. D. Huff, H. Hu, A. I. Vickrey, S. C. A. Nielsen, S. A Stringham, H. Hu, E. Willerslev, T. P. Gilbert, M. Yandell, G. Zhang, J. Wang, Genomic diversity and evolution of the head crest in the Rock Pigeon. *Science*. **339**, 1063–1067 (2013).

57. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **25**, 3389–3402 (1997).
58. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* 1303.3997 (2013).
59. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).
60. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26**, 841–842 (2010).
61. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
62. P. Danecek, A. Auton, G. Abencasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsake, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics*. **27**, 2156–2158 (2011).
63. K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. **505**, 43–49 (2014).
64. A. L. Delcher, A. Phillippy, J. Carlton, S. L. Salzberg, Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res*. **30**, 2478–2483 (2002).
65. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. **20**, 289–290 (2004).
66. D. J. Obbard, J. J. Welch, K.-W. Kim, F. M. Jiggins, Quantifying Adaptive Evolution in the *Drosophila* Immune System. *PLoS Genet*. **5**, e1000698 (2009).
67. T. I. Gossmann, B.-H. Song, A. J. Windsor, T. Mitchell-Olds, C. J. Dixon, M. V. Kapralov, D. A. Filatov, A. Eyre-Walker, Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol. Biol. Evol.* **27**, 1822–1832 (2010).

68. N. Stoletzki, A. Eyre-Walker, Estimation of the neutrality index. *Mol. Biol. Evol.* **28**, 63–70 (2011).
69. J. A. Shapiro, W. Huang, C. Zhang, M. J. Hubisz, J. Lu, D. A. Turissini, S. Fang, H.-Y. Wang, R. R. Hudson, R. Nielsen , Adaptive genic evolution in the *Drosophila* genomes. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2271–2276 (2007).
70. J. H. McDonald, M. Kreitman, Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*. **351**, 652–654 (1991).
71. L. Excoffier, H. E. L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour*. **10**, 564–567 (2010).
72. A. G. F. Teacher, D. J. Griffiths, HapStar: automated haplotype network layout and visualization. *Mol. Ecol. Resour*. **11**, 151–153 (2011).
73. M. Hofreiter, V. Jaenicke, D. Serre, A. von Haeseler, S. Pääbo, DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **29**, 4793–4799 (2001).
74. T. Lindahl, Instability and decay of the primary structure of DNA. *Nature*. **362**, 709–715 (1993).
75. P. D. Heintzman, A. E. R. Soares, D. Chang, B. Shapiro, Paleogenomics. *Reviews in Cell Biology and Molecular Medicine*. **1**, 243–267 (2015).
76. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. **29**, 1682–1684 (2013).
77. N. Backström, W. Forstmeier, H. Schielzeth, H. Mellenius, K. Nam, E. Bolund, M. T. Webster, T. Öst, M. Schneider, B. Kempenaers, H. Ellegren, The recombination landscape of the zebra finch Taeniopygia guttata genome. *Genome Res*. **20**, 485–495 (2010).
78. T. Kawakami, L. Smeds, N. Backström, A. Husby, A. Qvarnström, C. F. Mugal, P. Olason, H. Ellegren, A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol. Ecol.* **23**, 4035–4058 (2014).
79. S. Singhal, E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, A. I. Strand, Q. Li, B. Raney, C. N. Balakrishnan, S. C. Griffith, G. McVean, M. Przeworski, Stable recombination hotspots in birds. *Science*. **350**, 928–932 (2015).
80. A. Auton, G. McVean, Recombination rate estimation in the presence of hotspots. *Genome Res*. **17**, 1219–1227 (2007).

81. H. Brunschwig, L. Levi, E. Ben-David, R. W. Williams, B. Yakir, S. Shifman, Fine-scale maps of recombination rates and hotspots in the mouse genome. *Genetics*. **191**, 757–764 (2012).
82. J. C. Stanton, Present-day risk assessment would have predicted the extinction of the Passenger Pigeon (*Ectopistes migratorius*). *Biol. Conserv*. **180**, 11–20 (2014).
83. H. Ellegren, L. Smeds, R. Burri, P. I. Olason, N. Backström, T. Kawakami, A. Künstner, H. Mäkinen, K. Nadachowska-Brzyska, A. Qvarnström, S. Uebbing, J. B. W. Wolf, The genomic landscape of species divergence in Ficedula flycatchers. *Nature*. **491**, 756–760 (2012).
84. G. Zhang, C. Li, Q. Li, B. Li, D. M. Larkin, C. Lee, J. F. Storz, A. Antunes, M. J. Greenwold, R. W. Meredith, A. Ödeen, J. Cui, Q. Zhou, L. Xu, H. Pan, Z. Wang, L. Jin, P. Zhang, H. Hu, W. Yang, J. Hu, J. Xiao, Z. Yang, Y. Liu, Q. Xie, H. Yu, J. Lian, P. Wen, F. Zhang, H. Li, Y. Zeng, Z. Xiong, S. Liu, L. Zhou, Z. Huang, N. An, J. Wang, Q. Zheng, Y. Xiong, G. Wang, B. Wang, J. Wang, Y. Fan, R. R. da Fonseca, A. Alfaro-Núñez, M. Schubert, L. Orlando, T. Mourier, J. T. Howard, G. Ganapathy, A. Pfenning, O. Whitney, M. V. Rivas, E. Hara, J. Smith, M. Farré, J. Narayan, G. Slavov, M. N. Romanov, R. Borges, J. P. Machado, I. Khan, M. S. Springer, J. Gatesy, F. G. Hoffmann, J. C. Opazo, O. Håstad, R. H. Sawyer, H. Kim, K.-W. Kim, H. J. Kim, S. Cho, N. Li, Y. Huang, M. W. Bruford, X. Zhan, A. Dixon, M. F. Bertelsen, E. Derryberry, W. Warren, R. K. Wilson, S. Li, D. A. Ray, R. E. Green, S. J. O'Brien, D. Griffin, W. E. Johnson, D. Haussler, O. A. Ryder, E. Willerslev, G. R. Graves, P. Alström, J. Fjeldså, D. P. Mindell, S. V. Edwards, E. L. Braun, C. Rahbek, D. W. Burt, P. Houde, Y. Zhang, et al., Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*. **346**, 1311–1320 (2014).
85. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
86. J. Charlesworth, A. Eyre-Walker, The McDonald–Kreitman Test and Slightly Deleterious Mutations. *Mol. Biol. Evol*. **25**, 1007–1015 (2008).
87. J. A. Novembre, Accounting for background nucleotide composition when measuring codon usage bias. *Mol. Biol. Evol.* **19**, 1390–1394 (2002).
88. F. Wright, The "effective number of codons" used in a gene. *Gene*. **87**, 23–29 (1990).
89. H. Akashi, Inferring weak selection from patterns of polymorphism and divergence at" silent" sites in *Drosophila* DNA. *Genetics*. **139**, 1067–1076 (1995).
90. S. Karlin, J. Mrázek, What drives codon choices in human genes? *J. Mol. Biol.* **262**, 459–472 (1996).
91. J. C. Fay, C. I. Wu, Hitchhiking under positive Darwinian selection. *Genetics*. **155**, 1405–1413 (2000).

92. F. Tajima, The Effect of Change in Population Size on DNA Polymorphism. *Genetics*. **123**, 597–601 (1989).
93. K. Zeng, Y.-X. Fu, S. Shi, C.-I. Wu, Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics*. **174**, 1431–1439 (2006).
94. K. Zeng, S. Shi, C.-I. Wu, Compound tests for the detection of hitchhiking under positive selection. *Mol. Biol. Evol*. **24**, 1898–1908 (2007).
95. R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H.-Y. Fritz, N. F. Hansen, E. Y. Durand, A.-S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prüfer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Höber, B. Höffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Ž. Kucan, I. Gušic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. Pääbo, A Draft Sequence of the Neandertal Genome. *Science*. **328**, 710–722 (2010).
96. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
97. P. W. Messer, D. A. Petrov, Frequent adaptation and the McDonald–Kreitman test. *Proceedings of the National Academy of Sciences*. **110**, 8615–8620 (2013).
98. P. U. Clark, A. S. Dyke, J. D. Shakun, A. E. Carlson, J. Clark, B. Wohlfarth, J. X. Mitrovica, S. W. Hostetler, A. M. McCabe, The Last Glacial Maximum. *Science*. **325**, 710–714 (2009).
99. M. Karikoski, H. Irjala, M. Maksimow, M. Miiluniemi, K. Granfors, S. Hernesniemi, K. Elima, G. Moldenhauer, K. Schledzewski, J. Kzhyshkowska, S. Goerdt, M. Salmi, S. Jalkanen, Clever-1/Stabilin-1 regulates lymphocyte migration within lymphatics and leukocyte entrance to sites of inflammation. *Eur. J. Immunol*. **39**, 3477–3487 (2009).
100. M. Scherma, J. Medalie, W. Fratta, S. K. Vadivel, A. Makriyannis, D. Piomelli, E. Mikics, J. Haller, S. Yasar, G. Tanda, S. R. Goldberg, The endogenous cannabinoid anandamide has effects on motivation and anxiety that are revealed by fatty acid amide hydrolase (FAAH) inhibition. *Neuropharmacology*. **54**, 129–140 (2008).
101. S. Strindberg, T. W. Nielsen, Â. M. Ribeiro, B. Wiinberg, A. T. Kristensen, M. F. Bertelsen, Thromboelastography in Selected Avian Species*. J. Avian Med. Surg.* **29**, 282–289 (2015).
102. Y. Meng, W. Zhang, J. Zhou, M. Liu, J. Chen, S. Tian, M. Zhuo, Y. Zhang, Y. Zhong, H. Du, X. Wang, Genome-wide analysis of positively selected genes in seasonal and non-seasonal breeding species. *PLoS One*. **10**, e0126736 (2015).

103. M. Ben Khelifa, C. Coutton, R. Zouari, T. Karaouzène, J. Rendu, M. Bidart, S. Yassine, V. Pierre, J. Delaroche, S. Hennebicq, D. Grunwald, D. Escalier, K. Pernet-Gallay, P.-S. Jouk, N. Thierry-Mieg, A. Touré, C. Arnoult, P. F. Ray, Mutations in DNAH1, which encodes an inner arm heavy chain dynein, lead to male infertility from multiple morphological abnormalities of the sperm flagella. *Am. J. Hum. Genet.* **94**, 95–104 (2014).
104. S. Labeit, C. A. C. Ottenheijm, H. Granzier, Nebulin, a major player in muscle health and disease. *FASEB J.* **25**, 822–829 (2011).
105. Z. Lu, Z. Yuan, T. Miyoshi, Q. Wang, Z. Su, C. C. Chang, W. Shi, Identification of Soat1 as a quantitative trait locus gene on mouse chromosome 1 contributing to hyperlipidemia. *PLoS One*. **6**, e25344 (2011).
106. M. P. Richards, S. M. Poch, C. N. Coon, R. W. Rosebrough, C. M. Ashwell, J. P. McMurtry, Feed restriction significantly alters lipogenic gene expression in broiler breeder chickens. *J. Nutr.* **133**, 707–715 (2003).
107. M. Benn, Apolipoprotein B levels, APOB alleles, and risk of ischemic cardiovascular disease in the general population, a review. *Atherosclerosis*. **206**, 17–30 (2009).
108. S. Liu, E. D. Lorenzen, M. Fumagalli, B. Li, K. Harris, Z. Xiong, L. Zhou, T. S. Korneliussen, M. Somel, C. Babbitt, G. Wray, J. Li, W. He, Z. Wang, W. Fu, X. Xiang, C. C. Morgan, A. Doherty, M. J. O'Connell, J. O. McInerney, E. W. Born, L. Dalén, R. Dietz, L. Orlando, C. Sonne, G. Zhang, R. Nielsen, E. Willerslev, J. Wang, Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell*. **157**, 785–794 (2014).
109. M. Proszkowiec-Weglarz, M. P. Richards, R. Ramachandran, J. P. McMurtry, Characterization of the AMP-activated protein kinase pathway in chickens. *Comp. Biochem. Physiol. B Biochem. Mol. Biol*. **143**, 92–106 (2006).
110. S. O. Zhang, S. Mathur, G. Hattem, O. Tassy, O. Pourquié, Sex-dimorphic gene expression and ineffective dosage compensation of Z-linked genes in gastrulating chicken embryos. *BMC Genomics*. **11**, 13 (2010).
111. K. Silander, K. L. Mohlke, L. J. Scott, E. C. Peck, P. Hollstein, A. D. Skol, A. U. Jackson, P. Deloukas, S. Hunt, G. Stavrides, P. S. Chines, M. R. Erdos, N. Narisu, K. N. Conneely, C. Li, T. E. Fingerlin, S. K. Dhanjal, T. T. Valle, R. N. Bergman, J. Tuomilehto, R. M. Watanabe, M. Boehnke, F. S. Collins, Genetic variation near the hepatocyte nuclear factor-4 alpha gene predicts susceptibility to type 2 diabetes. *Diabetes*. **53**, 1141–1149 (2004).
112. S. Vasu, S. Shah, A. Orjalo, M. Park, W. H. Fischer, D. J. Forbes, Novel vertebrate nucleoporins Nup133 and Nup160 play a role in mRNA export. *J. Cell Biol*. **155**, 339–354 (2001).

113. S. Tang, D. C. Presgraves, Evolution of the Drosophila nuclear pore complex results in multiple hybrid incompatibilities. *Science*. **323**, 779–782 (2009).
114. S. Cappello, M. J. Gray, C. Badouel, S. Lange, M. Einsiedler, M. Srour, D. Chitayat, F. F. Hamdan, Z. A. Jenkins, T. Morgan, N. Preitner, T. Uster, J. Thomas, P. Shannon, V. Morrison, N. Di Donato, L. Van Maldergem, T. Neuhann, R. Newbury-Ecob, M. Swinkells, P. Terhal, L. C. Wilson, P. J. G. Zwijnenburg, A. J. Sutherland-Smith, M. A. Black, D. Markie, J. L. Michaud, M. A. Simpson, S. Mansour, H. McNeill, M. Götz, S. P. Robertson, Mutations in genes encoding the cadherin receptor-ligand pair DCHS1 and FAT4 disrupt cerebral cortical development. *Nat. Genet.* **45**, 1300–1308 (2013).
115. E. Sadeqzadeh, C. E. de Bock, R. F. Thorne, Sleeping giants: emerging roles for the fat cadherins in health and disease. *Med. Res. Rev.* **34**, 190–221 (2014).
116. A. R. R. Eagle, R. L. Hanson, W. Jiang, X. Han, G. L. Matters, G. Imperatore, W. C. Knowler, J. S. Bond, Meprin β metalloprotease gene polymorphisms associated with diabetic nephropathy in the Pima Indians. *Hum. Genet.* **118**, 12 (2005).
117. U. D. P. Lam, E. Lerchbaum, N. Schweighofer, O. Trummer, K. Eberhard, B. Genser, T. R. Pieber, B. Obermayer-Pietsch, Association of MEP1A gene variants with insulin metabolism in central European women with polycystic ovary syndrome. *Gene*. **537**, 245–252 (2014).
118. C. Albrecht, E. Viturro, The ABCA subfamily—gene and protein structures, functions and associated hereditary diseases. *Pflugers Arch*. **453**, 581–589 (2007).
119. K. Tatebe, A. Zeytun, R. M. Ribeiro, R. Hoffmann, K. S. Harrod, C. V. Forst, Response network analysis of differential gene expression in human epithelial lung cells during avian influenza infections. BMC *Bioinformatics*. **11**, 170 (2010).
120. C. Pattaro, A. De Grandi, V. Vitart, C. Hayward, A. Franke, Y. S. Aulchenko, A. Johansson, S. H. Wild, S. A. Melville, A. Isaacs, O. Polasek, D. Ellinghaus, I. Kolcic, U. Nöthlings, L. Zgaga, T. Zemunik, C. Gnewuch, S. Schreiber, S. Campbell, N. Hastie, M. Boban, T. Meitinger, B. A. Oostra, P. Riegler, C. Minelli, A. F. Wright, H. Campbell, C. M. van Duijn, U. Gyllensten, J. F. Wilson, M. Krawczak, I. Rudan, P. P. Pramstaller, A meta-analysis of genome-wide data from five European isolates reveals an association of COL22A1, SYT1, and GABRR2 with serum creatinine level. *BMC Med. Genet.* **11**, 41 (2010).
121. S. S. Reemers, D. A. van Haarlem, M. J. Groot Koerkamp, L. Vervelde, Differential gene-expression and host-response profiles against avian influenza virus within the chicken lung due to anatomy and airflow. *J. Gen. Virol.* **90**, 2134–2146 (2009).
122. M. R. Marcello, J. P. Evans, Multivariate analysis of male reproductive function in Inpp5b-/- mice reveals heterogeneity in defects in fertility, sperm--egg membrane

interaction and proteolytic cleavage of sperm ADAMs. *Mol. Hum. Reprod.* **16**, 492–505 (2010).

123. A. Gardin, J. White, The Sanger Mouse Genetics Programme: High Throughput Characterisation of Knockout Mice*. Acta Ophthalmol.* **89**, s248 (2011).