Title:

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

Authors:

Gemma G. R. Murray[1*], André E. R. Soares[1*], Ben J. Novak[1,2], Nathan K. Schaefer[3], James A. Cahill[1], Allan J. Baker[4†], John R. Demboski[5], Andrew Doll[5], Rute R. Da Fonseca[6], Tara L. Fulton[1,7], M. Thomas P. Gilbert[6,8], Peter D. Heintzman[1,9], Brandon Letts[10], George McIntosh[11], Brendan L. O'Connell[3], Mark Peck[5], Marie-Lorraine Pipes[12], Edward S. Rice[3], Kathryn M. Santos[11], A. Gregory Sohrweide[13], Samuel H. Vohr[3], Russell B. Corbett-Detig[3,14], Richard E. Green[3,14], Beth Shapiro[1,14‡].

Affiliations:

1. Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064, USA.
2. Revive & Restore, Sausalito, CA 94965, USA.
3. Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, USA.
4. Department of Natural History, Royal Ontario Museum, Toronto, ON M5S 2C6, Canada.
5. Department of Zoology, Denver Museum of Nature & Science, Denver, CO 80205, USA.
6. Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark.
7. Environment and Climate Change Canada, 9250-49th Street, Edmonton, Alberta T6B 1K5, Canada.
8. NTNU University Museum, 7491 Trondheim, Norway.

9. Tromsø University Museum, UiT - The Arctic University of Norway, 9037 Tromsø, Norway.
10. Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA.
11. Collections Department, Rochester Museum & Science Center, Rochester, NY 14607, USA.
12. Marie-Lorraine Pipes, Zooarchaeologist Consultant, Victor, NY 14564, USA.
13. A. Gregory Sohrweide D.D.S., Baldwinsville, NY 13027, USA.
14. UCSC Genomics Institute, 1156 High Street, Santa Cruz, CA 95064.

* These authors contributed equally to this work

† deceased author

‡ Corresponding author. Email: beth.shapiro@gmail.com.

One Sentence Summary:

The passenger pigeon's abundance and recombination landscape led to natural selection dominating genome-wide neutral site evolution.

Abstract:

The extinct passenger pigeon was once the most abundant bird in North America, and possibly the world. While theory predicts that large populations will be more genetically diverse, passenger pigeon genetic diversity was surprisingly low. To investigate this, we analysed 41 mitochondrial and 4 nuclear genomes from passenger pigeons and 2 genomes from band-tailed pigeons, which are passenger pigeons' closest living relatives. Passenger pigeons' large population size appears to have allowed for faster adaptive evolution and removal of harmful mutations, driving a huge loss in their neutral genetic diversity. These results demonstrate the impact selection can have on a vertebrate genome, and contradict results that suggested population instability contributed to this species' surprisingly rapid extinction.

Main text:

The passenger pigeon (*Ectopistes migratorius*) numbered between 3 and 5 billion individuals prior to its 19th century decline and eventual extinction (*1*). Passenger pigeons were highly mobile, bred in large social colonies, and their population lacked clear geographic structure (*2*). Few vertebrates have populations this large and cohesive and, according to the neutral model of molecular evolution, this should lead to a large effective population size (N_e) and high genetic diversity (*3*). Preliminary analyses of passenger pigeon genomes have, however, revealed surprisingly low genetic diversity (*4*). This has been interpreted within the framework of the neutral theory of molecular evolution as the result of a history of dramatic demographic fluctuations (*4*). However, in large populations, natural selection may be particularly important in shaping genetic diversity: population genetic theory predicts that selection will be more effective in large populations (*3*), and selection on one locus can cause a loss of diversity at other loci, particularly those that are closely linked (*5*–*8*). It has

been suggested that this could explain why the genetic diversity of a species is poorly predicted by its population size (*9*–*11*).

We investigated the impact of natural selection on passenger pigeon genomes through comparative genomic analyses of both passenger pigeons and band-tailed pigeons (*Patagioenas fasciata*). While ecologically and physiologically similar to passenger pigeons, band-tailed pigeons have a present-day population size three orders of magnitude smaller than their close relative the passenger pigeon (*2*, *12*, *13*).

We applied a Bayesian skyline model of ancestral population dynamics to the mitochondrial genomes of 41 passenger pigeons from across their former breeding range (Fig. 1A and table S1) (*14*). This returned a most recent effective population size (N_e) of 13 million (95% HPD: 2-58 million) and similar, stable N_e for the previous 20,000 years (Fig. 1B). While this N_e is much lower than the (census) population size (N_c), it is greater than previous estimates from analyses of nuclear genomes (*4*), and is likely to be conservative (*14*).

We compared nucleotide diversity (π) in the passenger pigeon nuclear genome to π in the band-tailed pigeon nuclear genome. We analysed four high-coverage passenger pigeon genome assemblies (two newly sequenced and two from published raw data; table S2), and two high-coverage band-tailed pigeon genome assemblies. π was greater in passenger pigeons (average π = 0.008) than in band-tailed pigeons (average π = 0.004), but this difference is less than expected given their population sizes (it suggests that N_e/N_c was 0.0002 for passenger pigeons compared to 0.2 for band-tailed pigeons; *14*). We estimated π for non-overlapping 5 Mb windows across the genome, and found that these species exhibit a correlated regional variation in π, but with greater variation in passenger pigeons (Fig. 2A and figs. S1-4).

To explore this variation, we mapped our scaffolds to the chicken genome assembly *(14)*, which approximates chromosomal structure since karyotype and synteny are strongly conserved across birds *(15)*. We found that low genetic diversity regions of the passenger pigeon genome are generally in the centres of macrochromosomes, while the edges of macrochromosomes and microchromosomes have higher diversity (Fig. 2B). Although this pattern is largely absent from the band-tailed pigeon genome, it is unlikely to be an artefact of ancient DNA damage as our assemblies had high coverage depth (table S2), we used conservative cut-offs for calling variants, and we recovered similar patterns after excluding variants more likely to have resulted from damage (fig. S5; *14*). Instead, the pattern mirrors the recombination landscape of the bird genome, where recombination rates are lower in the centers of macrochromosomes, relative both to their edges and the microchromosomes *(14, 15)*.

We next investigated the impact of selection on the evolution of protein-coding regions of the genome in both species. We calculated the rate of adaptive substitution relative to the rate of neutral substitution (ω_a) and the ratio of nonsynonymous to synonymous polymorphism (*pN*/*pS*) for 5 Mb windows across the genome. A higher ω_a suggests stronger or more efficient positive selection, and a lower *pN*/*pS* suggests stronger or more efficient selective constraint. ω_a was higher (Mann-Whitney U test, $p = 1.3\text{x}10^{-5}$) and *pN*/*pS* lower ($p = 8.2\text{x}10^{-12}$) in passenger pigeons than band-tailed pigeons (Fig. 3 and fig. S6). We also found that ω_a was higher ($p = 2.2\text{x}10^{-8}$) and *pN*/*pS* lower ($p = 4.1\text{x}10^{-6}$) in high-diversity regions of the passenger pigeon genome compared to low-diversity regions (Fig. 3 and fig. S6). In addition, codon usage bias, which is thought to reflect selection for translational optimization (*16*), was greater in passenger pigeons than in band-tailed pigeons, and greater in high-diversity regions (figs. S19, S20).

We also estimated the difference between the proportions of substitutions and polymorphisms that are nonsynonymous (the direction of selection, DoS) for individual genes, where a positive DoS indicates adaptive evolution. DoS was more often positive in passenger pigeons than in band-tailed pigeons and, in passenger pigeons, DoS was correlated with diversity (fig. S7). McDonald-Kreitman tests (*17*) identified 32 genes with evidence of adaptive evolution in passenger pigeons (table S3). Among them are genes associated with immune defense (e.g. *CPD*), seasonal consumption of high-sugar foods in passerine birds (*SI*), and stress modulation (*FAAH*). Selection on these gene functions is consistent with the distinctive diet of passenger pigeons, their sociability, and their population size and density (*2*), which could have led to an increased burden of transmissible pathogens (*18*) and increased stress (*19*).

Differences in the impact of selection between passenger pigeons and band-tailed pigeons could derive from differences in recombination rate, mutation rate, and the distribution of fitness effects. However, the close relationship between these species makes substantial differences in most of these factors unlikely and the most parsimonious explanation is their different population sizes. Theory predicts that larger populations will experience a greater impact of natural selection, both because they generate more mutations per generation, and because selection is more effective in overcoming random drift when N_e is large (*3*).

A greater impact of selection on nonsynonymous sites could also increase the impact of selection on neutral sites due to linkage. In linked regions, selection on one site can lead to reduced diversity at neutral sites and a reduced efficiency of selection at other selected sites (*3*, *20*). The impact of this will be greater where recombination rates are low since linked regions will be larger. Therefore, the large population size of the passenger pigeon,

148 assuming a typical avian recombination landscape, may have resulted in an overall
149 increased neutral genetic diversity and efficiency of selection, but reduced genetic diversity
150 and efficiency of selection in genomic regions with lower recombination rates, due to linkage
151 with selected variants. This explains the pattern of diversity across the passenger pigeon
152 genome (Fig. 2), including the low diversity in the mitochondrial genome (Fig. 1B; *14*). It is
153 also supported by other avian studies, which report a correlation between recombination rate
154 and both diversity (*21*, *22*) and the efficiency of selection (*23*, *24*). However, it has been
155 argued that the observed correlation between recombination and the efficiency of selection
156 could be an artefact of GC-biased gene conversion (gBGC) (*25*).
157
158 Regions of the genome with higher recombination rates are expected to accumulate GC
159 substitutions faster as a result of gBGC. gBGC promotes the fixation of A/T to G/C mutations
160 and the loss of G/C to A/T mutations by preferentially replacing A/T bases with G/C bases
161 when recombination occurs at a heterozygous locus (*26*). gBGC is predicted to have a
162 greater influence when N_e is large (*26*). We observe a higher GC-content in high-
163 recombination regions of both pigeon species' genomes (fig. S8), indicating a long-term
164 influence of gBGC. We also observe a higher rate of A/T to G/C substitution and a lower rate
165 of G/C to A/T substitution in passenger pigeons than in band-tailed pigeons, indicating a
166 greater influence of gBGC in passenger pigeons (Fig. 4A,B).
167
168 The purging of deleterious G/C mutations or fixing of beneficial A/T mutations could create
169 the appearance of a greater efficiency of selection in passenger pigeons (*25*). This is
170 apparent in our observation that in regions of the passenger pigeon genome with high
171 recombination rates (and high diversity) there is a both a higher rate of nonsynonymous
172 substitution relative to synonymous substitution (*dN*/*dS*) for substitutions opposed by gBGC
173 and a lower *dN*/*dS* for substitutions promoted by gBGC (Fig. 4C,D and fig. S9). We also find

that gBGC influences ω_a and *pN*/*pS* (figs. S10 and S11). To test whether our inference of more efficient selection in passenger pigeons is an artefact of gBGC, we estimated ω_a and *pN*/*pS* separately for G/C to G/C and A/T to A/T mutations, which are unaffected by gBGC. For these mutations, we again observed higher ω_a and lower *pN*/*pS* in passenger pigeons than in band-tailed pigeons (figs. S10 and S11), confirming that passenger pigeons experience more efficient selection. However, when comparing high- and low-diversity regions of the passenger pigeon genome, we only observe a difference in *pN*/*pS*. This indicates that differences in ω_a across the passenger pigeon genome may have been driven by gBGC.

Passenger pigeons' low genetic diversity has been explained as the result of drastic population fluctuations driven by resource availability on the basis of Pairwise Sequentially Markovian Coalescent (PSMC) analyses of the nuclear genome (*4*; *14*). In contrast, our analyses reveal both population stability preceding the species' extinction and a surprisingly pervasive influence of natural selection. Moreover, the extent of the influence of selection across the passenger pigeon genome indicates that analyses such as PSMC are unlikely to reliably inform us of demographic history (*14*). Our results therefore do not support the hypothesis that natural demographic fluctuations contributed to the passenger pigeon's extinction, and instead suggest that following the onset of the commercial harvest, traits that were adaptive when their population size was large may have made it more difficult for passenger pigeons to survive when their population size was diminished (*2*). More broadly, our results suggest that even species with large and stable population sizes can be at risk of extinction following a sudden environmental change.

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), *Current Ornithology*, pp. 1–36.

3. B. Charlesworth, Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009).

4. C.-M. Hung *et al.*, Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10636–10641 (2014).

5. J. Maynard 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. J. H. Gillespie, Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics*. **155**, 909–919 (2000).

8. B. Charlesworth, The Effects of Deleterious Mutations on Evolution at Linked Sites. *Genetics*. **190**, 5–22 (2012).

9. R. C. Lewontin, *The Genetic Basis of Evolutionary Change* (Columbia University Press, 1974).

10. 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).

11. E. M. Leffler *et al.*, Revisiting an Old Riddle: What Determines Genetic Diversity Levels

within Species? *PLoS Biol.* **10**, e1001388 (2012).

12. 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).

13. 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).

14. Materials and methods are available as supplementary materials.

15. H. Ellegren, Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol. Evol.* **25**, 283–291 (2010).

16. R. Hershberg, D. A. Petrov, Selection on codon bias. *Annu. Rev. Genet.* **42**, 287–299 (2008).

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

18. B. T. Grenfell, A. P. Dobson, *Ecology of Infectious Diseases in Natural Populations* (Cambridge University Press, 1995).

19. S. Creel, B. Dantzer, W. Goymann, D. R. Rubenstein, The ecology of stress: effects of the social environment. *Funct. Ecol.* **27**, 66–80 (2013).

20. W. G. Hill, A. Robertson, The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).

21. B. M. Van Doren *et al.*, Correlated patterns of genetic diversity and differentiation across an avian family. *Mol. Ecol.* (2017), doi:10.1111/mec.14083.

22. L. Dutoit *et al.*, 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).

23. K. Nam *et al.*, Molecular evolution of genes in avian genomes. *Genome Biol.* **11**, R68 (2010).

24. 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).

25. 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).

26. L. Duret, A. Eyre-Walker, N. Galtier, A new perspective on isochore evolution. *Gene*. **385**, 71–74 (2006).

27. H. N. Poinar, A. Cooper, Ancient DNA: do it right or not at all. *Science*. **5482**, 1139 (2000).

28. 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).

29. J. Dabney *et al.*, 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).

30. 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).

31. B. Shapiro *et al.*, Flight of the dodo. *Science*. **295**, 1683 (2002).

266 32. T. L. Fulton, M. Stiller, “PCR Amplification, Cloning, and Sequencing of Ancient DNA”, in
267 *Ancient DNA*, B. Shapiro, M. Hofreiter, Eds. (Humana Press), *Methods in Molecular*
268 *Biology*, pp. 111–119 (2012).

269 33. N. Rohland, D. Reich, Cost-effective, high-throughput DNA sequencing libraries for
270 multiplexed target capture. *Genome Res.* **22**, 939–946 (2012).

271 34. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed
272 target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448 (2010).

273 35. A. E. R. Soares *et al.*, Complete mitochondrial genomes of living and extinct pigeons
274 revise the timing of the columbiform radiation. *BMC Evol. Biol.* **16**, 230 (2016).

275 36. C.-M. Hung *et al.*, The de novo assembly of mitochondrial genomes of the extinct
276 passenger pigeon (*Ectopistes migratorius*) with next generation sequencing. *PLoS One*.
277 **8**, e56301 (2013).

278 37. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high
279 throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

280 38. M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: A multiplatform graphical user
281 interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**,
282 221–224 (2010).

283 39. L. Excoffier, H. E. L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform
284 population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–
285 567 (2010).

286 40. A. G. F. Teacher, D. J. Griffiths, HapStar: automated haplotype network layout and
287 visualization. *Mol. Ecol. Resour.* **11**, 151–153 (2011).

288 41. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with

289 BEAUti and the BEAST 1.7*. Mol. Biol. Evol*. 29, **8**, 1969–1973 (2012).

290 42. A. J. Drummond, A. Rambaut, B. Shapiro, O. G. Pybus, Bayesian coalescent inference
291 of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192
292 (2005).

293 43. S. Y. W. Ho, B. Shapiro, Skyline-plot methods for estimating demographic history from
294 nucleotide sequences. *Mol. Ecol. Resour.* **11**, 423–434 (2011).

295 44. B. Nabholz, R. Lanfear, J. Fuchs, Body mass-corrected molecular rate for bird
296 mitochondrial DNA. *Mol. Ecol.* **25**, 4438–4449 (2016).

297 45. S. Ho, G. Larson, Molecular clocks: when timesare a-changin'. *Trends Genet.* **22**, 79–83
298 (2006).

299 46. A. G. Rambaut, A. J. Drummond, Tracer v1.6, Available from
300 *http://beast.bio.ed.ac.uk/Tracer* (2014).

301 47. A. Rambaut, A. J. Drummond, TreeAnnotator v1. 7.0. *University of Edinburgh, Institute*
302 *of Evolutionary Biology* (2013).

303 48. E. Bazin, S. Glémin, N. Galtier, Population size does not influence mitochondrial genetic
304 diversity in animals. *Science*. **312**, 570–572 (2006).

305 49. J. A. Chapman *et al.*, Meraculous: de novo genome assembly with short paired-end
306 reads. *PLoS One*. **6**, e23501 (2011).

307 50. N. H. Putnam *et al.*, Chromosome-scale shotgun assembly using an in vitro method for
308 long-range linkage. *Genome Res.* **26**, 342–350 (2016).

309 51. D. Kim *et al.*, TopHat2: accurate alignment of transcriptomes in the presence of
310 insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).

52. M. Stanke *et al.*, AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–9 (2006).

53. L. W. Hillier *et al.*, Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. **432**, 695–716 (2004).

54. M. D. Shapiro *et al.*, Genomic diversity and evolution of the head crest in the Rock Pigeon. *Science*. **339**, 1063–1067 (2013).

55. S. F. Altschul *et al.*, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

56. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013), (available at http://arxiv.org/abs/1303.3997).

57. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).

58. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26**, 841–842 (2010).

59. M. A. DePristo *et al.*, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

60. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics*. **27**, 2156–2158 (2011).

61. K. Prüfer *et al.*, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. **505**, 43–49 (2014).

62. 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).

63. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. **20**, 289–290 (2004).

64. C. H. Langley *et al.*, Genome variation in natural populations of *Drosophila melanogaster*. *Genetics*. **192**, 533-598 (2012).

65. 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).

66. T. Lindahl, Instability and decay of the primary structure of DNA. *Nature*. **362**, 709–715 (1993).

67. P. D. Heintzman, A. E. R. Soares, D. Chang, B. Shapiro, Paleogenomics. *Reviews in Cell Biology and Molecular Medicinem*. **1**, 243–267 (2015).

68. 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).

69. 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).

70. T. I. Gossmann, P. D. Keightley, A. Eyre-Walker, The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol. Evol.* **4**, 658–667 (2012).

71. N. Stoletzki, A. Eyre-Walker, Estimation of the Neutrality Index. *Mol. Biol. Evol.* **28**, 63–70 (2011).

72. J. A. Shapiro *et al.*, Adaptive genic evolution in the *Drosophila* genomes. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2271–2276 (2007).

356 73. J. A. Novembre, Accounting for background nucleotide composition when measuring
357 codon usage bias. *Mol. Biol. Evol.* **19**, 1390–1394 (2002).

358 74. F. Wright, The “effective number of codons” used in a gene. *Gene*. **87**, 23–29 (1990).

359 75. H. Akashi, Inferring weak selection from patterns of polymorphism and divergence at“
360 silent” sites in *Drosophila* DNA. *Genetics*. **139**, 1067–1076 (1995).

361 76. S. Karlin, J. Mrázek, What drives codon choices in human genes? *J. Mol. Biol.* **262**,
362 459–472 (1996).

363 77. N. Backström *et al.*, The recombination landscape of the zebra finch *Taeniopygia*
364 *guttata* genome. *Genome Res.* **20**, 485–495 (2010).

365 78. T. Kawakami *et al.*, A high-density linkage map enables a second-generation collared
366 flycatcher genome assembly and reveals the patterns of avian recombination rate
367 variation and chromosomal evolution. *Mol. Ecol.* **23**, 4035–4058 (2014-8).

368 79. S. Singhal *et al.*, Stable recombination hotspots in birds. *Science*. **350**, 928–932 (2015).

369 80. A. Auton, G. McVean, Recombination rate estimation in the presence of hotspots.
370 *Genome Res.* **17**, 1219–1227 (2007).

371 81. H. Brunschwig *et al.*, Fine-scale maps of recombination rates and hotspots in the mouse
372 genome. *Genetics*. **191**, 757–764 (2012).

373 82. H. Li, R. Durbin, Inference of human population history from whole genome sequence of
374 a single individual. *Nature*. **475**, 493–496 (2011).

375 83. D. R. Schrider, A. G. Shanku, A. D. Kern, Effects of linked selective sweeps on
376 demographic inference and model selection. *Genetics*, **204**, 3, 1207-1223 (2016).

377 84. R. Frankham, Effective population size/adult population size ratios in wildlife: a review.

Genet. Res. **89**, 491–503 (2007).

85. F. P. Palstra, D. J. Fraser, Effective/census population size ratio estimation: a compendium and appraisal. *Ecol. Evol.* **2**, 2357–2365 (2012).

86. T. Karasov, P. W. Messer, D. A. Petrov, Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLoS Genet.* **6**, e1000924 (2010).

87. N. Barton, Understanding adaptation in large populations. *PLoS Genet.* **6**, e1000987 (2010).

88. J. H. Gillespie, Is the population size of a species relevant to its evolution? *Evolution*. **55**, 2161–2169 (2001).

89. B. H. Good, A. M. Walczak, R. A. Neher, M. M. Desai, Genetic diversity in the interference selection limit. *PLoS Genet.* **10**, e1004222 (2014).

90. J. C. Stanton, Present-day risk assessment would have predicted the extinction of the passenger pigeon (*Ectopistes migratorius*). *Biol. Conserv.* **180**, 11–20 (2014).

91. H. Ellegren *et al.*, The genomic landscape of species divergence in Ficedula flycatchers. *Nature*. **491**, 756–760 (2012).

92. G. Zhang *et al.*, Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*. **346**, 1311–1320 (2014).

93. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

94. J. Charlesworth, A. Eyre-Walker, The McDonald–Kreitman test and slightly deleterious mutations. *Mol. Biol. Evol.* **25**, 1007–1015 (2008).

95. K. Nam *et al.*, Evidence that the rate of strong selective sweeps increases with

400 population size in the great apes. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 1613–1618
401 (2017).

402 96. S. W. Schaeffer, Molecular population genetics of sequence length diversity in the *Adh*
403 region of *Drosophila pseudoobscura*. *Genet. Res.* **80**, 163–175 (2002).

404 97. K. J. Schmid, S. Ramos-Onsins, H. Ringys-Beckstein, B. Weisshaar, T. Mitchell-Olds, A
405 multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure
406 from a neutral model of DNA sequence polymorphism. *Genet.* **169**, 1601–1615 (2005).

407 98. F. Tajima, The effect of change in population size on DNA polymorphism. *Genetics*.
408 **123**, 597–601 (1989).

409 99. J. C. Fay, C. I. Wu, Hitchhiking under positive Darwinian selection. *Genetics*. **155**,
410 1405–1413 (2000).

411 100. R. E. Green *et al.*, A Draft Sequence of the Neandertal Genome. *Science*. **328**, 710–
412 722 (2010).

413 101. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, Testing for ancient admixture
414 between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).

415 102. P. W. Messer, D. A. Petrov, Frequent adaptation and the McDonald–Kreitman test.
416 *Proceedings of the National Academy of Sciences*. **110**, 8615–8620 (2013).

417 103. P. U. Clark *et al.*, The Last Glacial Maximum. *Science*. **325**, 710–714 (2009).

418 104. M. Karikoski *et al.*, Clever-1/Stabilin-1 regulates lymphocyte migration within
419 lymphatics and leukocyte entrance to sites of inflammation. *Eur. J. Immunol.* **39**, 3477–
420 3487 (2009).

421 105. S. Tong, J. Li, J. R. Wands, Carboxypeptidase D Is an avian Hepatitis B virus

receptor. *J. Virol.* **73**, 8696–8702 (1999).

106. 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).

107. 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).

108. M. Scherma *et al.*, 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).

109. S. Strindberg *et al.*, Thromboelastography in Selected Avian Species. *J. Avian Med. Surg.* **29**, 282–289 (2015).

110. Y. Meng *et al.*, Genome-wide analysis of positively selected genes in seasonal and non-seasonal breeding species. *PLoS One*. **10**, e0126736 (2015).

111. M. Ben Khelifa *et al.*, 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).

112. S. Labeit, C. A. C. Ottenheijm, H. Granzier, Nebulin, a major player in muscle health and disease. *FASEB J.* **25**, 822–829 (2011).

113. Z. Lu *et al.*, Identification of Soat1 as a quantitative trait locus gene on mouse chromosome 1 contributing to hyperlipidemia. *PLoS One*. **6**, e25344 (2011).

114. M. P. Richards *et al.*, Feed restriction significantly alters lipogenic gene expression in broiler breeder chickens. *J. Nutr.* **133**, 707–715 (2003).

115. M. Benn, Apolipoprotein B levels, APOB alleles, and risk of ischemic cardiovascular disease in the general population, a review. *Atherosclerosis*. **206**, 17–30 (2009).

116. S. Liu *et al.*, Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell*. **157**, 785–794 (2014).

117. 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).

118. 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).

119. K. Silander *et al.*, Genetic variation near the hepatocyte nuclear factor-4 alpha gene predicts susceptibility to type 2 diabetes. *Diabetes*. **53**, 1141–1149 (2004).

120. S. Vasu *et al.*, Novel vertebrate nucleoporins Nup133 and Nup160 play a role in mRNA export. *J. Cell Biol.* **155**, 339–354 (2001).

121. S. Tang, D. C. Presgraves, Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science*. **323**, 779–782 (2009).

122. S. Cappello *et al.*, Mutations in genes encoding the cadherin receptor-ligand pair DCHS1 and FAT4 disrupt cerebral cortical development. *Nat. Genet.* **45**, 1300–1308 (2013).

123. 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).

124. A. R. R. Eagle *et al.*, Meprin β metalloprotease gene polymorphisms associated with diabetic nephropathy in the Pima Indians. *Hum. Genet.* **118**, 12 (2005).

125. U. D. P. Lam *et al.*, Association of MEP1A gene variants with insulin metabolism in central European women with polycystic ovary syndrome. *Gene*. **537**, 245–252 (2014).

126. C. Albrecht, E. Viturro, The ABCA subfamily—gene and protein structures, functions and associated hereditary diseases. *Pflugers Arch - Eur J Physiol*. **453**, 581–589 (2007).

127. K. Tatebe *et al.*, Response network analysis of differential gene expression in human epithelial lung cells during avian influenza infections. *BMC Bioinformatics*. **11**, 170 (2010).

128. C. Pattaro *et al.*, 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).

129. 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).

130. 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).

Acknowledgments:

We thank L. Shiue, S. Weber, J. Kapp, M. Stiller, T. Kuhn, S. Wagner, and R. Shaw for assistance generating data. We thank J. Novembre for advice on analysing codon usage bias. Research was supported by the Packard Foundation, the Gordon and Betty Moore Foundation, and Revive & Restore. A.E.R.S. was funded by Ciência sem Fronteiras fellowship - CAPES, Brazil. Sequencing was supported by the Dean's Office, the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley (Berkeley sequencing supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303), and the Danish National Sequencing Centre in Copenhagen (sequencing supported by Lundbeck Foundation grant R52-5062). The sequence data generated in this study are archived in the relevant NCBI databases: the band-tailed pigeon assembly and RNA-seq reads used for its annotation can be found in Bioproject PRJNA308039 and reads from passenger pigeon samples in PRJNA381231 (accession numbers are provided in supplementary table 1).

Competing interests

The authors declare no competing interests.

Author contributions:

B.S. conceived and designed the study with critical input from G.G.R.M, A.E.R.S, R.E.G, and R.B.C-D.; B.S., T.L.F, and B.J.N. led sample collection; A.J.B., A.D., J.R.D., A.G.S., K.S., G.S., M.T.P.G., and M.P. provided samples; A.E.R.S., T.L.F., B.L., B.J.N, and R.R.DaF performed DNA extraction and library preparation experiments; A.E.R.S and P.D.H performed mitochondrial genome assembly and analyses; A.E.R.S, N.K.S, E.S.R, J.A.C., S.H.V., and P.D.H. performed nuclear genome assembly and analyses; G.G.R.M. designed and performed selection analyses; B.S., G.G.R.M, A.E.R.S, and R.E.G. wrote the paper; and all authors contributed to editing the manuscript.

Figures:

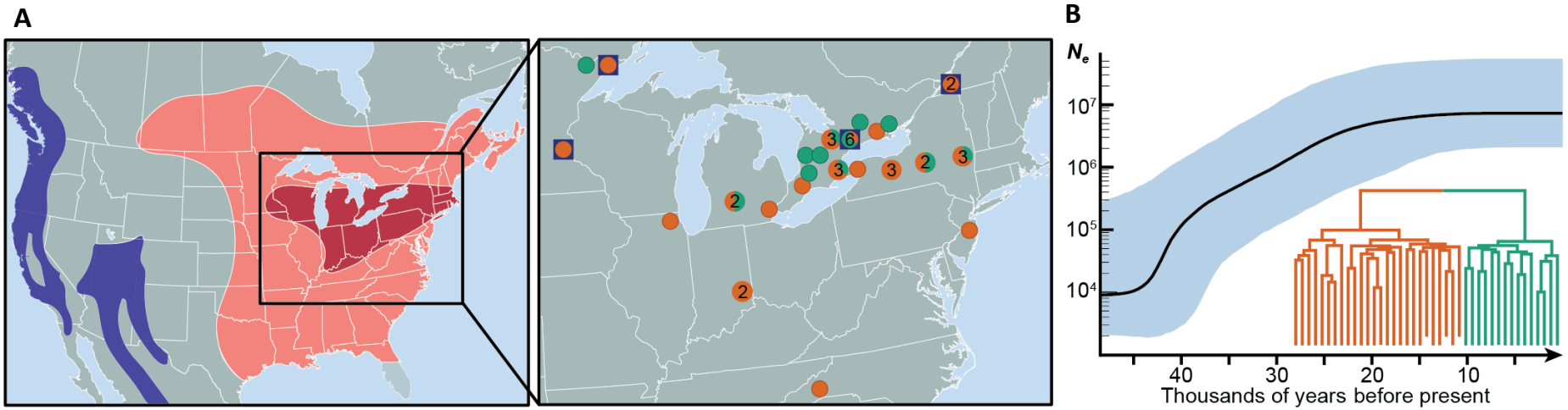
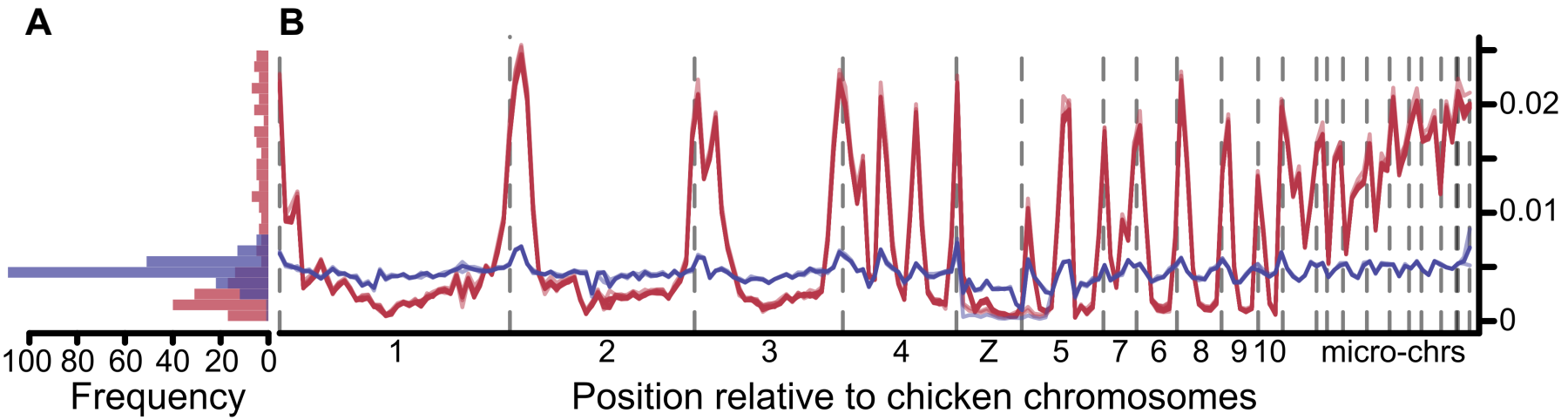


Fig. 1. Passenger pigeon range, sample origins, and N_e estimate from mitochondrial genomes. (**A**) Range of passenger pigeons at time of European contact (dark red: breeding range; light red: full range) (*1*) and current range of band-tailed pigeons (purple) (*12*), with inset showing the location of origin of the 41 passenger pigeon samples analyzed here. Locations of the four samples from which nuclear genomes were generated are indicated with a blue box. (**B**) Inferred N_e and mitochondrial phylogeny from a Bayesian coalescent analysis. Colors in (**A**) inset match the phylogeny in (**B**). The structure of the phylogeny does not correlate with geography, which is consistent with an absence of geographic population structure.



526

527

528 **Fig. 2**. ***π*** **across passenger pigeon and band-tailed pigeon genomes.** (**A**) A histogram describing mean π for 5 Mb windows across the

529 passenger pigeon (red) and band-tailed pigeon (blue) genomes. (**B**) Genomic distribution of individual pairwise estimates of mean π in 5 Mb

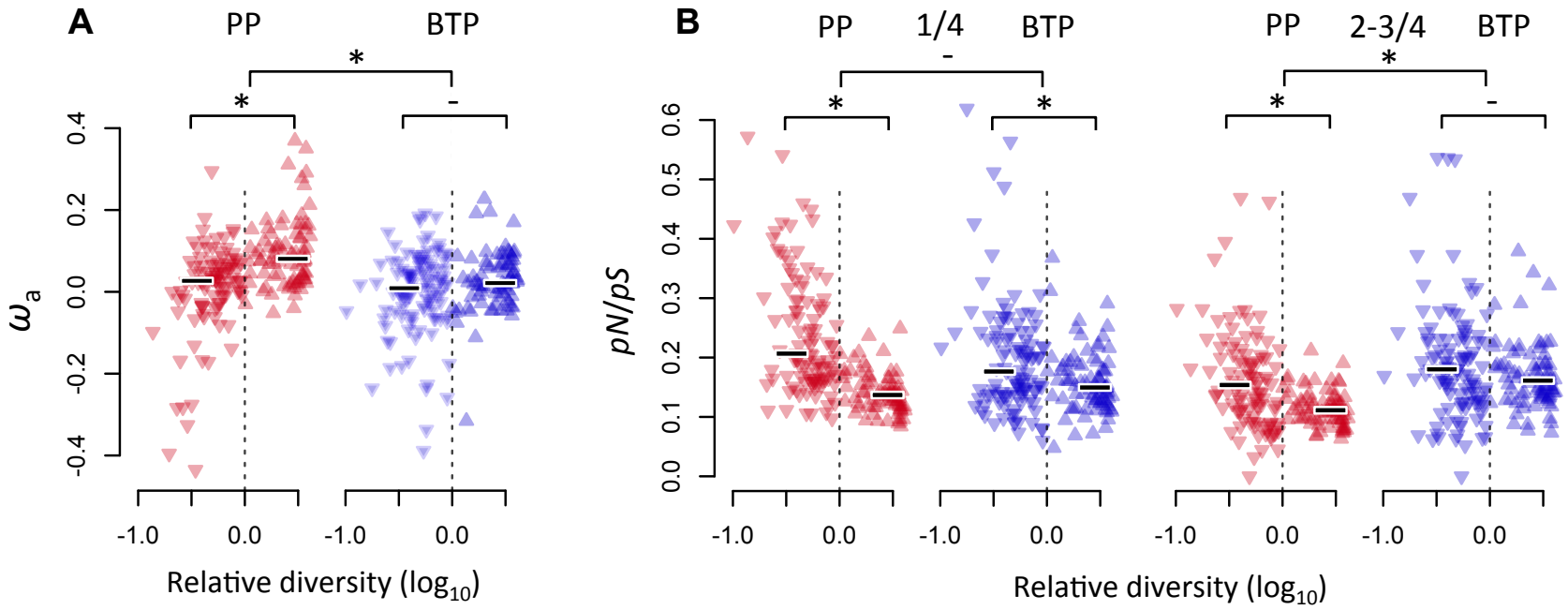
530 windows across the two species' genomes. Each between- and within-individual pairwise comparison is plotted as red (28 passenger pigeon

531 comparisons) or blue (6 band-tailed pigeon comparisons) lines. Chromosome boundaries are indicated as vertical dashed lines. Chromosomes

532 are ordered by their size in the chicken genome.

533

534



535

536 **Fig. 3. Estimates of ω_a and *pN/pS*.** Estimates are averages for 5 Mb windows and are plotted against the window's genetic diversity in
537 passenger pigeons relative to band-tailed pigeons (on a $\log_{10}$-scale). Comparisons are drawn between (**A**) ω_a and (**B**) *pN/pS* in passenger
538 pigeons (PP; red) and band-tailed pigeons (BTP; blue), and between low-diversity ($\pi_{PP} < \pi_{BTP}$; point-down triangles) and high-diversity ($\pi_{PP} >$
539 π_{BTP}; point-up triangles) windows (median values are shown as horizontal lines; '*' indicates $p \leq 1\text{x}10^{-4}$ and '-' $p \geq 0.1$ in a Mann-Whitney U test).
540 In (**B**) *pN/pS* estimates are for derived mutations present in 1/4 and 2-3/4 individuals. A higher *pN/pS* for lower frequency mutations could
541 reflect the slow purging of weakly deleterious mutations. Estimates are based on analyses of two individuals from each species (see figure S6
542 for estimates using all passenger pigeon samples).

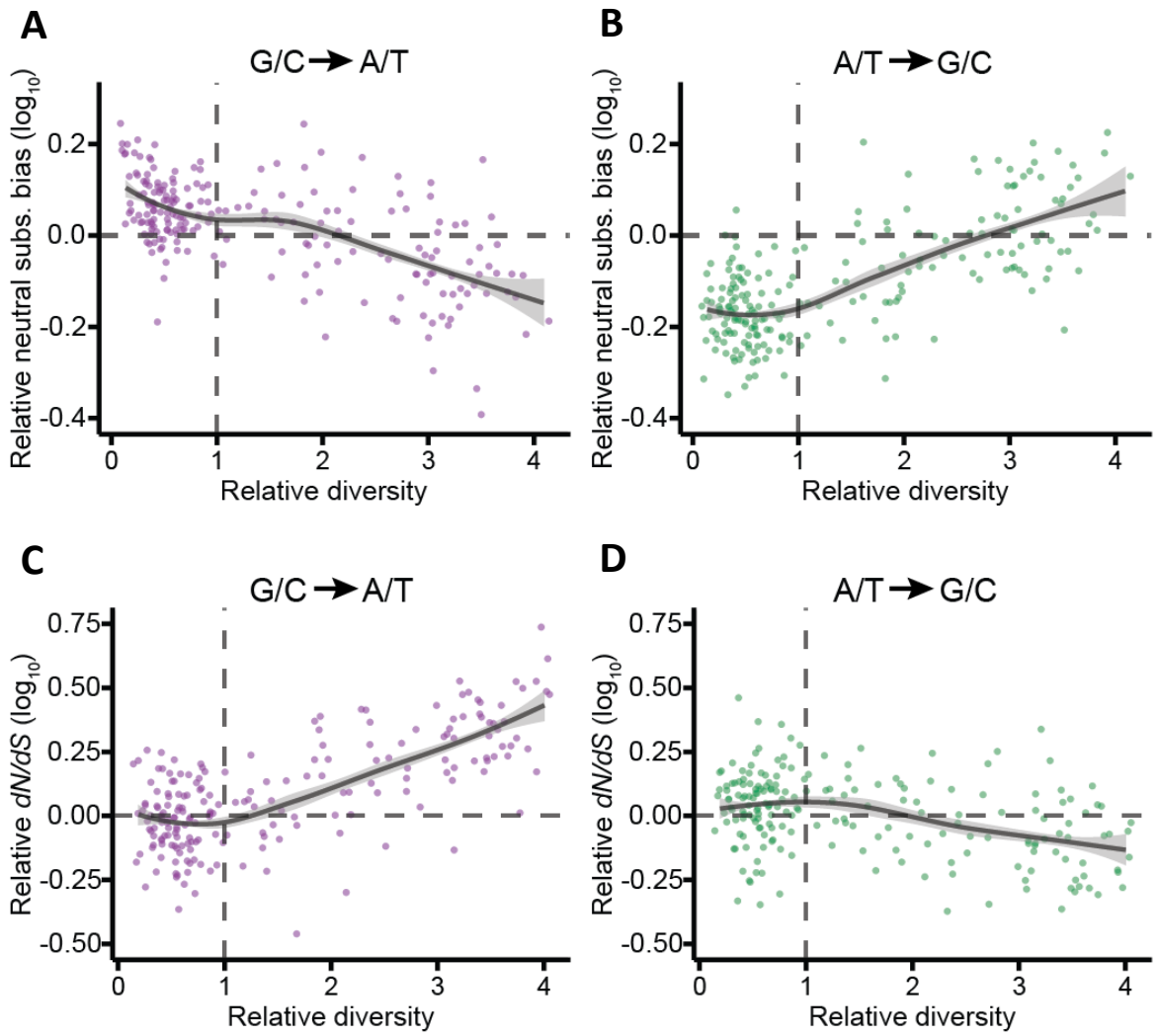


Fig. 4. Patterns of substitution for nucleotide base changes that are opposed (A, C) and promoted (B, D) by gBGC. (**A**) The rate of G/C to A/T substitution relative to G/C to G/C substitution in passenger pigeons, divided by the same parameter in band-tailed pigeons. (**B**) The rate of A/T to G/C substitution relative to A/T to A/T substitution in passenger pigeons lineage, divided by the same parameter in band-tailed pigeons. (**C**) *dN*/*dS* for G/C to A/T mutations in passenger pigeons, divided by the same parameter in band-tailed pigeons. (**D**) *dN*/*dS* for A/T to G/C mutations in passenger pigeons, divided by the same parameter in band-tailed pigeons. All estimates are for 5 Mb windows across the genome, and are plotted on a log_{10}-scale against diversity in passenger pigeons relative to band-tailed pigeons. Trend lines were estimated using the 'stat_smooth' function in *ggplot2* (method = 'loess') in *R*. Shading reflects 95% confidence limits around the trend lines.

Supplementary Materials

Contents

Text

Figures and Tables

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 SRS622857) and BMNH794 (SRA SRS622896 (*4*)). We chose to not include the two other samples from (*4*) in our analyses, since one had a much lower coverage than the other genomes, and both had shorter assembled genome lengths. Due to the additional filtering steps we applied prior to our reassembly of this data, the lower quality of these genomes would have resulted in the exclusion of a large number of sites from our analyses.

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 (*27*), 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 (*28*, *29*), and from toe pads using the Qiagen Blood and Tissue Kit, with ancient DNA-specific modifications (*30*). We purified the digested samples using either the Qiagen DNeasy extraction protocol (*30*), Qiagen Nucleotide Removal Kit, or "in-house" silica columns (*28*).

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 (*30*) and a 136 bp fragment of the mitochondrial cytochrome *b* gene, using the primers 5'-CAAAGAAACCTGAAACACAGG (*31*) (forward) and 5'-GGGACAGCCGAGAATAGGTT (reverse). We performed PCR following (*32*), but with an annealing temperature of 48˚C for cytochrome *b*. We cleaned PCR products using the MagNA bead protocol (*33*), and assessed damage and potential contamination via molecular cloning and Sanger sequencing (*32*). 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 (*34*). 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.

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 mapping assembler that uses an ancient DNA-specific substitution matrix. To avoid calling bases that could be the result of ancient DNA damage, we also required a minimum of three agreeing and independent reads to call a base at each site, and 2/3 agreement between mapped reads that exceeded the minimum 3x coverage. Sites not meeting these criteria were were classed 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). The high coverage obtained for the mitochondrial genomes, the iterative mapping strategy used by MIA, and the ancient DNA-specific substitution matrix greatly reduce the possibility of calling sites that could be result from ancient DNA damage.

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) (*35*), and assembled mitochondrial genomes as above from published short read data for BMNH1149 (SRA SRX555773) (*36*) and BMNH794 (SRA SRX555813) (*4*).

We aligned all 41 passenger pigeon mitochondrial genomes using MUSCLE (*37*) as implemented in SeaView v.4 (*38*). 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 0.0018 differences per site.

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 (*39*), which we visualised using HapStar v0.7 (*40*). The MSN displays star radiations, with four to sixteen nucleotide differences between haplotypes (fig. S13), which is consistent with a population expansion or a selective sweep.

The structure present in this network and in the phylogeny estimated as part of the Bayesian skyline plot analysis described below (Fig. 1B) suggests the presence of two clades, separated by 16 nucleotide differences. These two clades are not correlated with geography (Fig. 1A). This is consistent with 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 (*39*). A negative Tajima's D value indicates an excess of low frequency polymorphisms, which could result from population growth, a selective sweep or purifying selection. Similarly, a negative F_S value indicates an excess number of alleles, which could result from population growth or a selective sweep.

Bayesian skyline plot analysis

We estimated the coalescent history of the mitochondrial genomes using the Bayesian genealogical inference package, BEAST v1.8.1 (*41*). We assumed the HKY+Γ nucleotide substitution model and the skyline plot model of the coalescent process (*42*, *43*). 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 a lineage-specific evolutionary rate for passenger pigeons, both for all sites in the mitochondrial genome and for third codon positions only (*44*). Since these rate estimates are based on divergence between species and evolutionary rates are likely to be faster within a species than between species (*45*), neither are likely to be accurate for our data. However, because mitochondrial rates also vary across bird species (*46*), a rate estimate cannot be easily be obtained from another species. We therefore used the rate estimate for 3rd codon positions in the passenger pigeon mitochondria (1.25×10^{-8} substitutions/site/year) from (*44*), knowing that since most sites in the mitochondria evolve slower than 3rd codon positions, this rate is unlikely to be slower than the true rate. This means that our estimates of N_e are likely to be lower than the true values and our estimates of the dates of demographic changes younger than the true dates. 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 (*46*) and calculated the maximum clade credibility tree using TreeAnnotator v1.8 (*47*).

For comparison, we also ran the Bayesian skyline plot analysis using the long-term rate estimate for all sites (3.0×10^{-9} substitutions/site/year) (fig. S14). This is likely to represent the slowest possible rate at which the variable sites present in our data evolved. 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.

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 (*35*)) 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.

In addition to this evidence of purifying selection affecting diversity in the passenger pigeon mitochondria, evidence has been found that mitochondrial diversity is generally a poor predictor of population size due to the impact of selective sweeps (*48*). An impact of selection (either selective sweeps or purifying selection) on passenger pigeon mitochondrial genomes means that the results of our analysis of the passenger pigeon mitochondria using the skyline plot model may not reflect demographic change, i.e. the increase in N_e we infer may not reflect a population expansion, but might instead reflect the impact of a selective sweep or purifying selection on genetic diversity. Moreover, it could explain why the N_e we infer from the mitochondrial genome is so much lower than passenger pigeons' census population size. Nevertheless, we consider it likely that the results of the Bayesian skyline plot analysis reflect a minimum bound to the recent N_e of the passenger pigeon population, and the length of time over which this N_e was stable, since both selective sweeps and purifying selection act to reduce N_e (and reduce estimates of historic N_e).

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 (*49*). We provided HMW DNA from BTP2014 to Dovetail Genomics (Santa Cruz, CA), who prepared and sequenced a Chicago library (*50*) from this extract. They used the resulting data to scaffold our *de novo* assembly with their software, HiRise (*50*). 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 (*51*). We then used these alignments to generate intron annotation hints using the bam2hints script in the AUGUSTUS package with default options (*52*). 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) (*53*) and rock dove (*Columba livia*) (*54*) genomes. We performed reciprocal BLASTP (*55*) 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.

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 (*56*). We sorted, indexed, and removed duplicates using samtools (*57*) and calculated genome coverage statistics using genomeCoverageBed (*58*).

We genotyped each individual using GATK 3.3 UnifiedGenotyper (*59*). 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 (*60*) and a custom Python program to filter our variant call set (available at https://github.com/Paleogenomics/DNA-Post-Processing/blob/master/vcf2fa_allsites.py), in an approach similar to Prüfer *et al. (61)*. 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 5x. 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 (*57*) 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 (*54*). 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 (*62*). 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.

5. Estimation of diversity and divergence within and between genes

We estimated nucleotide 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* (*63*) to estimate pairwise distances between pseudo-haploid genomes. 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 (*54*) to infer the most likely ancestral state. We counted only differences that were fixed in both passenger pigeons and band-tailed pigeons, and 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.

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 (*52*) and 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 biallelic polymorphic sites at different frequencies in each species, using the rock dove to polarize variants. 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.

To assess the potential impact of variation in mutation rate on variation in genetic diversity across the genome, we estimated the rate of divergence between one genome from each

391 species, for non-coding regions. We partitioned rates onto the passenger pigeon and band-
392 tailed pigeon lineages. We estimated the rate of divergence for all types of nucleotide
393 mutations and the rate for only those that would have been unaffected by GC-biased gene
394 conversion (i.e. G/C to G/C and A/T to A/T). We estimated rates for 5 Mb windows along the
395 scaffolds, ordered according to our mapping to the chicken genome.
396
397 We found that while the variance in nucleotide diversity across the non-coding regions of the
398 passenger pigeon genome is 4.89×10^{-5}, the variance in the rate of divergence along the
399 passenger pigeon lineage is only 7.52×10^{-6} for all mutations and 1.75×10^{-7} for mutations
400 unaffected by GC-biased gene conversion (fig. S15). This suggests that GC-biased gene
401 conversion is driving some variation in the rate of divergence across the passenger pigeon
402 genome. Therefore, considering only mutations unaffected by GC-biased gene conversion,
403 we find that the variance in nucleotide diversity across the passenger pigeon genome is 279
404 times greater than the variance in the rate of divergence.
405
406 The variance in nucleotide diversity in the non-coding regions of the band-tailed pigeon
407 genome is 1.04×10^{-6} (47 times less than in the passenger pigeon). The variance in the rate
408 of divergence along the band-tailed pigeon lineage is 1.70×10^{-6} for all mutations and
409 6.37×10^{-8} for mutations unaffected by GC-biased gene conversion (fig. S15). Therefore,
410 considering only mutations unaffected by GC-biased gene conversion, the variance in
411 nucleotide diversity across the band-tailed pigeon genome is 16 times greater than variance
412 in the rate of divergence.
413
414 We found evidence of a positive correlation between diversity and divergence for both
415 species (fig. S15). This is not unexpected as variation in mutation rates across the genome
416 will contribute to variation in levels of genetic diversity. Moreover, some correlation between
417 divergence and diversity is expected as a consequence of the action of linked selection on
418 levels of polymorphism in the ancestors of passenger pigeons and band-tailed pigeons (as in
419 e.g. *64*). However, the degree of variation in rates of divergence is far too small to explain
420 the variation in diversity we observe across the genomes of either species.
421
422 **6. Assessing the impact of ancient DNA damage**
423 Ancient DNA data can contain artefacts due to the process of DNA decay (*65*, *66*). Common
424 damage observed in ancient DNA data includes strand breakage and miscoding lesions,
425 with the latter primarily occurring through the hydrolytic deamination of cytosine to uracil,
426 particularly at the ends of sequences (*67*). If present, this type of damage can affect
427 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 (*57*). 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 (*67*). The amount of damage was found to be low; 4% or fewer of cytosines were deaminated at read ends (fig. S16).

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. S17). 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. S18). 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. It is possible that this difference between *Ts*/*Tv* in passenger pigeons and band-tailed pigeons was a consequence of the greater impact of GC-biased gene conversion on substitution rates in the passenger pigeon genome. GC-biased gene conversion affects all transitions, but only half of transversions, and therefore it is possible that it drove a lower *Ts*/*Tv* in passenger pigeons. 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.

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. S8), demonstrating that this is a robust result. We also tested for a correlation between genetic diversity and the average root mean squared map quality for 5 Mb windows across the genome (fig. S20). We did not find any evidence that poor map quality was associated with higher estimates of genetic diversity (Spearman's ρ = -0.033, p = 0.69).

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 $\omega_a = \alpha(dN/dS)$ (*69*, *70*) 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. S6). We defined high- and low-diversity regions as regions that have higher or lower genetic diversity in passenger pigeons than in band-tailed pigeons in regions outside of annotated coding regions (approximately dividing the genome in half). These statistics were inferred both along the passenger pigeon lineage and across the band-tailed pigeon lineage, using estimates of *pN* and *pS* from within both passenger pigeons and band-

tailed pigeons, and estimates of *dN* and *dS* estimated along both the passenger pigeon and band-tailed pigeon lineages (see Section 5).

We estimated these statistics for 5 Mb windows, instead of 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 (*35*), 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 genomes. These were defined as regions that have higher or lower nucleotide diversity outside of annotated coding regions of the genome in passenger pigeons compared to band-tailed pigeons. To mitigate any bias that might result from using genetic diversity as a proxy for recombination rate, we also used our estimates of population-scaled recombination in the band-tailed pigeon genome to divide the genome in two (fig. S12). We obtained similar results from both methods. We also estimated these statistics separately for different types of nucleotide base changes to compare biases across lineages.

We estimated the direction of selection (DoS = $dN/(dN+dS) - pN/(pN+pS)$) (*71*) 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 (*72*). 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 (*17*). We implemented this test using a two-tailed Fisher's exact test in *R*.

8. Comparisons of codon usage bias

Codon usage bias is likely to reflect both mutational biases and natural selection for translational optimization (*16*). 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 in those of band-tailed pigeons, and also was stronger in high-diversity regions of the 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 (*73*). In particular, we estimated the effective number of codons after accounting for background nucleotide composition (*73*, *74*), which is inversely proportional to the extent of nonuniform codon usage. We also estimated Akashi's scaled χ^2 (*75*) and the *B**(a) measure of Karlin and Mrazek (*76*), 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. S20), and higher in regions of the genome that have high-diversity in passenger pigeons (here defined as 5 Mb windows that have greater than or less than 0.004 nucleotide diversity in passenger pigeons) (fig. S21). While we observed this in both passenger pigeons and band-tailed pigeons, we found that the difference was greater in passenger pigeons.

9. Estimation of the population-scaled recombination rate

Bird genomes have an unusually stable chromosome structure, and are highly syntenic (*15*). 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 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. S22; Spearman's r^2 = 0.68, Spearman's rank correlation test $p < 2.2 \times 10^{-16}$). 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 across the band-tailed pigeon genome. However, we find little evidence of substantial variation in either diversity or the efficacy of selection across the band-tailed pigeon genome, which suggests that there is no substantial variation in N_e. This means that the correlation between ρ and diversity in the passenger pigeon genome is most likely to have been driven by variation in the recombination rate.

10. Estimation of effective population size from the nuclear genome

We estimated the effective population size (N_e) of both passenger pigeons and band-tailed pigeons from their nuclear genomes, first based on an estimate of the average genetic diversity across their nuclear genomes, and second through application of a Pairwise Sequentially Markovian Coalescent (PSMC) model (*82*), which estimates the distribution of coalescence times across the genome, and can be used to infer changes in N_e over time. A previous study of passenger pigeon nuclear genomes (*4*) used a PSMC analysis of passenger pigeon nuclear genomes to infer the demographic history of the species. Here, we demonstrate that natural selection had such a strong and widespread impact on diversity in the passenger pigeon nuclear genome that genome-wide genetic diversity and the distribution of coalescence times across the genome, cannot be interpreted purely as a consequence of the demographic history of the species (*83*).

610 N_e determines both the equilibrium level of neutral diversity in a population ($\pi = 4N_e\mu$, where
611 π = nucleotide diversity, and μ = mutation rate) and the influence of selection in determining
612 whether a beneficial mutation is fixed and a deleterious mutation purged (*3*). N_e can be
613 defined as the census size (N_c) of the equivalent Wright-Fisher population (*3*). The N_e of a
614 population is often lower than its N_c. There are several possible reasons for this, such as
615 population structure, changes in population size over time, inbreeding, and natural selection
616 (*3*). However, it is often assumed that selection has a limited influence on genetic diversity in
617 vertebrate populations, and therefore low N_e/N_c ratios are often interpreted as solely a
618 consequence of population structure (e.g. geographic structure) or historic population
619 fluctuations (e.g. bottlenecks). Our results demonstrate that in some vertebrate populations
620 the influence of selection on N_e can be substantial, and so a low N_e/N_c ratio may not solely
621 be a consequence of demography. We further demonstrate the impact of selection on the
622 results of PSMC analyses (*43*). Our results suggest that the influence of selection on
623 genomic diversity can be sufficiently great that historic changes in N_e inferred from such
624 analyses may not reflect demographic change.
625
626 *Estimation of average N_e across the genome*
627 Using the relationship $\pi = 4N_e\mu$ *(3)*, average estimates of π across the unannotated regions
628 of the nuclear genomes, and the estimate of the substitution rate for the rock dove genome
629 in place of the true mutation rate (2.84×10^{-9} substitutions/site/generation) (*54*), we estimated
630 that the N_e of the passenger pigeon nuclear genome was 7.0×10^5 and the N_e of the band-
631 tailed pigeon nuclear genome is 3.9×10^5. Our estimate of N_e for the passenger pigeon is
632 approximately twice that of Hung et al. (*4*). Their underestimation of N_e is likely due to an
633 inappropriate use of the parameter "-C50" while calling consensus with samtools mpileup.
634 This penalizes the mapping quality of reads containing mismatches, which are expected
635 when mapping reads against a different species. This would have resulted in many variable
636 sites being excluded from their analyses.
637
638 For both species, our estimates of N_e are less than their N_c. Since the band-tailed pigeon N_c
639 is estimated to be approximately 2 million individuals (*12*), N_e/N_c in band-tailed pigeons is
640 approximately 0.2. Since the passenger pigeon N_c is estimated to have been approximately
641 3-5 billion individuals prior to their rapid decline, we find that N_e/N_c in passenger pigeons was
642 approximately 0.0002, which is much lower than estimates for any other vertebrate species
643 (*84*). Comparisons across a wide range of taxa suggest that on average N_e/N_c is 0.1 (*84*, *85*),
644 and estimates for several bird species suggest that birds do not tend to have unusually low
645 N_e/N_c ratios; in (*84*) the bird N_e/N_c ratios range from 0.05 to 0.8.
646

Variation in N_e across the passenger pigeon genome

Variation in π across the passenger pigeon genome suggests regional differences in N_e. While we are unable to distinguish between the impact of different types of selection on diversity across the passenger pigeon genome, we consider it likely that both selective sweeps and background selection contributed to this variation. Our observation that π in the passenger pigeon genome is less than π in the band-tailed pigeon genome in the centers of macrochromosomes suggests that although on average N_e was greater in passenger pigeons than in band-tailed pigeons, N_e was lower in passenger pigeons for some regions of the genome. However, all the possible targets of selection in the passenger pigeon genome may not be included within our annotation of the band-tailed pigeon genome, and therefore some of the sites used to estimate neutral diversity may have been targets of selection. This could mean that our lower estimates of genetic diversity for the passenger pigeon than for the band-tailed pigeon do not actually reflect a lower N_e for these genomic regions.

Another factor that may have contributed to the reduction in diversity in some regions of the passenger pigeon genome to below that of the band-tailed pigeon genome may have been the fixation of strongly beneficial mutations. The fixation probability of a strongly beneficial mutation is largely independent of the interference among linked sites that can affect local N_e *(86, 87).* Also, the rate of fixation of strongly beneficial mutations is correlated with N_c because a beneficial mutation is more likely to appear in a larger population. Selective sweeps of strongly beneficial mutations therefore can result in an independence of Ne from Nc *(7, 88)*. After correcting for GC-biased gene conversion, we find no evidence that the rate of adaptive substitution relative to synonymous substitution is depressed in regions of low diversity (or low recombination) in the passenger pigeon genome (fig. S10). This is consistent with the local rate of adaptive substitution having been largely independent of local reductions in N_e in the passenger pigeon genome.

We also note that after correcting for GC-biased gene conversion, we still find evidence for a difference in the impact of purifying selection across high and low diversity regions of the genome (fig. S10,11). This suggests that low rates of recombination in some regions of the passenger pigeon genome resulted in weakly deleterious mutations segregating in the population for longer, while still eventually being purged. This could drive greater interference among selected sites, and might further reduce diversity at neutral sites *(89)*.

Comparison between nuclear and mitochondrial genomes

When we use the relationship $\pi = 4N_e\mu$ to estimate N_e in the passenger pigeon mitochondrial genome (using the long-term substitution rate obtained for all sites or for only

3rd codon positions (*44*)) we obtain similar (although slightly lower) values than we do for the nuclear genome after scaling for the expected difference in N_e between mitochondrial and nuclear genomes (5.9×10^5 using the rate estimate for all sites, and 1.4×10^5 using the rate estimate for 3rd codon positions). These estimates of N_e do not, however, take into account the distribution of coalescence rates across the genome, which can inform us about how N_e has changed over time.

Application of the PSMC model to the nuclear genomes

We applied the PSMC to the nuclear genomes of both passenger pigeons and band-tailed pigeons. For this, 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 applied a PSMC analysis to high and low-diversity regions separately (using genetic diversity as a proxy for recombination rate). 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 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 (*90*) and a mutation rate of 5.68×10^{-9} substitutions/site/generation (equivalent to 1.42×10^{-9} substitutions/site/year (*54*)). We ran 100 bootstrap replicates for each analysis.

The results of our PSMC analyses of the passenger pigeons whole nuclear genomes suggest that the rate of coalescence varies across the genome in a way that is not predicted by a constant population size evolving under neutrality (figs. S23A,B and S24A,B). If we were to assume neutral evolution, we would interpret these results as evidence that the passenger pigeon N_e fluctuated in size several times between 20 million and 100,000 years ago (figs. S23A,B and S24A,B). However, the results of PSMC analyses can only be interpreted as as reflecting demographic change over time if most sites in the genome have evolved neutrally; otherwise they may return results that reflect the impact of selection in addition to demography (*83*). Since our results suggest that a large proportion of sites across the genome were affected by selection, the results of PSMC analyses cannot be used to reliably infer the size of the passenger pigeon population, or its stability over time.

Our PSMC results for the whole nuclear genomes of the four passenger pigeons consistently differ from previously reported PSMC plots for passenger pigeons (*4*), both in the timing and the scale of population size fluctuations (fig. S25). 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. As mentioned above, in (*4*) the authors use an option that 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, excluding 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 therefore explains the different results observed by Hung *et al. (4)*.

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 due to differences in the rate of recombination. To explore this, we compared the results of PSMC analyses on the high- and low-diversity regions of the passenger pigeon genome separately (where diversity is used as a proxy for recombination rate). 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 might expect that PSMC analyses of low-diversity regions will reflect the impact of selective forces to a greater extent than PSMC analyses of high-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. S23 and S24). These results suggest an historic expansion in N_e for the high-diversity regions (figs. S23C,D and S24C,D), and an historic contraction in N_e for low-diversity regions (figs. S23E,F and S24E,F). In particular, the inferred recent elevation in the N_e of low-diversity regions is consistent with the expected impact of recurrent selective sweeps (i.e. an excess of low-frequency, recent mutations).

In contrast, the results of our PSMC analyses of the homologous regions of one of our band-tailed pigeons (AMNH DOT 14025) are very similar (fig. S26C,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. S26D,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. Differences in the impact of population structure across the genome might drive differences in our PSMC reconstructions following recent admixture (*91*).

11. 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. S6) 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 (*92*): 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 (93)*: 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 (*94*).

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.

12. The influence of gene density and gene proximity on diversity

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 windows with higher genetic diversity, these genes tended to be shorter (fig. S27). 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. S27). 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.

Comparisons of genetic diversity within and between genes

For each 5 Mb window across the genome we estimated π for all sites in the window, for all sites within protein-coding genes, and for only 4-fold degenerate (4fd) sites within genes (at which no change results in an amino acid change). For all these sites we observed the same regional variation in π: higher π at the edges of macrochromosomes and on microchromosomes, and lower π in the centres of macrochromosomes. As expected, we found that the π in protein-coding regions is almost always less than π in non-coding regions (fig. S28A,C). While π is higher at 4fd sites than it is at other sites within genes, it is still generally lower than in non-coding regions (fig. S28B,D). We observe a similar pattern in the band-tailed pigeon genome (fig. S28, blue). Lower diversity at 4fd sites could result from their greater proximity to sites under selection, or from selection for preferred codons.

Testing for a reduction in diversity around genes

If variation in diversity across the genome was the result of selection, we might expect that diversity is both lower within genes, and in regions closer to genes. Nam *et al*. (*95*) show that diversity increases with distance from the nearest gene in great ape species, and that this correlation spans greater physical distances in species with larger effective population sizes (causing a greater overall loss in genome-wide diversity). Nam *et al*. (*95*) find that in orangutans diversity continues to increase up to 1 Mb from a gene. We might expect that the influence of selection was further reaching in passenger pigeons.

In great apes, there are regions of the genome that are 1Mb away from the closest gene (these have been described as 'gene deserts') (*95*). Around 50 Mb of the great ape genome is >823 kb away from a gene (*95*). Considering each scaffold of our band-tailed pigeon genome separately, we find that the maximum distance between genes is 759 kb. This suggests that no region of the band-tailed pigeon genome is >380 kb from a gene. Moreover, <1% of sites are further than 200 kb from a gene, <6% are further than 100 kb from a gene, and >50% are within 15 kb of a gene (fig. S29). Therefore, if the influence of selection on one gene is as far reaching or even further reaching than it is in orangutans, we might expect that diversity at all intergenic sites in the passenger pigeon genome was influenced by selection on more than one gene, and therefore a less strong correlation with distance from closest gene in passenger pigeons than in apes.

Testing for a correlation between π and gene distance

Using the longest scaffolds in our genome assembly, we tested for a correlation between π and distance from the closest gene in both passenger pigeons and band-tailed pigeons. To do this we estimated π for 10 kb windows along the scaffold in both species, and for windows in which there is no annotated gene we calculated the distance from that window to the closest gene. We chose a larger window size than Nam et al. (*95*) because we have a high proportion of sites that cannot be called in all individuals (which are excluded from our estimates of diversity), and because some regions of the passenger pigeon genome have very low diversity.

We found evidence of a weak negative correlation between π and distance from the closest gene in both species (passenger pigeons: Kendall's ρ = -0.020, p = 0.0055; band-tailed pigeons: Kendall's ρ = -0.13, $p < 2x10^{-16}$), and a weak positive correlation between relative π across the two species (π_{PP}/π_{BTP}) and distance from the closest gene (Kendall's ρ = 0.05, $p = 6x10^{-13}$). While a positive correlation between relative π and distance from the closest gene is consistent with gene proximity affecting the impact of selection on genetic diversity, the weakness of this correlation and the weakly negative correlations we obtain when considering the two species individually, suggest that other factors are more important in determining genetic diversity (such as recombination rate). Our failure to detect a strong positive correlation may also in part be a result of the absence of 'gene deserts'.

13. 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 (*99*).

We estimated these statistics for 5 Mb windows across the nuclear genomes of both passenger and band-tailed pigeons. We weighted both statistics by their theoretical minimum values for the number of segregating sites in each window and for the number of individuals in our data sets, in order to facilitate comparisons across genomic regions and across species, following Schaeffer (*96*) and Schmid *et al.* (*97*). We calculated D/D_{min} as described in Tajima (*98*) and Schaeffer (*98*) and H/H_{min} as described in Fay and Wu (*99*) and Schmid *et al.* (*97*). We found that Tajima's D was, on average, negative for both passenger

and band-tailed pigeons. D/D_{min} was more negative in the passenger pigeon than the band-tailed pigeon, and more negative in low-diversity genomic regions than in high-diversity genomic regions. The H-statistic was, on average, positive for the passenger pigeon, but negative for the band-tailed pigeon. Compared to D/D_{min}, H/H_{min} was relatively uniform across high and low diversity regions of the genome, although with greater variation, and a slight decline in low-diversity regions for synonymous sites. These results are consistent with both background selection and selective sweeps having an impact on genetic diversity across the passenger pigeon genome, and a greater impact on the passenger pigeon genome than across the band-tailed pigeon genome.

14. 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 statistic (*100*, *101*). 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 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 analyses to transversion sites that are not affected by cytosine deamination damage, characteristic of ancient DNA (*64*).

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 , was consistent with only a difference of 0.6% in band-tailed pigeon ancestry among passenger pigeons (=0.006) (table S7). In the case of band-tailed introgression into passenger pigeons, we observe D-statistic values much greater than , 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 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.

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.

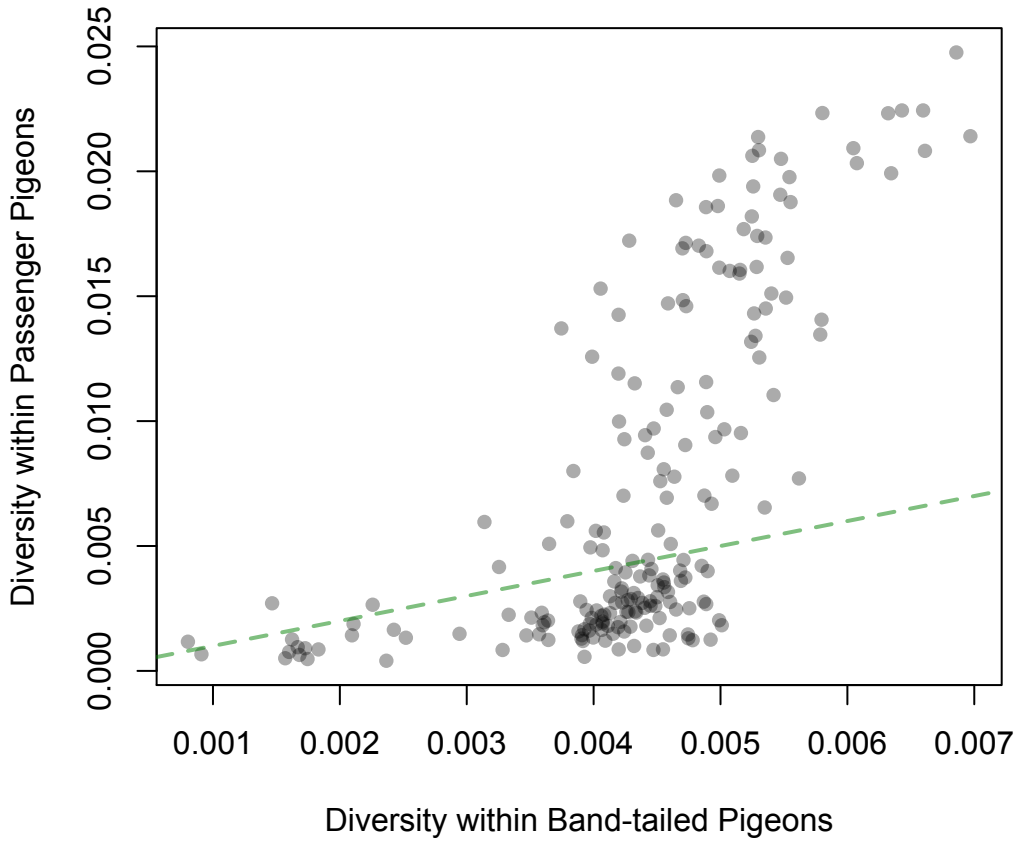


Figure S2. Comparisons of estimates of heterozygosity across the genomes of different passenger pigeons. Estimates of heterozygosity for each passenger pigeon plotted against one another on a $\log_{10}$ scale, estimated for 5 Mb windows across the genome. Diversity is calculated as the mean proportions of sites that are heterozygous. The dashed line represents equality. Outliers are the result of differences in heterozygosity on the z-chromosome between males (BMNH794 & BMNH1149) and females (ROM43.3.23.2 & ROM40360) (fig. S17).

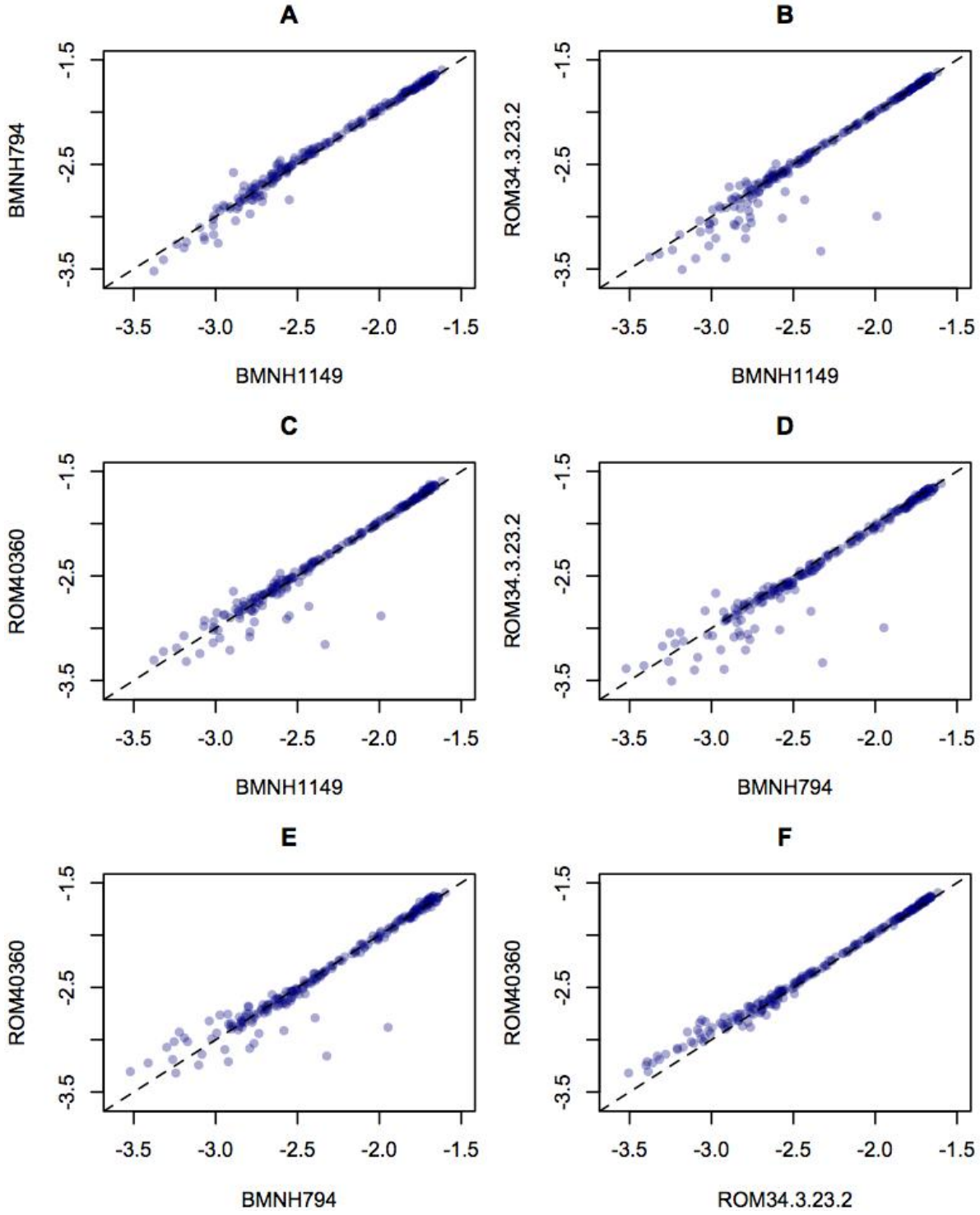


Figure S3. Comparisons of estimates of heterozygosity across the genomes of different band-tailed pigeons. Estimates of heterozygosity for each band-tailed pigeon plotted against one anothermon a $\log_{10}$ scale, estimated for 5 Mb windows across the genome. Diversity is calculated as the mean proportions of sites that are heterozygous. The dashed line represents equality. Outliers are the result of differences in heterozygosity on the z-chromosome between a male (BTP2013) and female (BTP2012) pigeon (fig. S17).

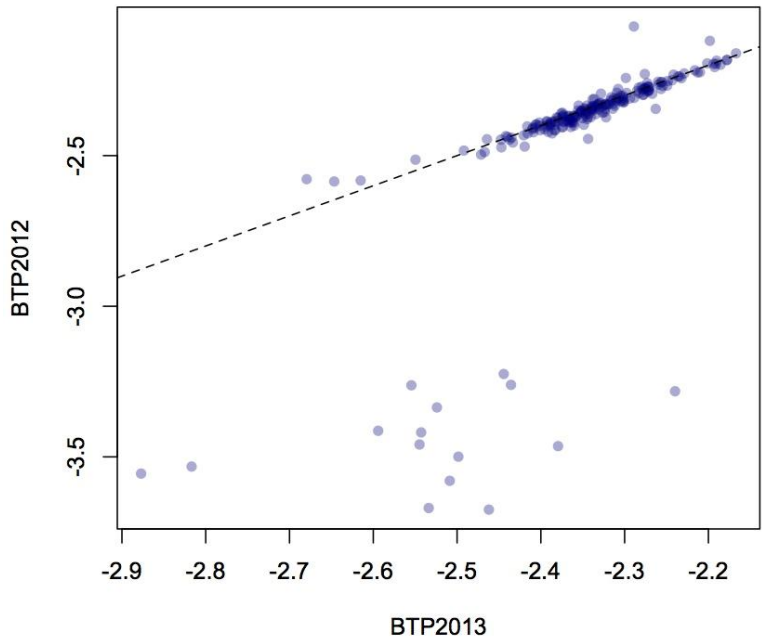


Figure S4. Comparisons of estimates of pairwise nucleotide difference across the haplodized genomes of passenger pigeons and band-tailed pigeons). Minimum (green) and maximum (blue, triangles) estimates of pairwise differences between each pair of passenger pigeon haploid genome sequences (left) and band-tailed pigeon haploid genome seqeunces (right) (excluding within individual comparisons) plotted against the mean of these comparisons, estimated for 5 Mb windows across the genome. The dashed line represents equality.

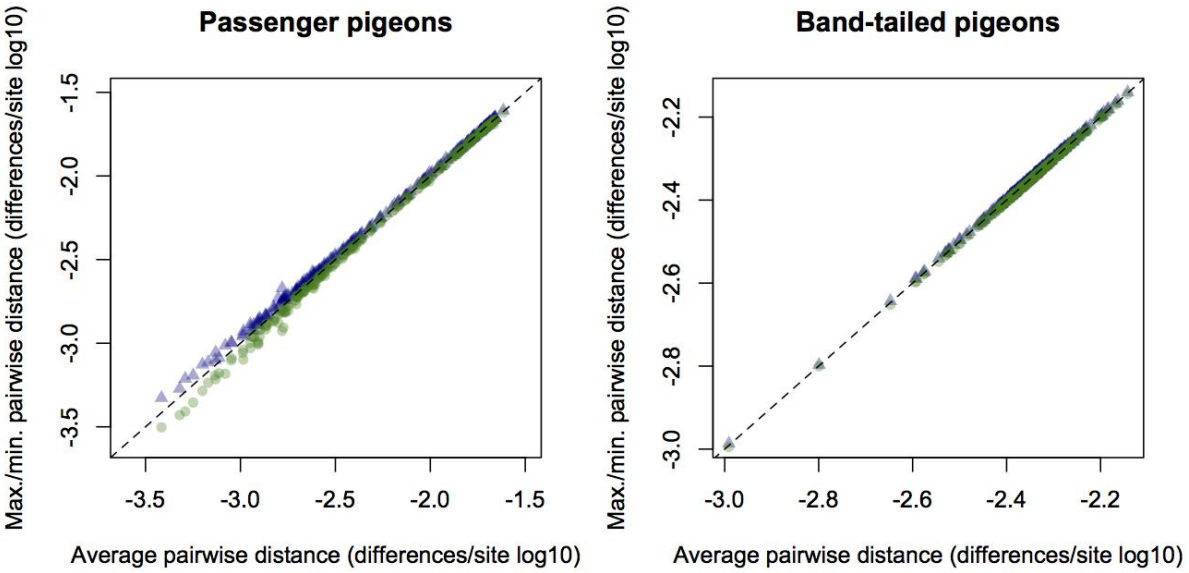


Figure S5. Estimates of diversity across the passenger pigeon and band-tailed pigeon genomes omitting certain types of variant. The proportion of synonymous sites that are bi-allelic polymorphisms, that are (**a**) the result of transitions (dashed lines) or transversions (solid lines), (**b**) at a 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.

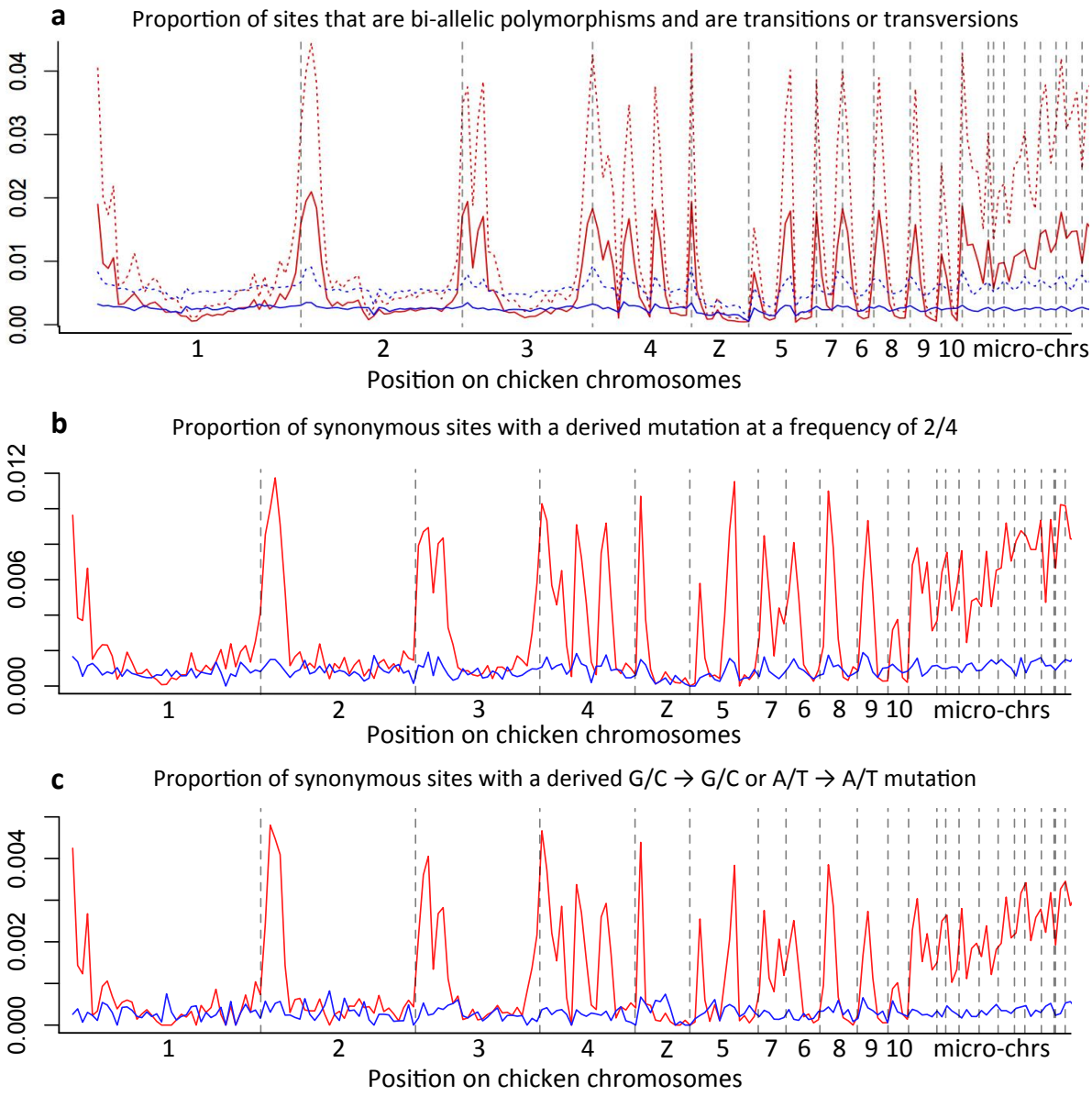
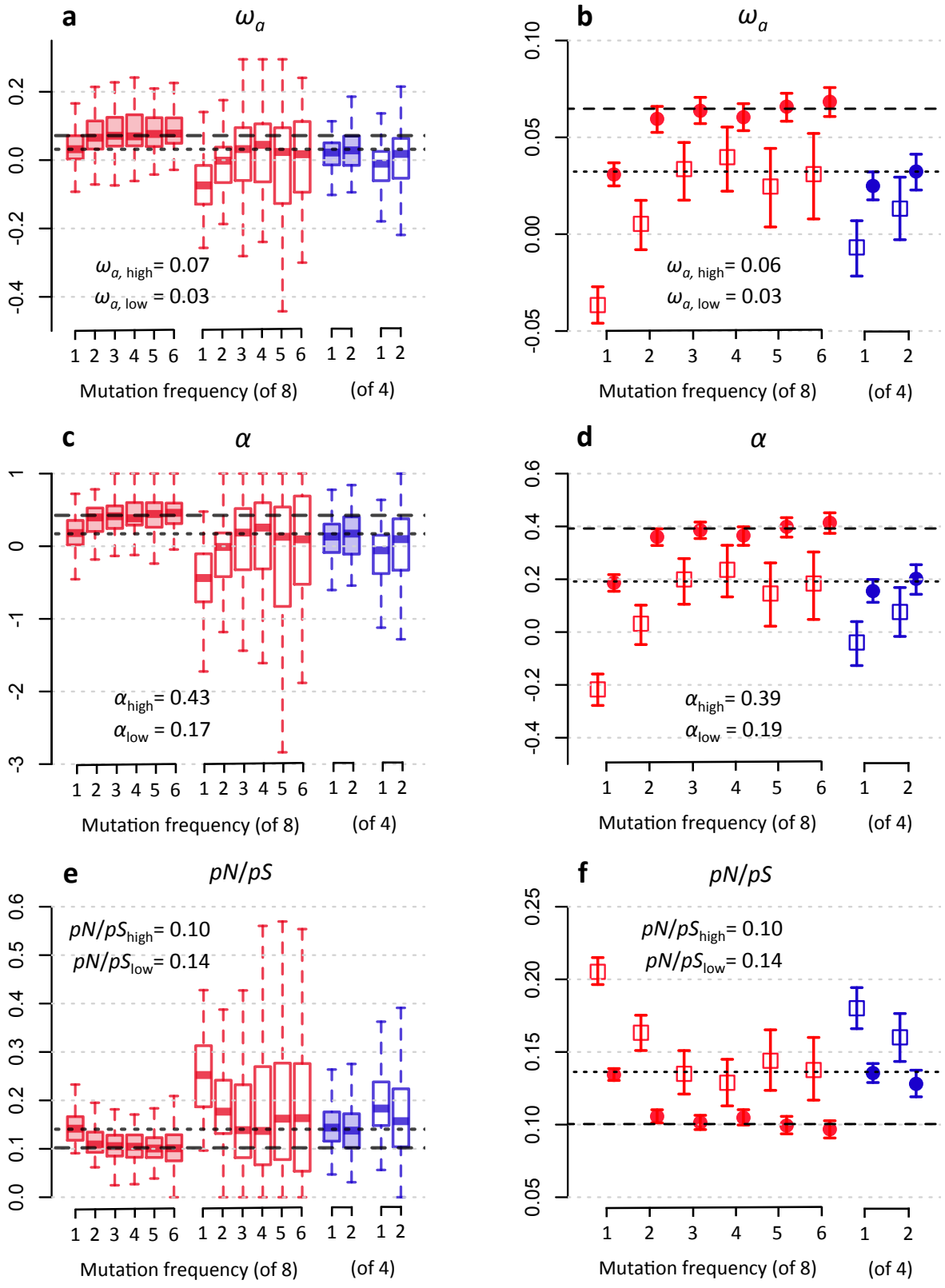


Figure S6. Estimates of ω_a and α, and *pN*/*pS* for different frequencies of derived mutations in passenger pigeons. (a) and (b) show estimates of the rate of adaptive evolution (ω_a). (c) and (d) show estimates of the proportion of amino-acid substitutions that were driven to fixation by positive selection (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 (*102*). 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.


a
ω_a
$\omega_{a,\,high}$= 0.07
$\omega_{a,\,low}$ = 0.03
Mutation frequency (of 8)
(of 4)
b
ω_a
$\omega_{a,\,high}$= 0.06
$\omega_{a,\,low}$ = 0.03
Mutation frequency (of 8)
(of 4)
c
α
α_{high}= 0.43
α_{low} = 0.17
Mutation frequency (of 8)
(of 4)
d
α
α_{high}= 0.39
α_{low} = 0.19
Mutation frequency (of 8)
(of 4)
e
pN/pS
pN/pS_{high}= 0.10
pN/pS_{low}= 0.14
Mutation frequency (of 8)
(of 4)
f
pN/pS
pN/pS_{high}= 0.10
pN/pS_{low}= 0.14
Mutation frequency (of 8)
(of 4)

1012

Figure S7. 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 are included (due to the likelihood of inaccurate estimates in genes with fewer changes). 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.

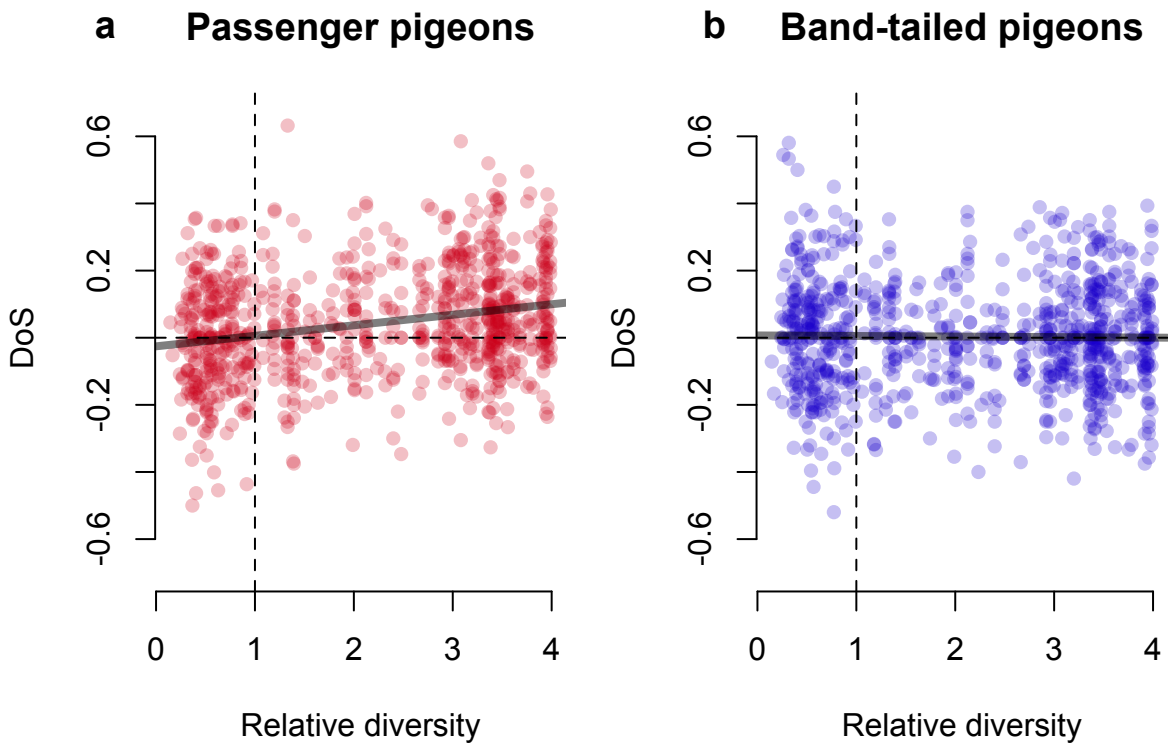


Figure S8. GC-content and neutral substitution biases across the passenger pigeon and band-tailed pigeon genomes. (a) and (b) show the GC-content of the passenger pigeon (a) and band-tailed pigeon (b) genomes against relative diversity across the two species, estimated for 5Mb windows across the genome. (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.

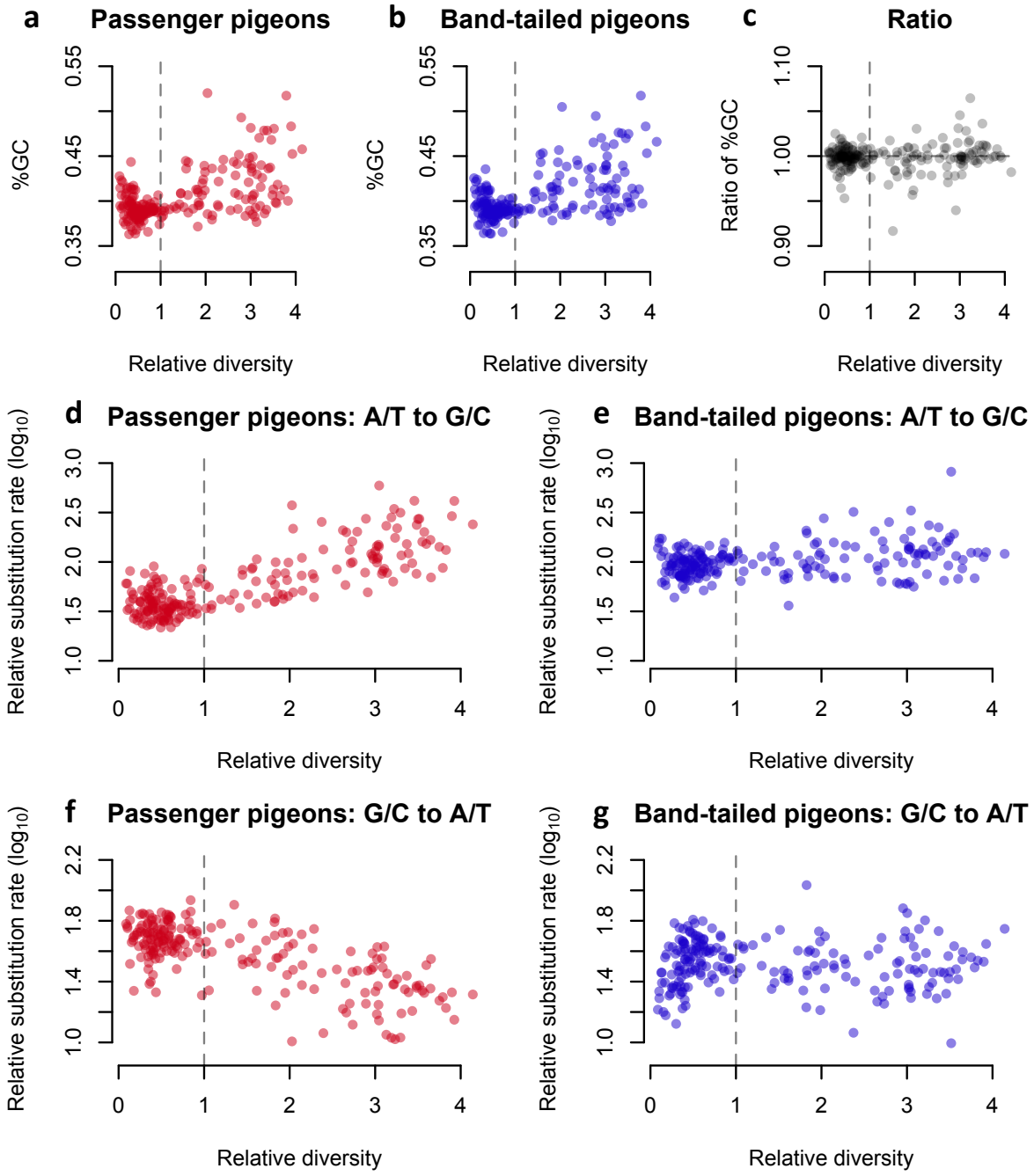


Figure S9. The proportion of substitutions that are nonsynonymous along the passenger pigeon and band-tailed pigeon lineages against relative diversity across their genomes. (a) and (b) show *dN*/ (*dN*+*dS*) for all types of substitution. (c) and (d) show *dN*/ (*dN*+*dS*) for just substitutions that 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.

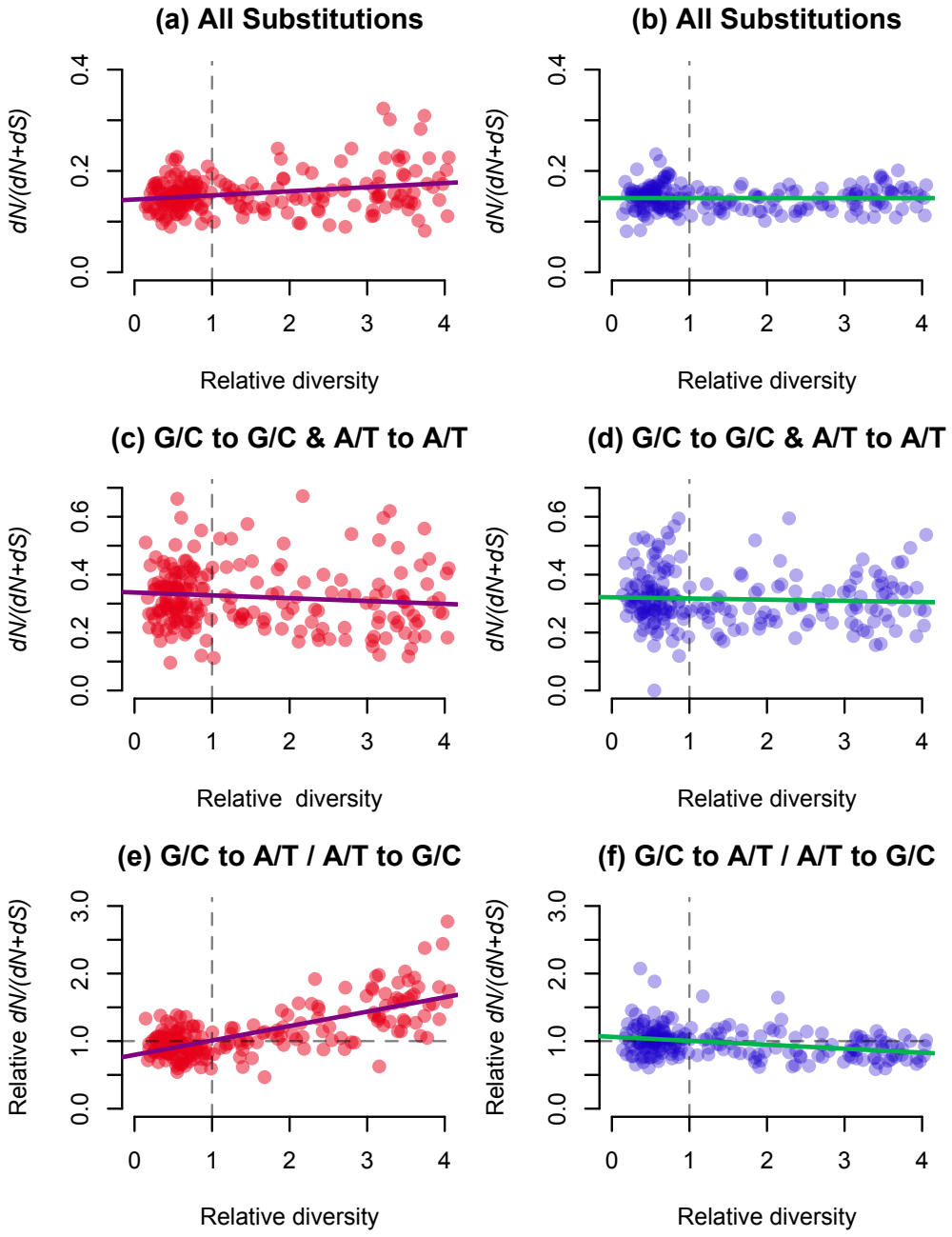


Figure S10. Uncorrected estimates of ω_a, α and *pN*/*pS* for different types of nucleotide base change. These estimates are not corrected for differences in base composition across nonsynonymous and synonymous sites, and across different regions of the genome, and so, with the exception of G/C to G/C 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 S6.

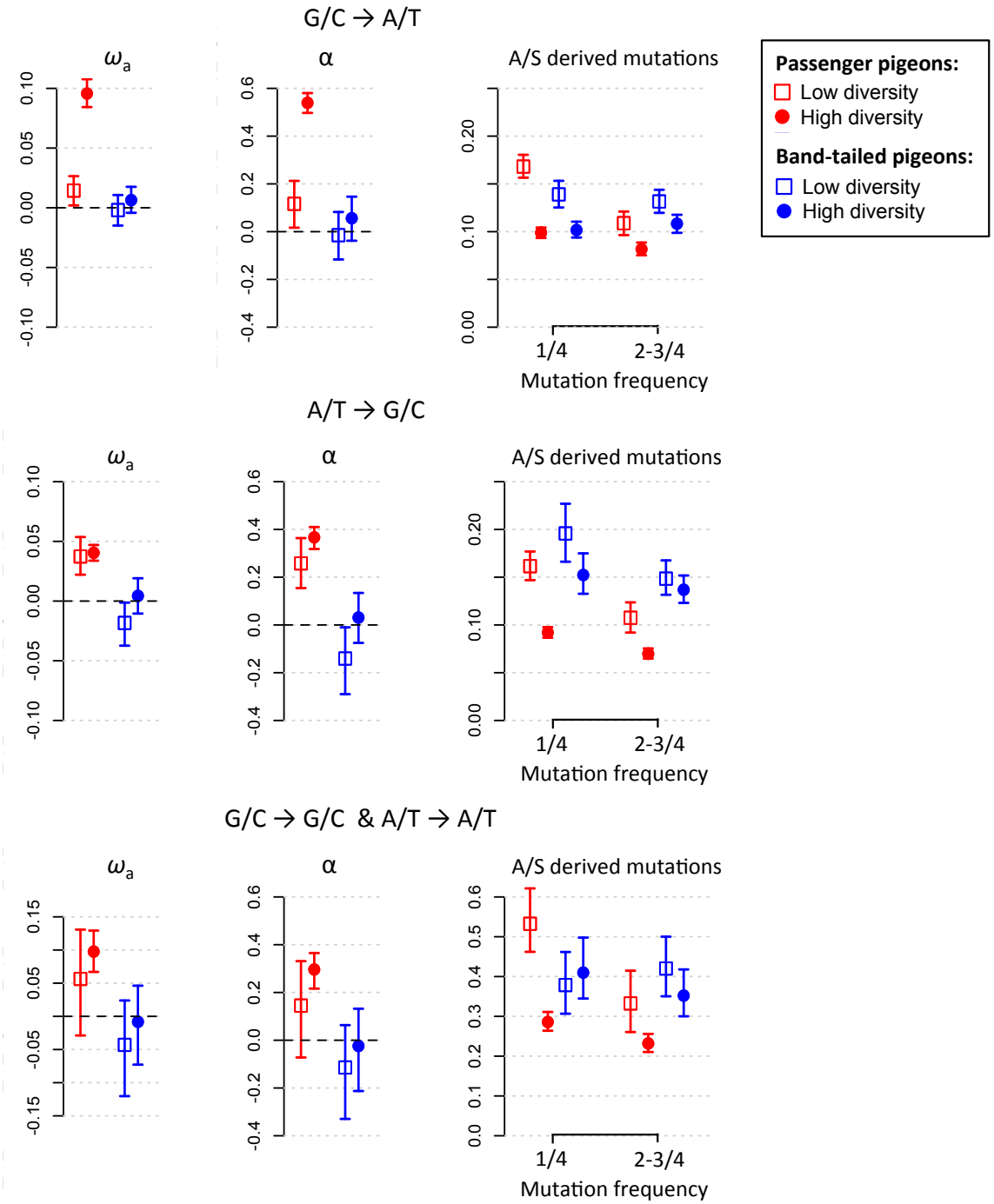


Figure S11. Comparisons of estimates of the ratio of nonsynonymous to synonymous counts of different types of derived nucleotide base change, at different frequencies in our sample, using all 8 passenger pigeon alleles. The estimates are calculated as described in figures S6 and S10.

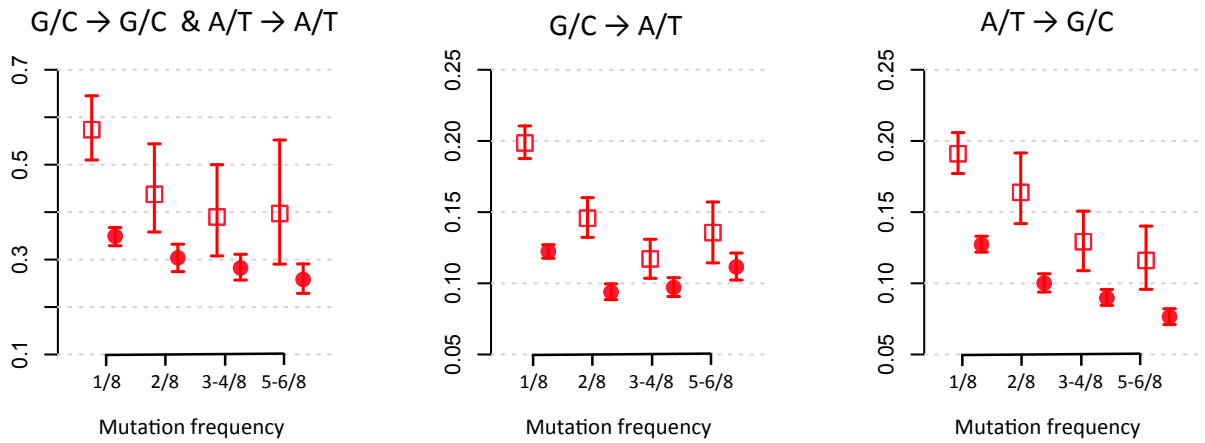


Figure S12. Estimates of ω_a and *pN*/*pS* plotted against estimates of the population-scaled recombination rate in band-tailed pigeons. Estimates are averages for 5 Mb windows and are plotted against the window's population-scaled recombination rate (ρ) in band-tailed pigeons. Comparisons are drawn between (**a**) ω_a and (**b**) *pN*/*pS* in passenger pigeons (PP; red) and band-tailed pigeons (BTP; blue), and between low ρ (ρ < median ρ; point-down triangles) and high ρ (ρ > median ρ; point-up triangles) windows (median values are shown as horizontal lines; '*' indicates $p \leq 1x10^{-3}$ and '-' $p \geq 0.1$ in a Mann-Whitney U test). In (**b**) *pN*/*pS* estimates are for derived mutations present in 1/4 and 2-3/4 individuals. A higher *pN*/*pS* for lower frequency mutations could reflect the slow purging of weakly deleterious mutations. Estimates are based on analyses of two individuals from each species (see figure S3 for estimates using all passenger pigeon samples). Only windows that do not cross scaffold boundaries are included, due to the requirements for estimating ρ.

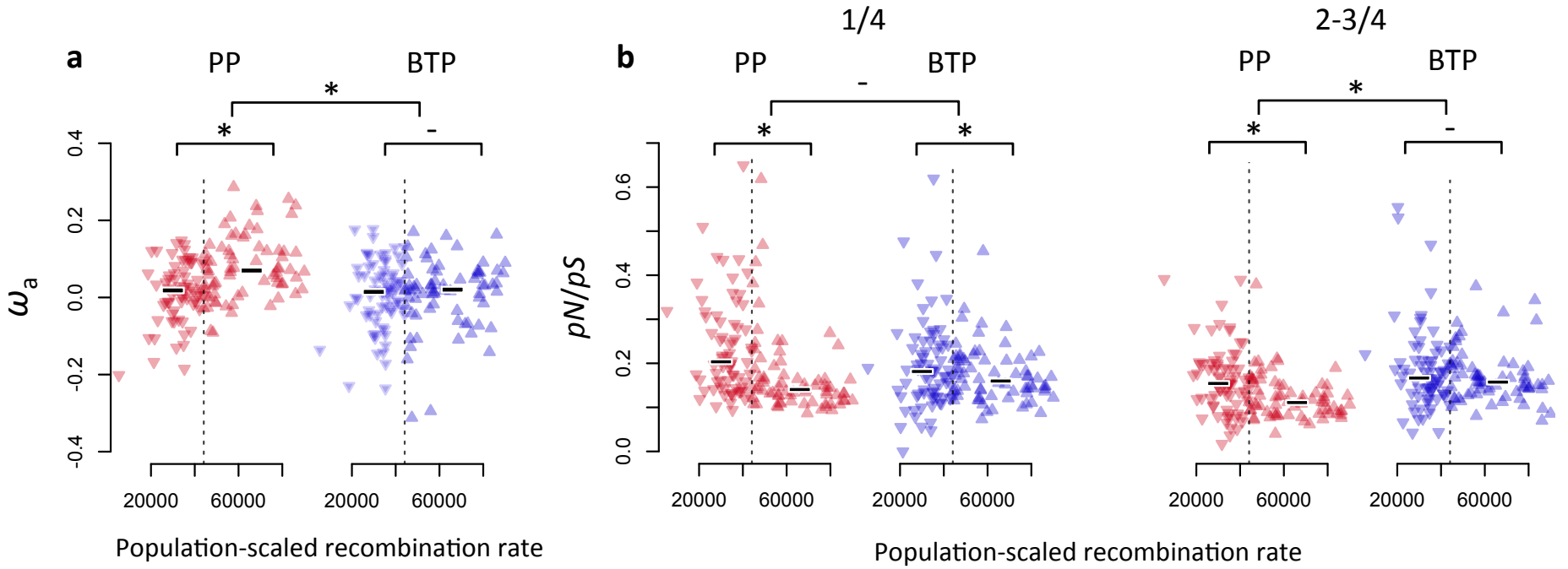


Figure S13. A minimum spanning network of the 41 passenger pigeon mitochondrial genomes. Large circles represent observed haplotypes and small circles represent inferred intermediate haplotypes. Each step between circles represents a single substitution. Colors follow those depicted in the mitochondrial phylogeny in Fig. 1B.

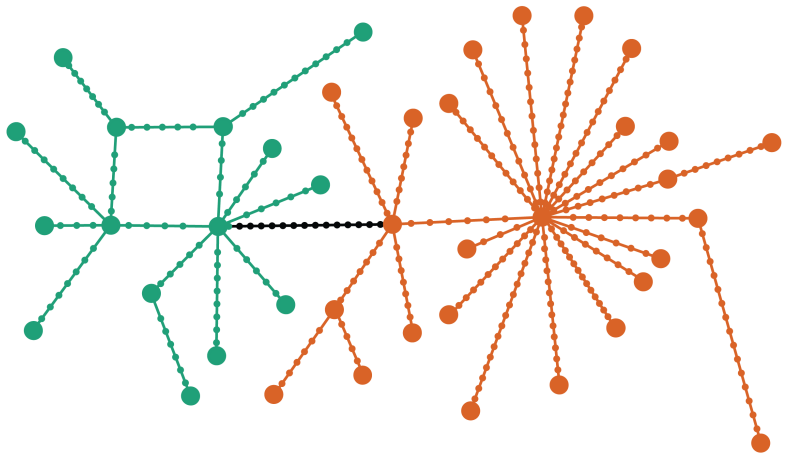


Figure S14. Inferred N_e estimated using using two different calibration rates. Both rates are lineage-specific estimates for passenger pigeons in (*44*): 3.00×10^{-9} substitutions/site/year rate, which was estimated for all sites, for the passenger pigeon lineage, and 1.25×10^{-8} substitutions/site/year 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 (*103*).

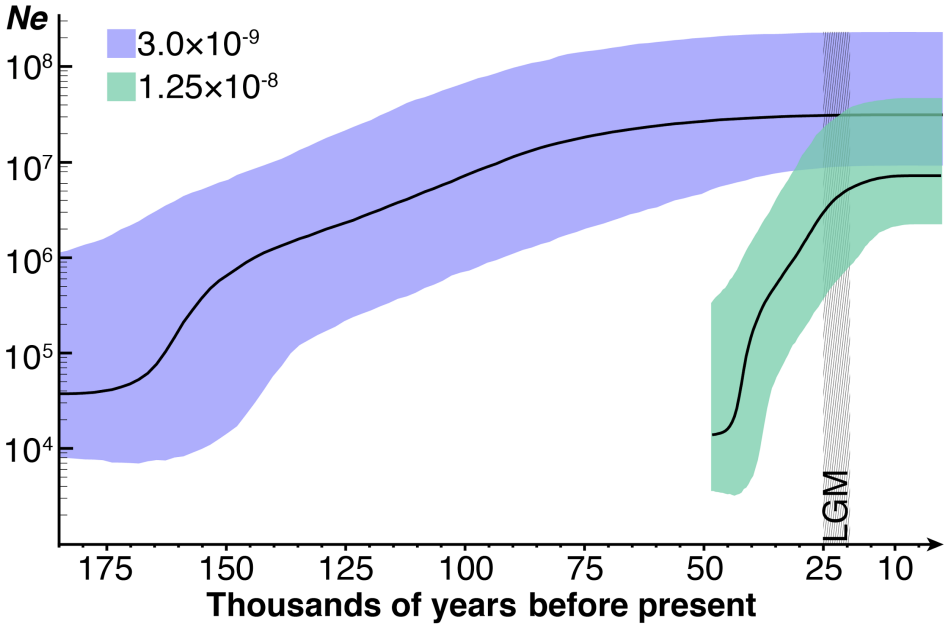


Figure S15. Estimates of nucleotide divergence from the common ancestor of passenger pigeons and band-tailed pigeons for passenger pigeons (A, C) and band-tailed pigeons (B, D), based on divergence between a single individual from each species, plotted against nucleotide diversity within each species. All estimates are for 5 Mb windows across the genome. In A and B, divergence is estimated for all types of nucleotide change, and in C and D, divergence is estimated only for A/T to A/T and G/C to G/C. There is a significant correlation between diversity and divergence in A (Spearman's ρ = 0.30, p = 1.1x10^{-5}), B (ρ = 0.44, p = 12.3x10^{-11}), C (ρ = 0.14, p = 0.043), and D (ρ = 0.17, p = 9.9x10^{-3}). However, variance in divergence is much lower than variance in diversity. Variance in diversity is 4.9x10^{-5} for passenger pigeons, and 1.0x10^{-6} for band-tailed pigeons. Variance in divergence is 7.5x10^{-6} for all sites in passenger pigeons, 2.7x10^{-6} for all sites in band-tailed pigeons, 1.7x10^{-7} for A/T to A/T and G/C to G/C mutations in passenger pigeons, and 6.4x10^{-8} for A/T to A/T and G/C to G/C mutations in band-tailed pigeons.

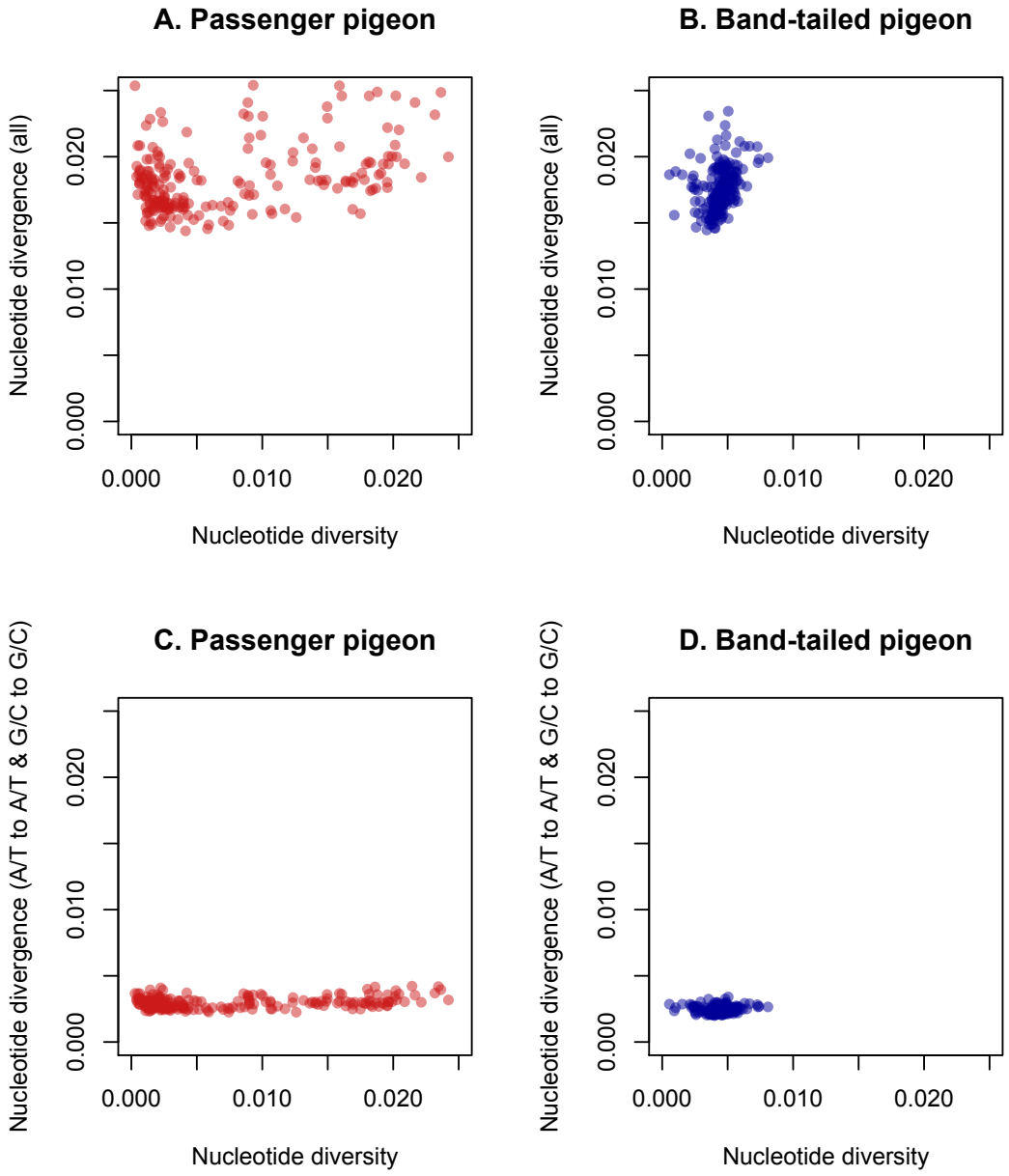
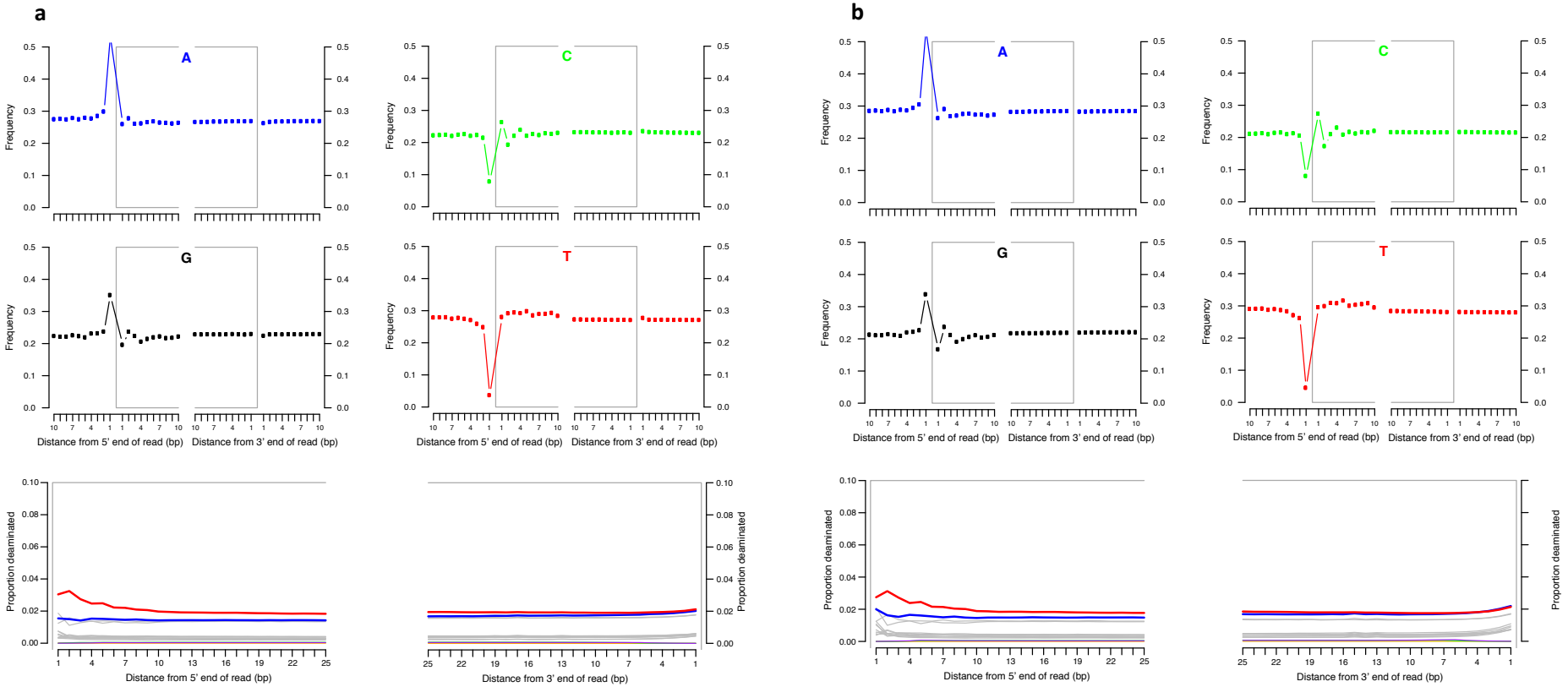


Figure S16. Characterization of damage patterns in genomic DNA from passenger pigeons based on mapping to the band-tailed pigeon genome: (A) BMNH794, (B) BMNH1149, (C) ROM 34.3.23.2, and (D) ROM 40360,. The increased frequency of purines (guanine, G; adenine, A) immediately, upstream of the 5', and a 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).



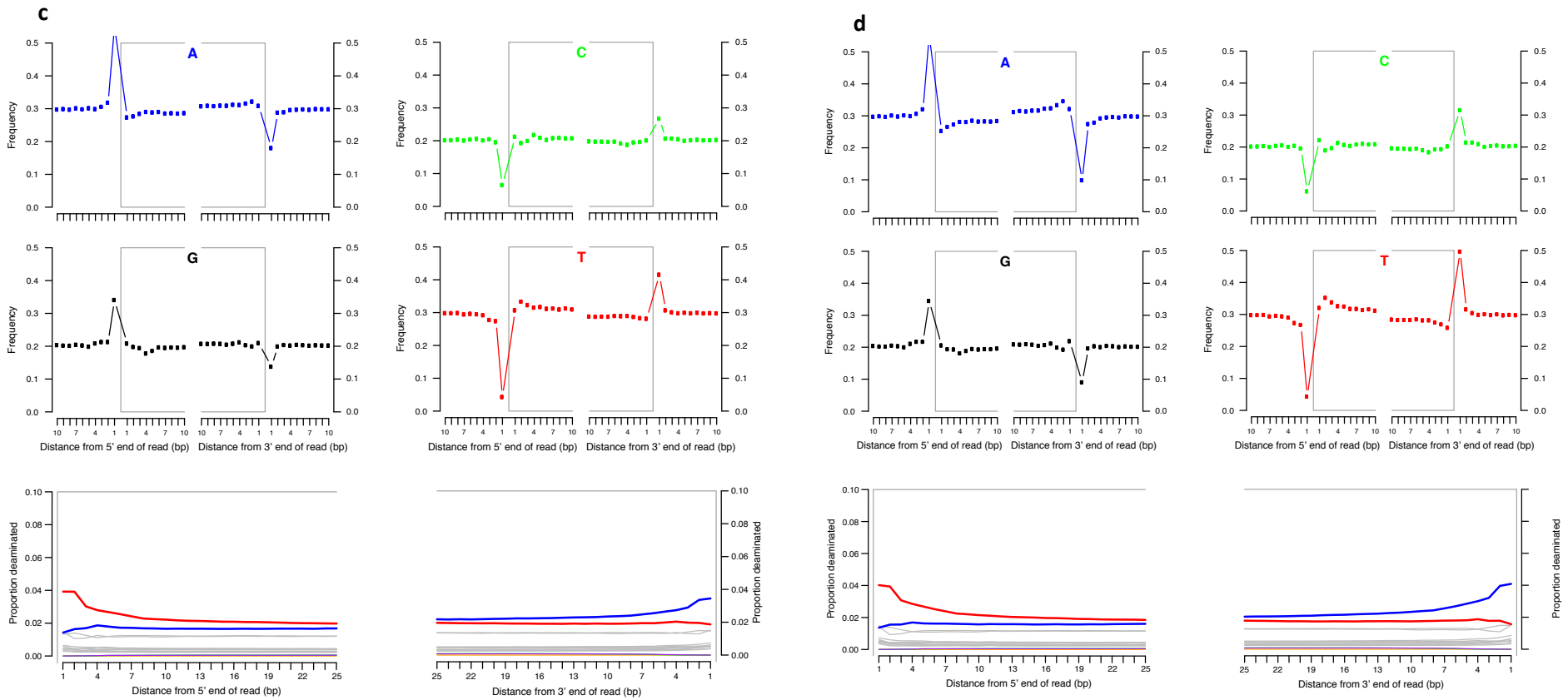

c
d
A
C
G
T
Frequency
Distance from 5' end of read (bp)
Distance from 3' end of read (bp)
Proportion deaminated

Figure S17. 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.

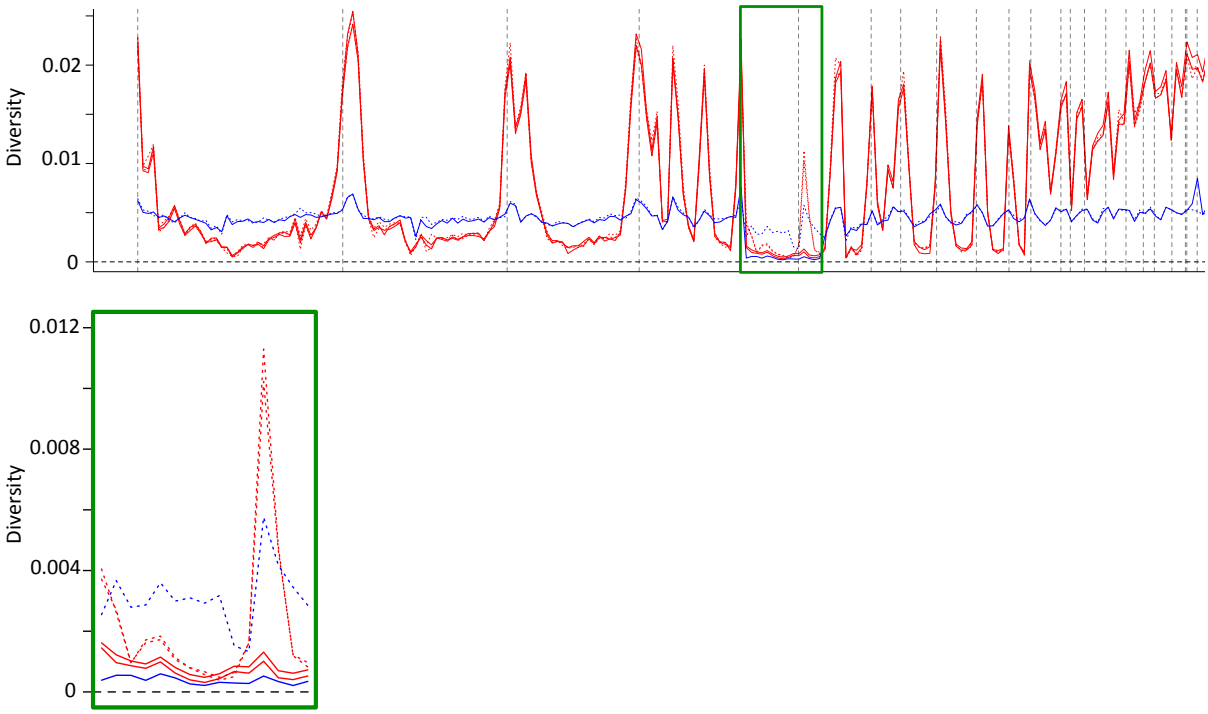


Figure S18. Transition to transversion substitution rates (Ts/Tv): for passenger pigeons (red) and band-tailed pigeons (blue), for 5 Mb windows along our scaffolds, ordered according to our mapping to the chicken genome. Horizontal dashed lines represent chromosome boundaries in the chicken genome.

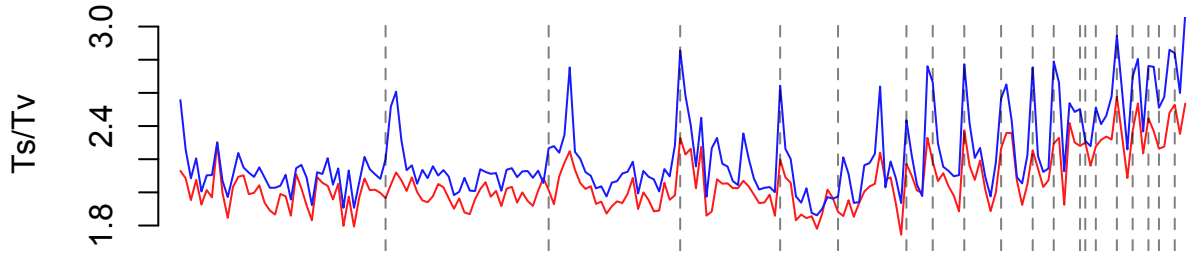


Figure S19. A comparison of estimates of nucleotide diversity against mean map quality for variants that passed our filters for 5Mb windows across the passenger pigeon genome.

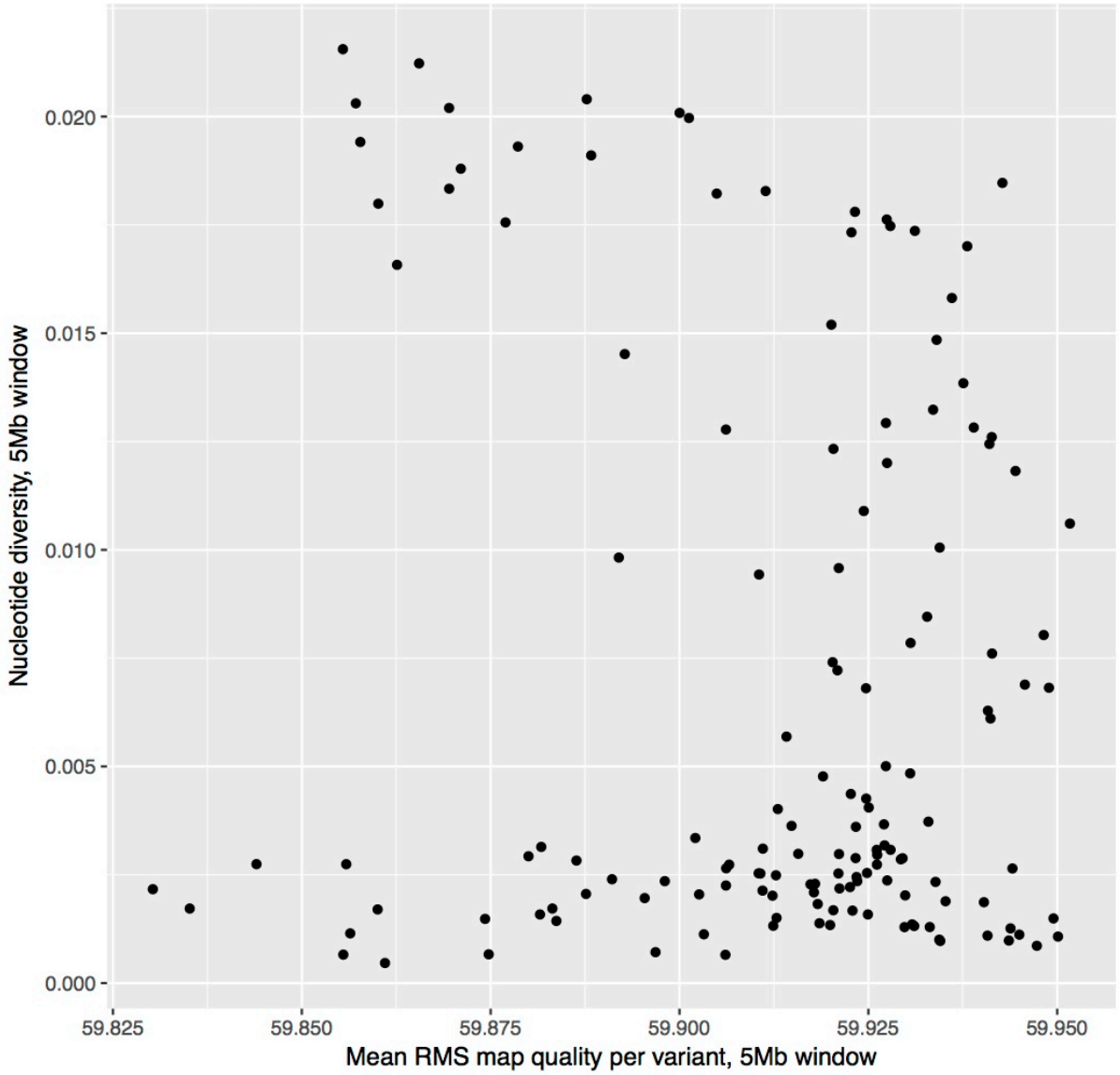


Figure S20. Comparisons of codon usage bias statistics across passenger pigeons and band-tailed pigeons. Histograms of summary statistics of codon usage bias for individual genes in passenger pigeons (red; A, C, E) and band-tailed pigeons (blue; A, C, E), and the difference between these statistics in the two species (B, D, F).

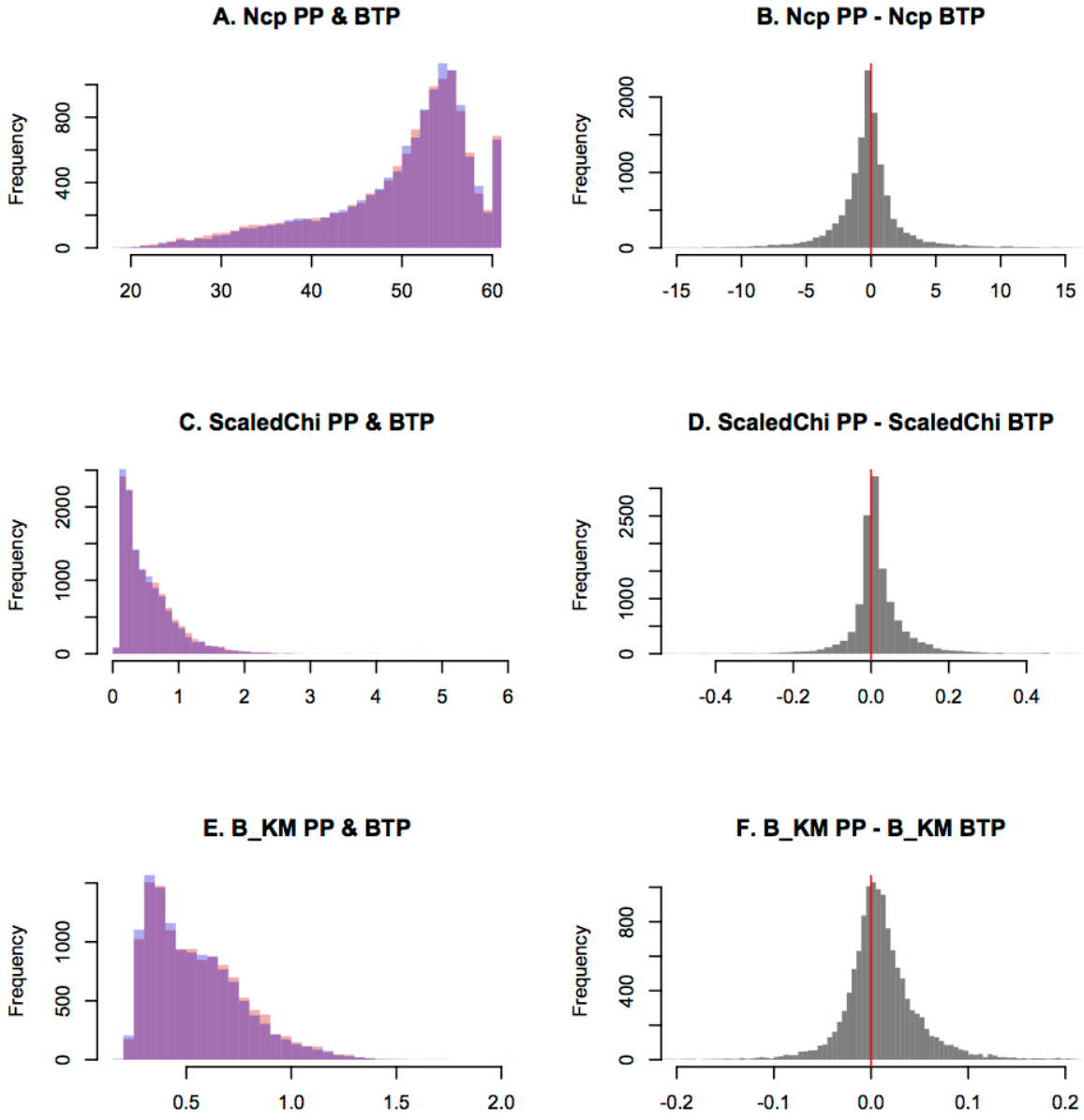


Figure S21. Comparisons of codon usage bias statistics across high- and low-diversity regions of the genome. Histograms of summary statistics of codon usage bias for individual genes in passenger pigeons (A, D, G) and band-tailed pigeons (B, E, H), and the difference between these statistics in the two species (C, F, I), for genes in high-diversity regions (green) and in low-diversity regions (blue) of the passenger pigeon genome.

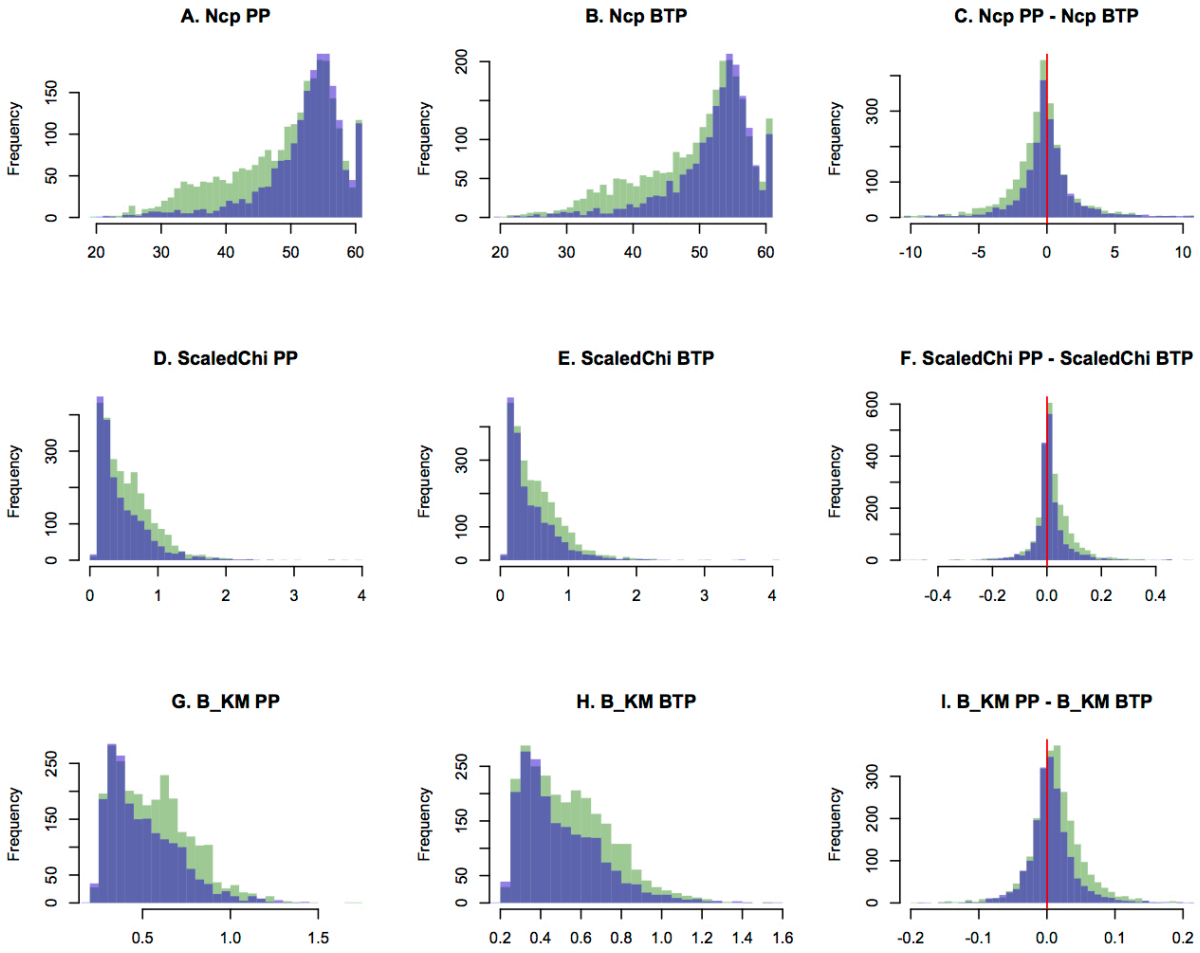
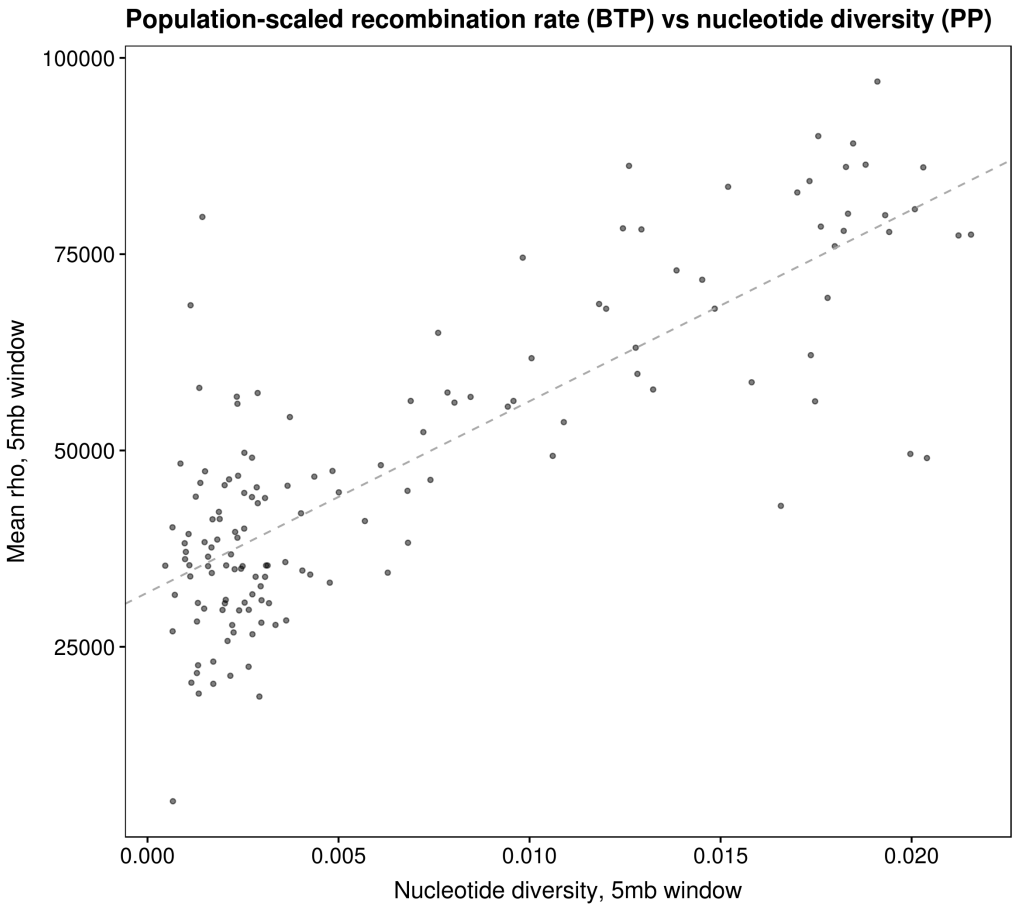


Figure S22. Population-scaled recombination rate (*ρ*) estimates from band-tailed pigeons using LDhat. (a) Mean *ρ* (in band-tailed pigeons) per 5 Mb window plotted against mean nucleotide diversity in passenger pigeons per 5 Mb window. (**b**) The same data shown mapped to the chicken genome, with *ρ* 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)

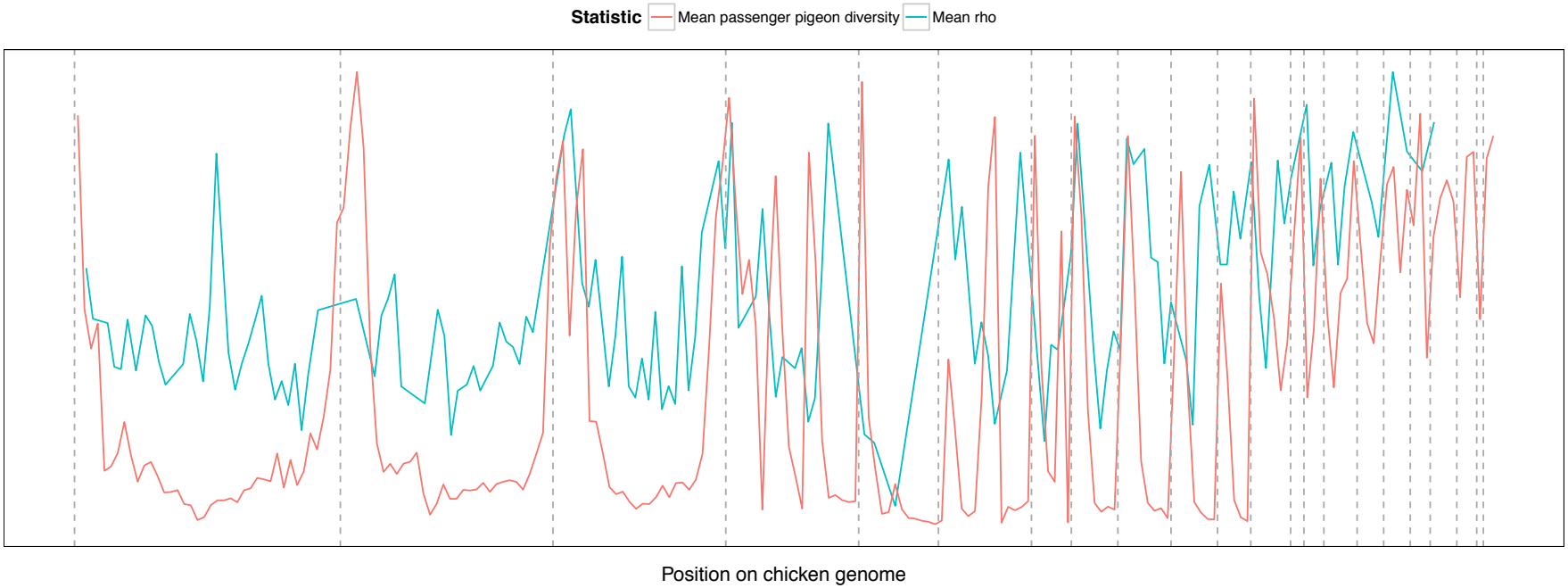


Figure S23. PSMC results for the passenger pigeons ROM 34.3.23.2 and ROM 40360. These plots show N_e over time, with the x-axis indicating years before the present, scaled according to a generation time of 4 years (*90*) and a mutation rate of 5.68×10^{-9} substitutions/site/generation. Every PSMC plot includes 100 bootstrap replicates. A and B) PSMC plots showing the results of N_e 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.

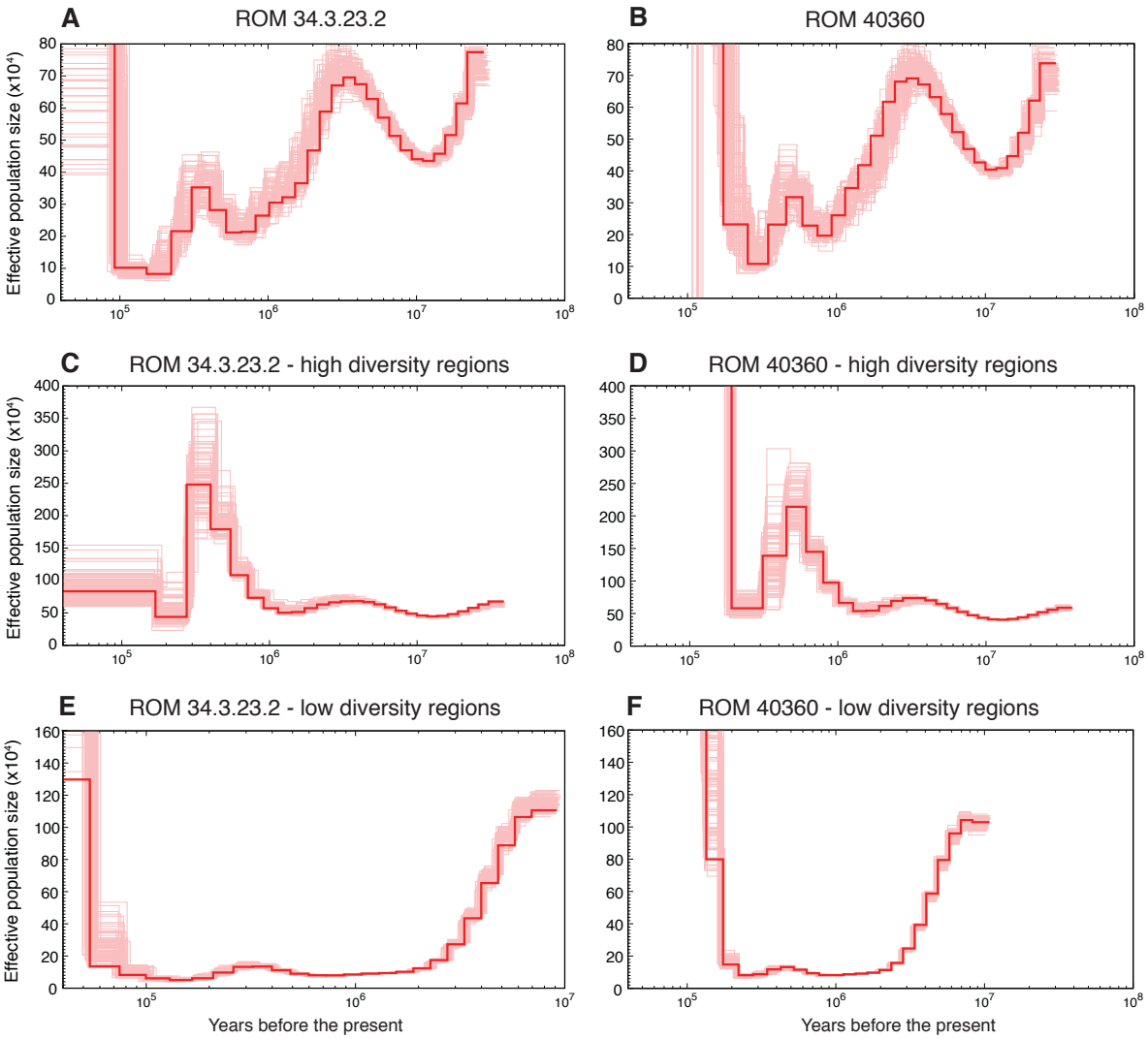


Figure S24. PSMC results for the passenger pigeons BMNH794 and BMNH1149. These plots show N_e over time, with the x-axis indicating years before the present, scaled according to a generation time of 4 years (*90*) and a mutation rate of 5.68×10^{-9} substitutions/site/generation. Every PSMC plot includes 100 bootstrap replicates. A and B) PSMC plots showing the results of N_e 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.

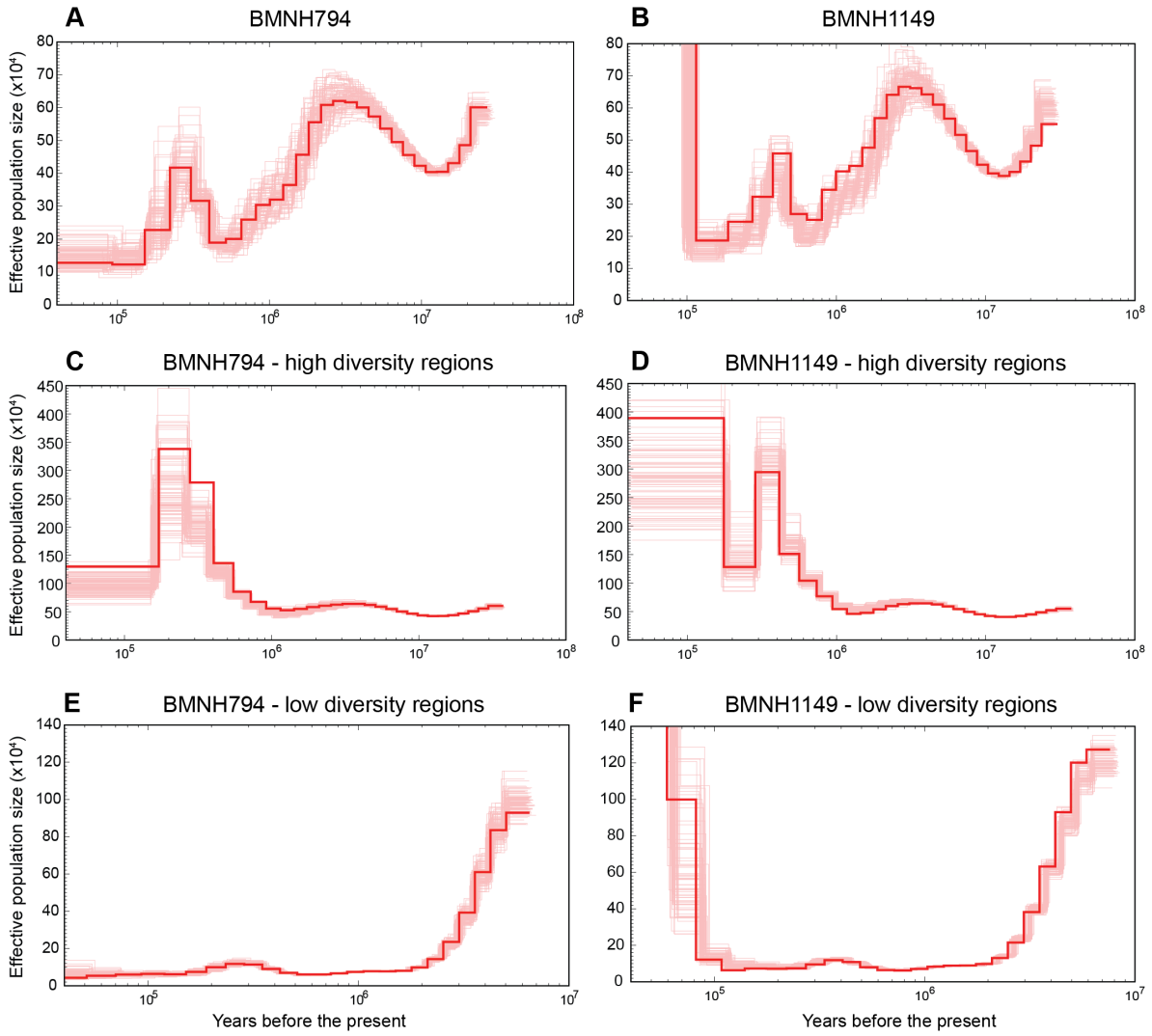


Figure S25. PSMC results for the whole-genome of passenger pigeon ROM 34.3.23.2 using two different parameter choices. The red line represents the result of the PSMC analysis when using the parameters described in Hung *et al. (4)* and the green line represents the result when using parameters that are more appropriate for this data (as discussed above), and as used in the analyses shown in figures S23-S24.

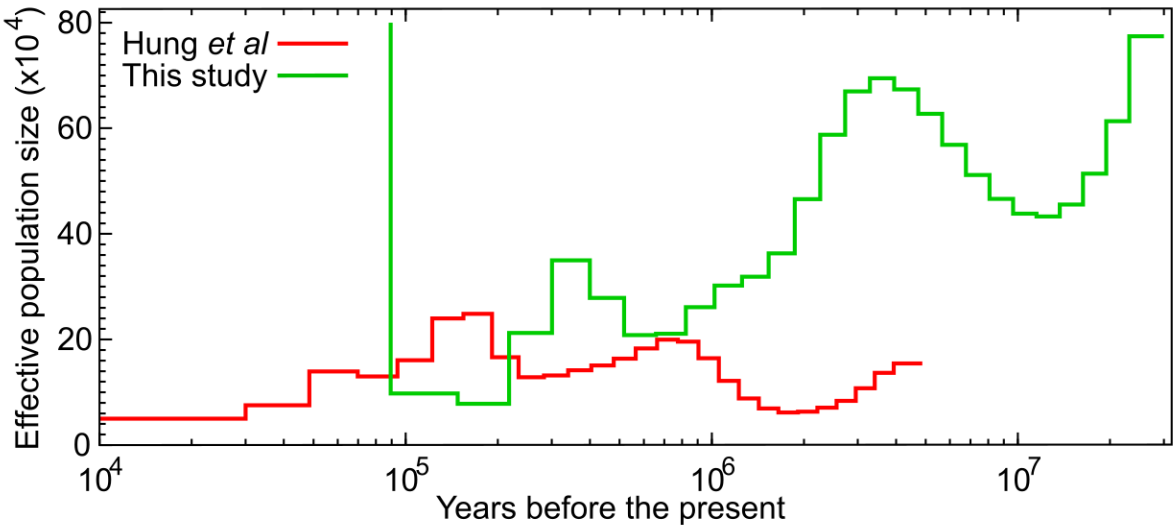


Figure S26. PSMC results for the band-tailed pigeons individuals AMNH DOT 14025 (the reference genome) and BTP2013. These plots show N_e over time, with the x-axis indicating years before the present, scaled according to a generation time of 4 years (*90*) and a mutation rate of 5.68×10^{-9} substitutions/site/generation. Every PSMC plot includes 100 bootstrap replicates. A and B) PSMC plots showing the results of N_e 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.

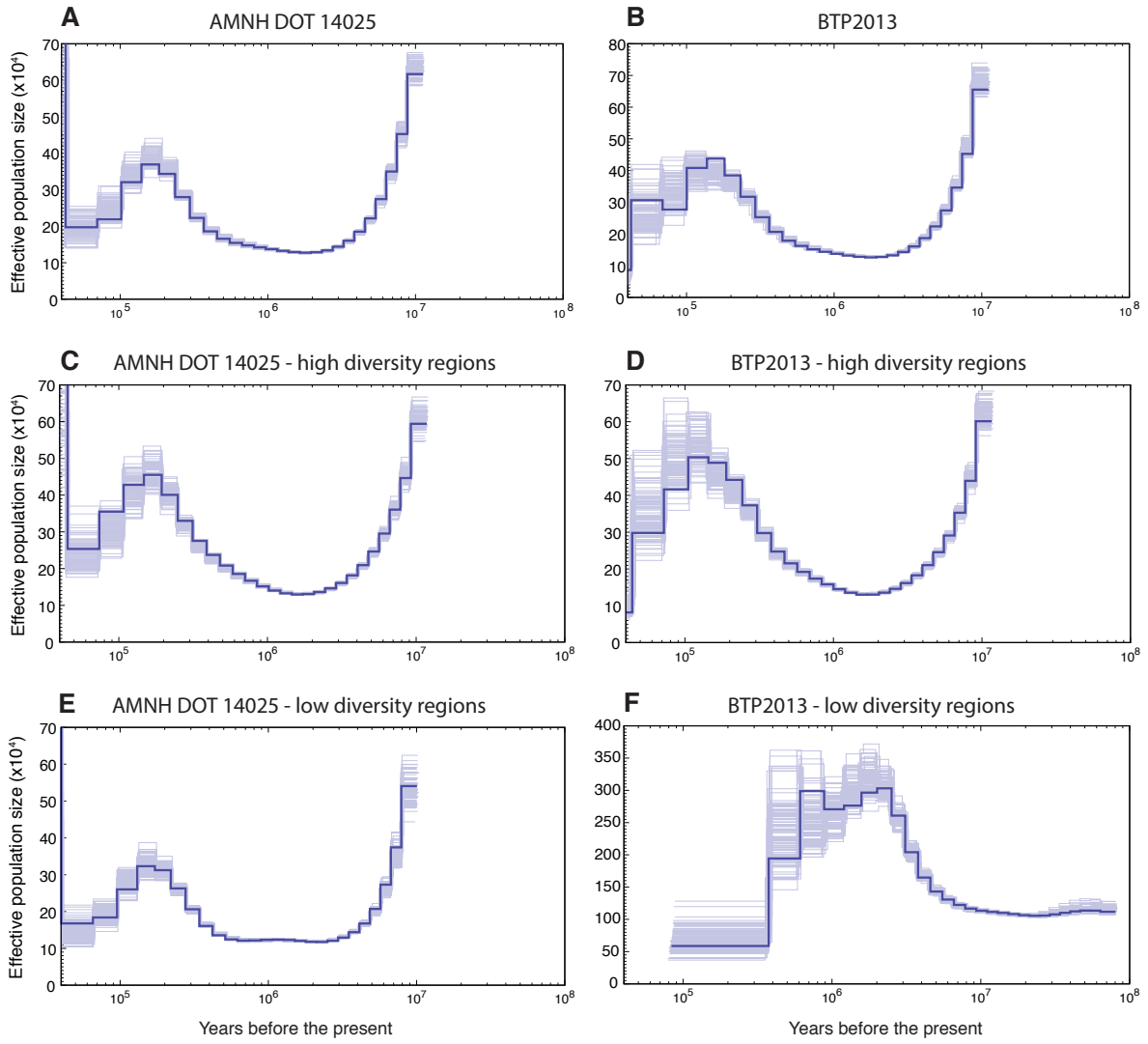


Figure S27. Comparisons of gene count, average gene length, and gene density and genetic diversity for 5 Mb windows across the passenger pigeon genome. Both gene count (Spearman's rank correlation; ρ = 0.51, p = 1.2x10^{-15}) and gene density (ρ = 0.31, p = 4.4x10^{-6}) are positively correlated with genetic diversity in passenger pigeons, whereas gene length is negatively correlated with genetic diversity in passenger pigeons (ρ = -0.47, p < 2.2x10^{-16}).

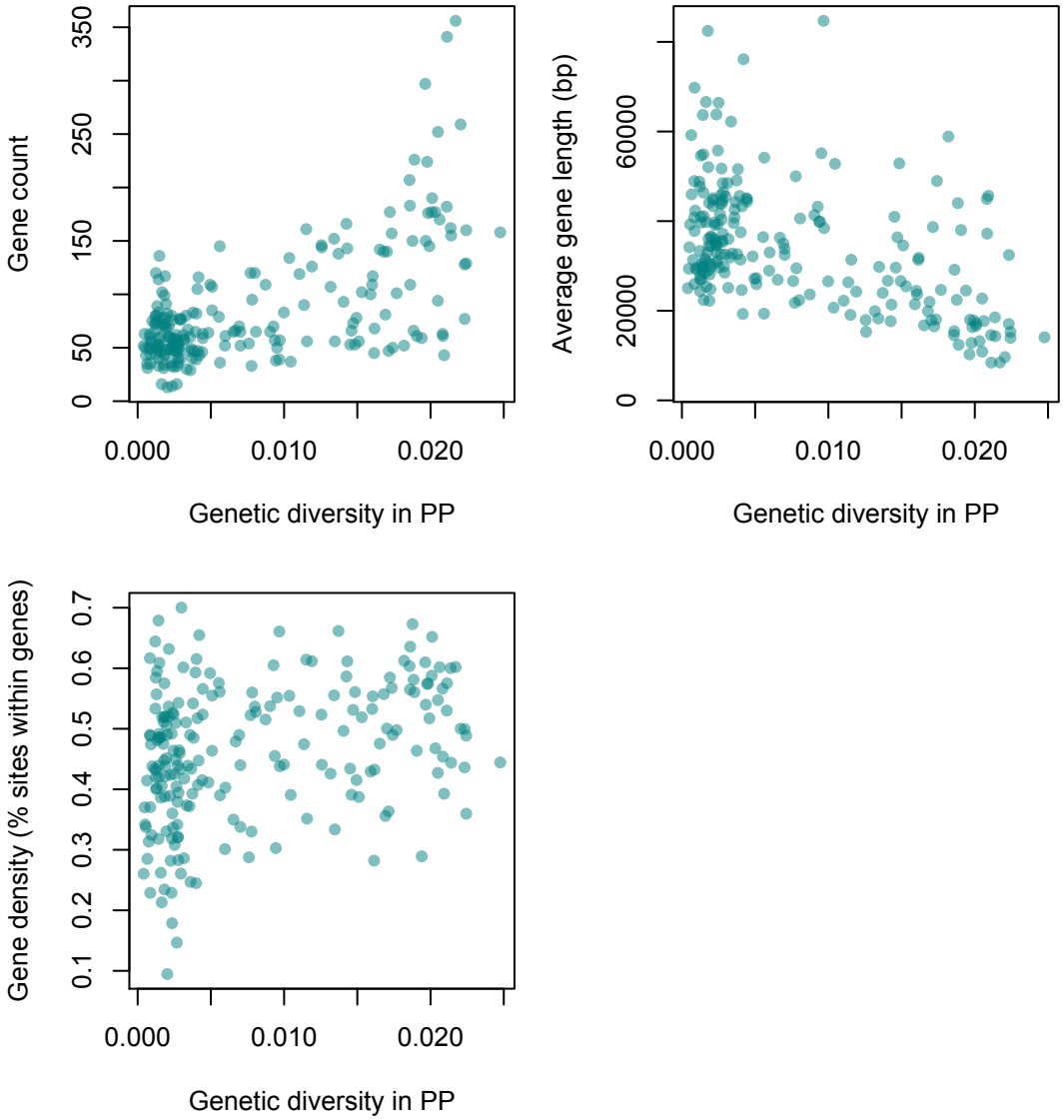


Figure S28. Comparisons of estimates of π at different classes of site. Estimates are based on sites within 5Mb windows across both the passenger pigeon (red) and band-tailed pigeon (blue) genomes. A and C compare π estimated for all sites within genes, to π estimated across all sites within a 5Mb window. B and C compare π estimated for 4fd sites within genes to π estimated for non-coding regions. C and D are on a log-10 scale. Dashed lines represent equality.

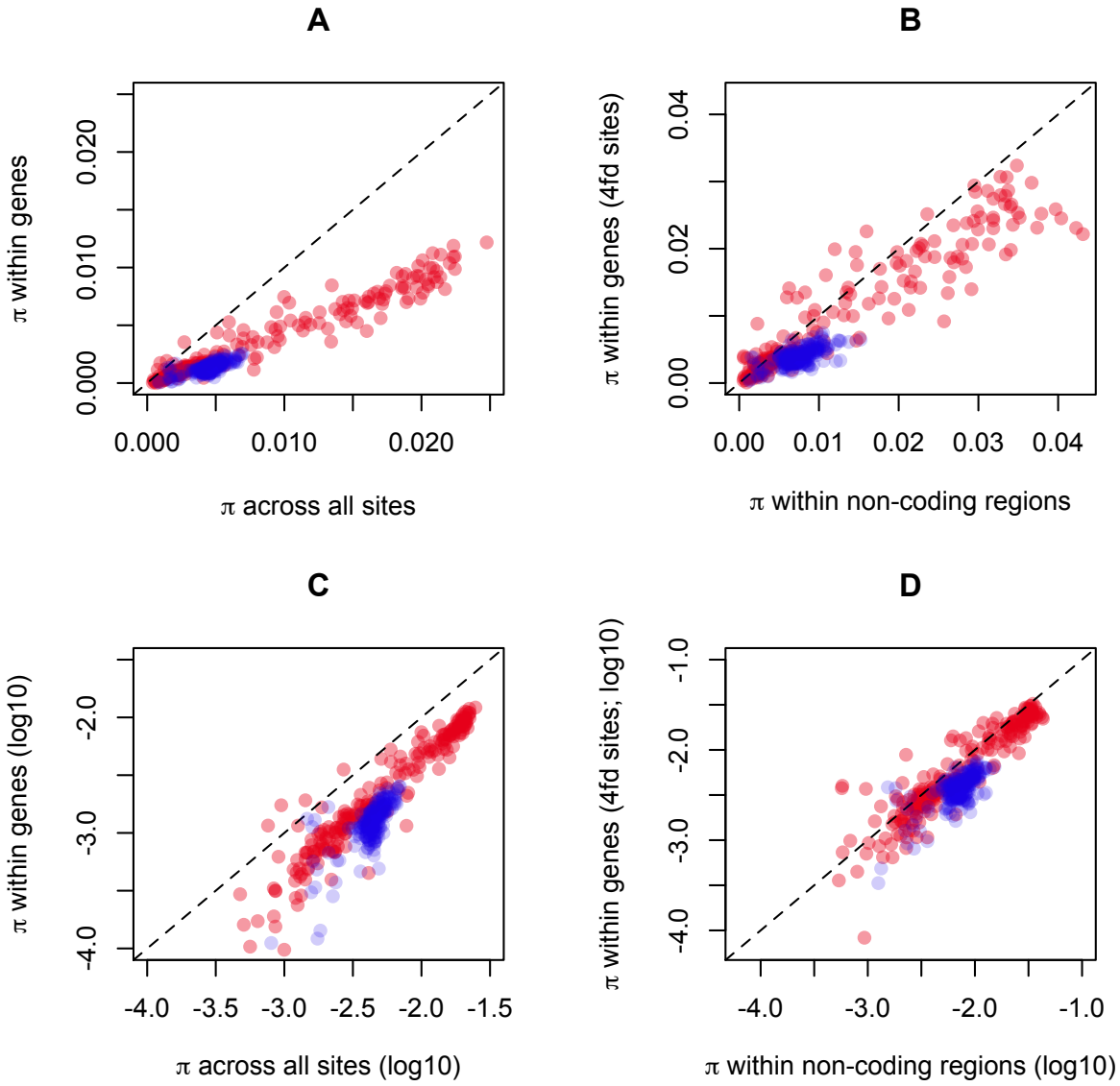


Figure S29. A histogram of the distances between genes in our band-tailed pigeon genome annotation.

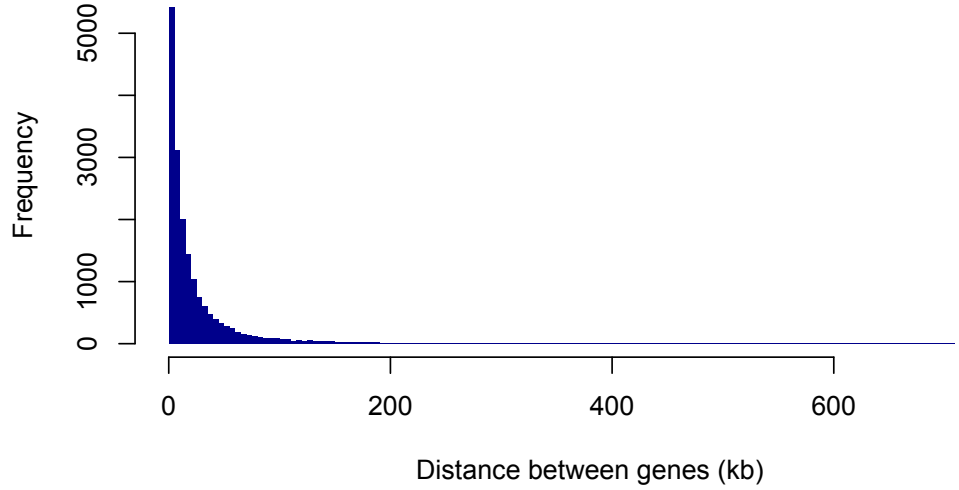


Figure S30. Estimates of Tajima's D/D_{min} and Fay and Wu's H/H_{min}: for synonymous (circles) and nonsynonymous (triangles) mutations plotted against against estimates of relative neutral diversity (diversity in the passenger pigeon divided by diversity in the band-tailed pigeon) for 5 Mb windows across the genome.

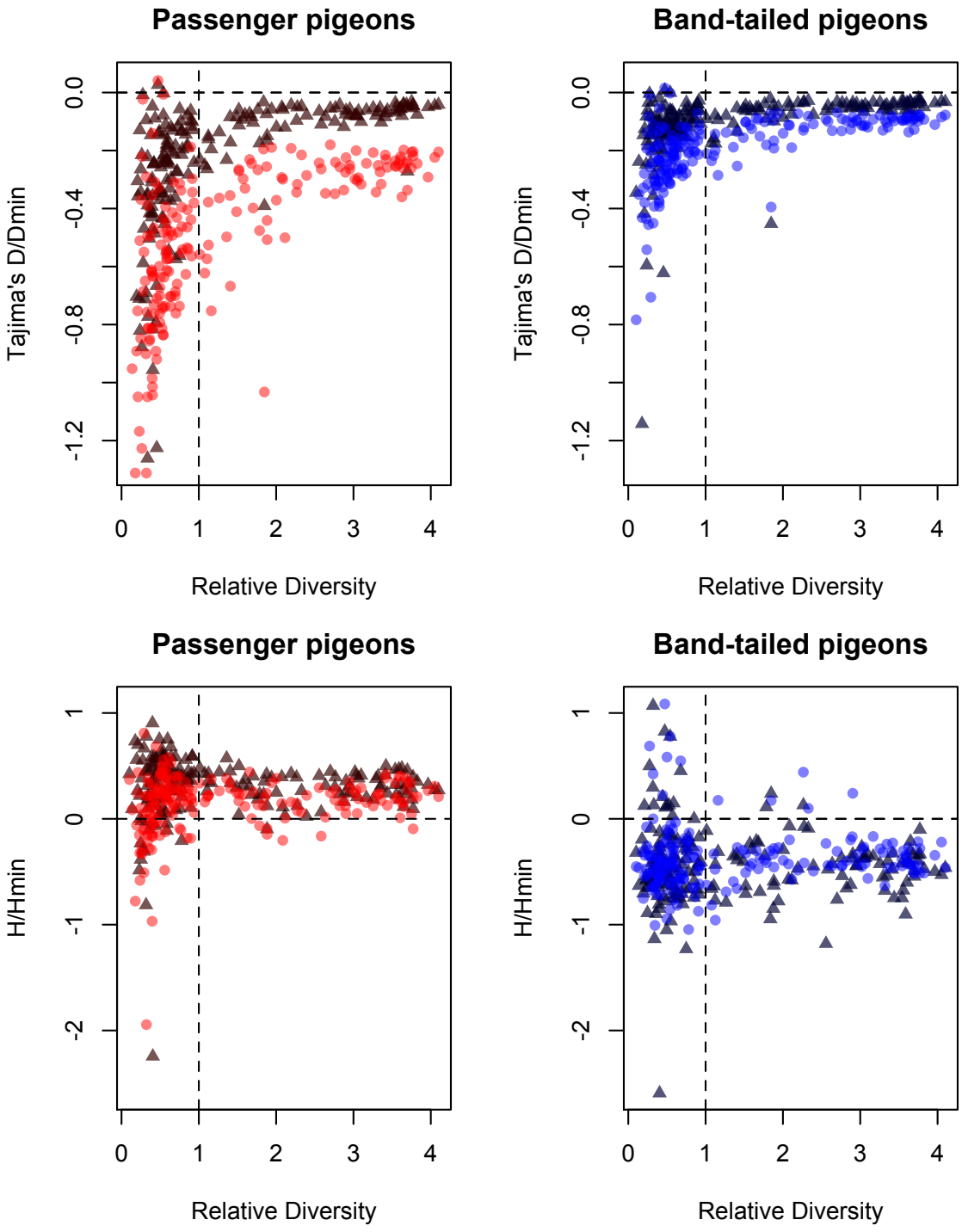
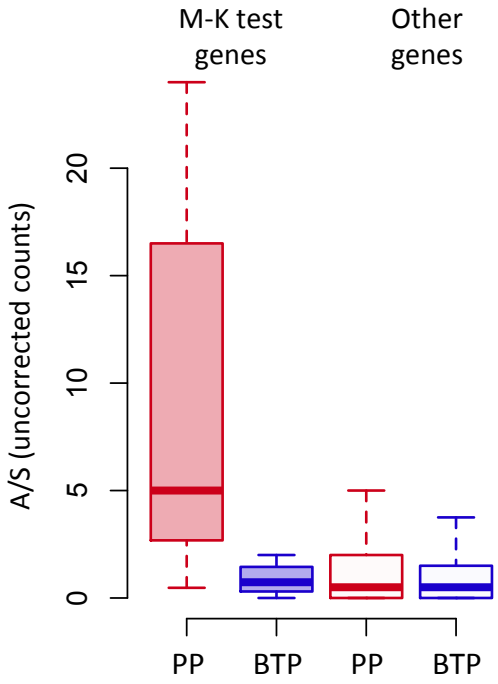


Figure S31. **The ratio of nonsynonymous to synonymous counts of fixed differences for the 32 genes identified as showing evidence of adaptive substitution.** Plots show the ratio of nonsynonymous to synonymous counts of fixed differences (A/S) along the passenger pigeon lineage (PP, red) and band-tailed 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).



Supplementary Tables

Table S1. Sample information. Species name, museum voucher number, sample type, locality, date of collection, extraction and sequencing method, mitochondrial genome average coverage, and GenBank accession number for all specimens used in this study. Provided as a separate excel spreadsheet.

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. Accession numbers and references for previously published data is available on Supplementary Table S1.

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
BMNH794	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 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 (see also fig. S31).

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 (*104*).	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 (*105*).	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 digestion and associated with different diets in birds (*106*).	carbohydrate metabolic process, polysaccharide digestion	8.2×10^{-11}	**Subs**	33	11	17	10
				Poly	14	71	6	9
HELZ2 Helicase with zinc finger domain 2	A nuclear transcriptional coactivator for several nuclear receptors,	DNA replication, RNA catabolic process, mRNA splicing, regulation of transcription	3.5×10^{-8}	**Subs**	26	11	25	21
				Poly	28	107	7	18

	including ones involved in metabolism in the liver and intestine.	from RNA polymerase II, tRNA metabolic process						
ALG13 Putative bifunctional UDP-N-acetylglucosamine transferase and deubiquitinase		dolichol-linked oligosaccharide biosynthetic process	3.5×10^{-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 (*107*, *108*).	fatty acid metabolic process, tRNA aminoacylation for protein translation	1.8×10^{-7}	**Subs**	15	1	2	3
				Poly	3	22	1	3
PRCP Lysosomal Pro-X carboxypeptidase	Associated with blood clotting strength in birds (*109*).	proteolysis	2.0×10^{-7}	**Subs**	9	0	1	4
				Poly	3	34	1	2
DNAH1 Dynein heavy chain 1, axonemal	Force generating protein of respiratory cilia. Involved in sperm motility and implicated in sperm flagellar assembly (*110*, *111*).	cellular component morphogenesis, cellular component movement, chromosome segregation, fertilization, intracellular protein transport, mitosis, spermatogenesis, vesicle-mediated transport	4.9×10^{-7}	**Subs**	17	4	13	16
				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	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

	system associated with the myofibrils (*112*).							
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	$7.6x10^{-7}$	**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 (*92*).	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 (*113*).	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. Expression associated with feed restriction in chickens (*114*). It enables the transport of fat molecules in blood plasma and lymph and facilitates the movement of molecules such as cholesterol into cells (*115*). Has been identified as a target of selection in polar bears (*116*).	Cholesterol transporter activity, lipid binding	4.0×10^{-6}	**Subs**	24	12	24	30
				Poly	48	139	5	13
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 (*117*). Also implicated in sexual differentiation in chickens (*118*).	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
HNF4A	Transcriptionally	cell cycle, lipid	9.5×10^{-6}	**Subs**	10	0	3	5

Hepatocyte nuclear factor 4-alpha	controlled transcription factor. May be essential for development of the liver, kidney and intestine. Associated with type 2 diabetes in humans (*119*).	metabolic process, sensory perception, visual perception, vitamin metabolic process, sex differentiation		**Poly**	6	24	2	6
NUP160 Nuclear pore complex protein Nup160	Involved in mRNA transport (*120*). Associated with reproductive isolation in *Drosophila*, where there is evidence of positive selection (*121*).	cellular process, nuclear transport	$1.1x10^{-5}$	**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 (*122*, *123*).	Nervous system development	$1.4x10^{-5}$	**Subs**	29	9	22	17
				Poly	93	151	12	10
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	angiogenesis, blood coagulation, ectoderm development localization, endocytosis, heart development, intracellular protein transport, nervous system development, proteolysis, sensory	$1.4x10^{-5}$	**Subs**	24	1	7	7
				Poly	18	22	6	6

	humans (*124*, *125*).	perception, skeletal system development, synaptic transmission, visual perception, vitamin transport						
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 ATP-binding cassette sub-family A member 5	May play a role in lysomoal trafficking and is thought to be expressed in skeletal muscle, kidney, liver and placenta (*126*).	catabolic process, cellular process, lipid transport, nitrogen compound metabolic process, nucleobase-containing compound metabolic process, phosphate-containing compound metabolic process, cholesterol efflux, reverse cholesterol transport	$1.5x10^{-5}$	**Subs**	10	16	6	12
				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	extracellular transport, immune system process, response to toxic substance	$1.7x10^{-5}$	**Subs**	13	19	5	11
				Poly	5	81	2	5

	for the export of bile acids and glucuronides from cholestatic hepatocytes (By similarity). Has been associated with response to avian influenza infection in humans (*127*)							
COL22A1 Collagen alpha-1 (XXII) chain	A structural protein. Associated with serum creatinine level (*128*), 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 (*129*).		$2.7x10^{-5}$	**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	Vitronectin is a cell adhesion and spreading factor	cell-matrix adhesion, extracellular matrix	$3.8x10^{-5}$	**Subs**	7	3	1	3
				Poly	4	53	1	3

	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.	organization, immune response, liver regeneration, oligodendrocyte differentiation, positive regulation of cell-substrate adhesion, protein polymerization						
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 (*130*).	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 closest match in other annotations is identical to the above gene)*			$4.1x10^{-5}$	**Subs**	28	4	12	3
				Poly	37	45	9	9
CFAP57 Cilia- and flagella-associated protein 57			$8.0x10^{-5}$	**Subs**	14	4	8	5
				Poly	24	65	9	4
NCAPH	Regulatory		$8.2x10^{-5}$	**Subs**	13	2	7	6

Condensin complex subunit 2	subunit of the condensin complex, a complex required for conversion of interphase chromatin into mitotic-like condensed chromosomes.			**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