MOLECULAR ECOLOGY

Molecular Ecology (2012) doi: 10.1111/j.1365-294X.2012.05711.x

A parapatric propensity for breeding precludes the completion of speciation in common teal (*Anas crecca*, sensu lato)

JEFFREY L. PETERS,*[1] KEVIN G. McCRACKEN,†‡ CHRISTIN L. PRUETT,§ SIEVERT ROHWER,¶ SERGEI V. DROVETSKI,** YURIY N. ZHURAVLEV,†† IRINA KULIKOVA,†† DANIEL D. GIBSON‡ and KEVIN WINKER‡‡‡‡

*Department of Biological Sciences, Wright State University, 3640 Colonel Glenn Hwy, Dayton, OH 45435, USA, †Institute of Arctic Biology, University of Alaska, Fairbanks, 902 N. Koyukuk Drive, Fairbanks, AK 99775, USA, ‡University of Alaska Museum, 907 Yukon Drive, Fairbanks, AK 99775, USA, §Department of Biological Sciences, Florida Institute of Technology, 150 W University Blvd, Melbourne, FL 32901, USA, ¶Department of Biology and Burke Museum, University of Washington, Seattle, WA 98195, USA, **Tromsø University Museum, NO-9037 Tromsø, Norway, ††Institute of Biology and Soil Science, Far East Branch, Russian Academy of Sciences, Vladivostok 690022, Russia

Abstract

Speciation is a process in which genetic drift and selection cause divergence over time. However, there is no rule dictating the time required for speciation, and even low levels of gene flow hinder divergence, so that taxa may be poised at the threshold of speciation for long periods of evolutionary time. We sequenced mitochondrial DNA (mtDNA) and eight nuclear introns (nuDNA) to estimate genomic levels of differentiation and gene flow between the Eurasian common teal (*Anas crecca crecca*) and the North American green-winged teal (*Anas crecca carolinensis*). These ducks come into contact in Beringia (north-eastern Asia and north-western North America) and have probably done so, perhaps cyclically, since the Pliocene–Pleistocene transition, ~2.6 Ma, when they apparently began diverging. They have diagnosable differences in male plumage and are 6.9% divergent in the mtDNA control region, with only 1 of 58 *crecca* and 2 of 86 *carolinensis* having haplotypes grouping with the other. Two nuclear loci were likewise strongly structured between these teal ($\Phi_{ST} \geq 0.35$), but six loci were undifferentiated or only weakly structured ($\Phi_{ST} = 0.0–0.06$). Gene flow between *crecca* and *carolinensis* was ~1 individual per generation in both directions in mtDNA, but was asymmetrical in nuDNA, with ~1 and ~20 individuals per generation immigrating into *crecca* and *carolinensis*, respectively. This study illustrates that species delimitation using a single marker oversimplifies the complexity of the speciation process, and it suggests that even with divergent selection, moderate levels of gene flow may stall the speciation process short of completion.

Keywords: divergent selection, Holarctic, isolation by distance, parapatric speciation, population genetics, speciation with gene flow

Received 30 December 2011; revision received 12 June 2012; accepted 14 June 2012

Introduction

Speciation is a process whereby genetic drift and selection cause populations to become more differentiated over time. However, even small levels of gene flow can retard this process, and there is no evolutionary rule for how long it takes for speciation to reach completion. Three major factors—population size, divergent selection and gene flow—can make speciation a complex process with an uncertain outcome. The classic 'dumbbell' model of allopatric speciation involves the separation of

Correspondence: Jeffrey L. Peters, Fax: 937 775 3320; E-mail: jeffrey.peters@wright.edu

[1]These authors contributed equally to this work.

populations by a barrier that at its extreme precludes gene flow (Mayr 1940, 1942; White 1978; Haffer 2007). If the barrier is incomplete or if the timing of separation and/or degree of divergent selection between populations prior to secondary contact are insufficient, then speciation may be incomplete. As Mayr's (1942) analysis of this model and subsequent decades of research have shown, speciation is most favoured when the ranges of diverging forms are discontinuous, reducing gene flow sufficiently for divergence to produce two lineages unable to reticulate. In speciation theory, if the 'handle' connecting the two populations in the dumbbell model is not broken (i.e. gene flow persists), then parapatric models apply (speciation with gene flow in a nonsympatric distribution; Gavrilets 2004). Parapatric speciation is well supported theoretically, but its frequency in nature is uncertain (Coyne & Orr 2004; Price 2008), and it is a relatively neglected area in speciation research (Gavrilets 2004).

Recently, however, speciation with gene flow has become a popular research topic in natural populations under rubrics such as 'ecological speciation' (Rundle & Nosil 2005; Nosil 2008), 'soft allopatry' (Pyron & Burbrink 2010) and 'disturbance–vicariance' (Lötters *et al.* 2010). Each of these areas basically considers how the 'handle' of gene flow connecting diverging populations affects the speciation process, asking whether the connection is continuous, discontinuous or intermittent, and how it changes through time. Despite its appeal and seeming likelihood, the number of accepted empirical examples for speciation with gene flow is at best modest (Coyne & Orr 2004; Rundle & Nosil 2005). This stems both from an inability to know that strict allopatry did not occur at some critical historical point (Endler 1982; Mayr 1982) and from difficulties in reconstructing temporal aspects of gene flow (e.g. Becquet & Przeworski 2009; Strasburg & Rieseberg 2011). Speciation with gene flow may be better understood by studying cases in which gene flow is unlikely to have ever ceased or in which it has long been recurrent.

Long-distance seasonally migratory birds often exhibit semiannual transcontinental and transoceanic movements that can prevent diverging populations from undergoing long periods of strict allopatry. Such lineages offer a potentially rich series of natural experiments for studying speciation with gene flow (Winker 2010). Here, we examine what is likely to be such a case. Two continental populations of teal, the Eurasian common teal (*Anas crecca crecca*) and the North American green-winged teal (*Anas crecca carolinensis*), breed widely across the Holarctic (Fig. 1) and come into contact in Beringia (which extends from the Russian Far East to western Canada) and, probably less frequently, elsewhere in Eurasia and North America (Johnson 1995). However, whether they breed assortatively in sympatry is not known, and the two taxa are currently recognized either as subspecies or as full species by different authorities (AOU 1998; Sangster *et al.* 2002). Modern estimates place population sizes at 3.7–4.7 and 2.9 million individuals of *crecca* and *carolinensis*, respectively (Delany & Scott 2002).

The Eurasian and North American forms have undergone phenotypic and mtDNA divergence. Evidence from mtDNA coding regions from one individual each of the two taxa suggested that *crecca* and *carolinensis* are about 5.9% divergent (Johnson & Sorenson 1999); Zink *et al.* (1995) also found a relatively deep divergence in mtDNA, but with haplotype sharing between

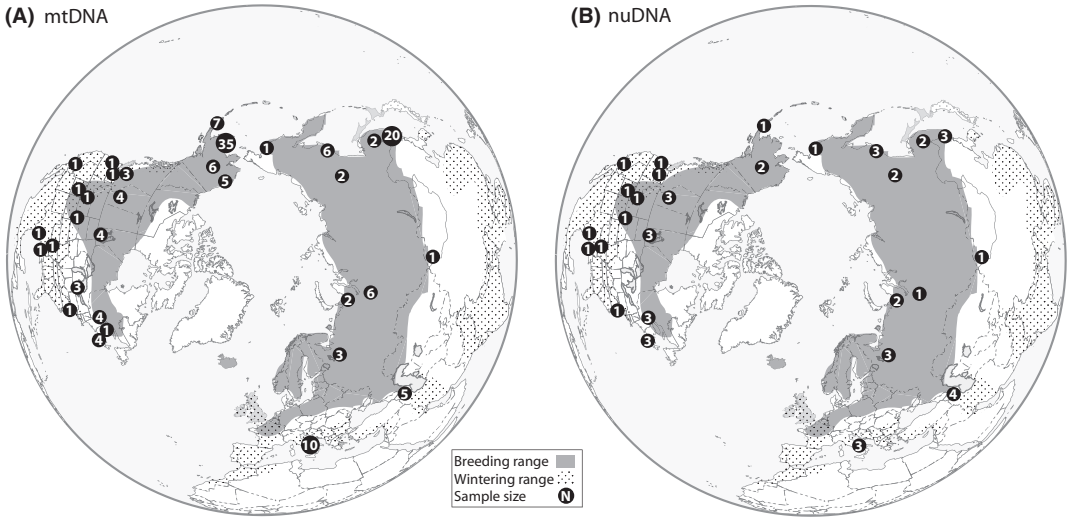

Fig. 1 Sampling localities of common and green-winged teal used for (A) mtDNA sequencing (n = 144) and (B) nuDNA sequencing (n = 50).

continents. Assuming that mtDNA cytochrome *b* diverges at a rate of ~2.29% per million years in waterfowl (Weir & Schluter 2008) suggests that these haplotypes diverged ~2.6 Ma, which corresponds to the Pliocene–Pleistocene transition and the onset of a long period of glacial–interglacial cycles. This deep mitochondrial divergence was further confirmed using larger sample sizes from eastern Russia and Alaska, but *crecca* and *carolinensis* were not significantly differentiated on the basis of amplified fragment length polymorphisms in nuclear DNA (Humphries & Winker 2011).

The mtDNA divergence between *crecca* and *carolinensis* is intriguing for two additional reasons. First, their haplotypes are not sister lineages; mtDNA data suggest that *carolinensis* is more closely related to the South American speckled teal (*Anas flavirostris*), which is morphologically well differentiated but only 2.6% divergent (Johnson & Sorenson 1999). Thus, *A. crecca* (sensu lato) haplotypes are paraphyletic. Second, the mtDNA divergence between *crecca* and *carolinensis* (5.9%) is similar to the divergence observed between some duck species that are strongly differentiated in morphology. For example, mallards (*Anas platyrhynchos*) and northern pintails (*Anas acuta*) are approximately 5.8% divergent (Johnson & Sorenson 1999). Livezey (1991) found that only one of 157 morphological characters differed between *crecca* and *carolinensis*, while 23 of these characters differed between *carolinensis* and *flavirostris* and 32 differed between *platyrhynchos* and *acuta*. Thus, mtDNA divergence between *crecca* and *carolinensis* is deeper than expected given the degree of morphological divergence.

The present breeding distributions of *crecca* and *carolinensis* are effectively parapatric, largely separated by the Bering Sea (Fig. 1). Current and historical distributions have probably been characterized by permeable barriers: presently, the Bering Sea and historically, (perhaps) extensive glaciers. The teal's biology, however, probably lowers the effectiveness of such barriers to dispersal. Both continental populations comprise long-distance seasonal migrants with strong flight capabilities, and individuals use coastal marine environments (Rave & Baldassarre 1989; Gibson & Byrd 2007). During the Quaternary (~2.6 Ma to the present), the two continents were intermittently connected by the Bering land bridge when massive continental glaciers caused sea levels to drop. This land bridge extended ~1600 km in width from north to south during the last glacial maximum (LGM), and was exposed at least nine times from 1.8 to 0.01 Ma (Hopkins 1967; Ruddiman *et al.* 1986). During this time, the Bering land bridge and great expanses to the east and west remained unglaciated (Kaufman & Manley 2004). Oceanic paleotemperature records of glacial–interglacial cycles suggest that more than 40 such cycles occurred in the Quaternary (Ruddiman *et al.* 1986). Fossil and palynological evidence reveals rich vertebrate and plant communities, even during the height of the Wisconsinan glacial episode (110 000–10 000 kya), one of the more severe (Elias 2001; Harrington 2003; Abbott *et al.* 2010), and that Beringia has hosted a rich assemblage of migratory waterfowl since the early Quaternary, including *Anas crecca* (Fitzgerald 1991).

In this study, we investigate the genetic attributes of the speciation process between *crecca* and *carolinensis* using a multilocus approach. Our overarching question is whether, as seems likely given this animal's biology and the history of Beringia, this is a case of speciation with gene flow. To determine this, we ask three questions: Are green-winged teal genetically distinct between continents? How much gene flow, if any, connects them? And do different loci exhibit heterogeneous levels of differentiation? We address these questions using DNA sequences from the mtDNA control region and eight nuclear introns.

Methods

Sampling

We sampled 58 individuals from Eurasia (*crecca*) and 86 individuals from North America (*carolinensis*) from widely distributed locations throughout their Holarctic distributions (Fig. 1A; Table S1, Supporting information). When vouchered specimens were available, we confirmed subspecific designations based on plumage (adult males only). Individuals were sampled during various times of the year and include breeding, migrating and wintering individuals. DNA sources included muscle tissues from museum specimens and hunter-harvested birds and feathers salvaged from nests. All 144 individuals were sequenced for the mtDNA control region, whereas we arbitrarily subsampled 25 individuals per continent (blind to mtDNA haplotype) for nuclear DNA (nuDNA) sequencing (Fig. 1B). Subsamples were chosen to maintain geographical coverage but to balance sample sizes among localities.

DNA sequencing

We sequenced 990–994 bp of the mitochondrial DNA control region and adjacent tRNA-Phe for all individuals using the primer pairs L78-H774 and L736-H1251 (Sorenson & Fleischer 1996; Sorenson *et al.* 1999). For the subsampled individuals, we sequenced seven autosomal introns, including CRYAB (334 bp), GRIN1 (330–331 bp), ENO1 (309–313 bp), ODC1 (345–349 bp), PCK1 (343 bp), FGB (448–459 bp) and LDHB (531–534 bp),

and a Z-chromosome, sex-linked intron, CHD1Z (308–310 bp) (Peters *et al.* 2012a). Each intron is located on a different chromosome within the chicken (*Gallus gallus*) genome and thus assumed to be unlinked in teal. PCR followed standard protocols, with an annealing temperature of 52°C for mtDNA and 58°C for nuDNA. PCR products were cycle-sequenced using BigDye v3.1 and an ABI3100 automated sequencer (Applied Biosystems, Foster City, CA, USA).

We resolved the gametic phase of alleles in three ways. (i) For sequences containing indels, we compared the ambiguous 5′-end with the unambiguous 3′-end of forward and reverse sequences to resolve the placement and composition of gaps and the linkage of polymorphisms to those gaps (Peters *et al.* 2007). (ii) We used the program PHASE to reconstruct the most likely gametic phase of each sequence containing multiple polymorphic sites (Stephens *et al.* 2001); input files were created using SEQPHASE (Flot 2010). (iii) When the probability of reconstructed alleles was <0.95, we developed allele-specific primers to independently amplify and sequence one of the alleles and then subtracted this allele from the heterozygous sequence to resolve the other allele. We re-ran PHASE, treating the newly resolved alleles as known alleles to verify that all reconstruction probabilities were ≥0.95.

Population differentiation

We constructed haplotype networks using the median-joining algorithm in NETWORK ver.4.5 (Bandelt *et al.* 1999). We calculated Φ_{ST}, the proportion of genetic variation partitioned between *crecca* and *carolinensis*, in ARLEQUIN ver. 3.1 (Excoffier *et al.* 2005). We also calculated Φ_{ST} between these northern teal and *Anas flavirostris* for five introns obtained from McCracken *et al.* (2009a). We used the clustering algorithm in the program STRUCTURE ver.2.2.3 to test for population differences using multilocus genotypes (Pritchard *et al.* 2000) with alleles numbered from 1 to n (n equals the number of different alleles observed). STRUCTURE uses a Bayesian approach to test for deviations from Hardy–Weinberg equilibrium and linkage disequilibrium to calculate the likelihood of a user-defined number of populations and to assign individuals to those populations. We used a no admixture model, assumed independent allele frequencies, tested models of varying numbers of populations (K = 1–5) and report results from the model with the highest likelihood. Only nuclear loci were included, and no a priori information was provided on individual origin. STRUCTURE was run for 100 000 generations of burn-in and 500 000 generations of sampling. Each analysis was replicated 10 times, all of which converged on similar values. We chose the best value of K using the Evanno *et al.* (2005) method in the program STRUCTURE HARVESTER (Earl & vonHoldt 2012).

To examine evidence of a cline in genotypic frequencies, we fit a sigmoidal curve to the assignment probabilities across the contact zone in Beringia. Because our total sample included breeding, migrating and wintering individuals, we defined locations for this analysis by longitude only. We chose 170° W as the approximate centre of the contact zone and set this longitude to zero; in this way, Eurasian localities had negative values and North American localities had positive values. We also performed this analysis on mtDNA by categorizing haplotypes as either 'Old World' (1) or 'New World' (0), depending on their haplogroup assignment. We restricted this analysis to the 50 individuals that were sequenced for all loci to equalize sample sizes between marker types.

Coalescent estimates of effective population size, gene flow and divergence times

We fit the sequence data to an isolation–migration model of neutral population divergence in the program IM (Hey & Nielsen 2004). We used IM to estimate six parameters scaled to the per-locus mutation rate (u), including θ_{crecca}, $\theta_{carolinensis}$ and θ_A (a measure of neutral genetic variation for *crecca*, *carolinensis* and their common ancestor, respectively, where $\theta = 4N_e u$ and N_e is the effective population size), t (Tu, where T is the number of years since divergence) and M_{crecca} and $M_{carolinensis}$ (m/u, where m/u is the ratio by which new alleles immigrate into a population relative to the mutation rate). We treated mtDNA and nuDNA in separate analyses because we suspected sex-biased dispersal (Peters *et al.* 2012b). In analyses of nuDNA, inheritance scalars were set to 1.0 for the seven autosomal loci and 0.75 for CHD1Z to reflect the different modes of inheritance. All 144 mtDNA haplotypes were included in the analysis.

To meet IM's assumption of no intralocus recombination, we first tested for recombination within nuDNA using the four-gamete test (Hudson & Kaplan 1985). For loci that showed evidence of recombination, we used the program IMgc to subsample our data (Woerner *et al.* 2007); IMgc optimizes the removal of copies of loci and/or base pairs to retain the maximum number of polymorphic sites that are consistent with no recombination. We ran the program iteratively, changing the weighted preference for removing chromosomes so that a maximum of 5% of copies were removed for each locus. The remaining data were used for IM analyses under an assumption of no recombination.

We initially ran IM using wide priors. Using these preliminary runs, we set priors containing the entirety

of the posterior distributions for the six parameters; these uniform priors were assumed to be uninformative. However, for mtDNA, the posterior distribution for θ_A was flat over a wide range of values, presumably because mtDNA is almost completely sorted and lacks information about ancestral N_e. We set the upper prior for θ_A as 40% of the upper prior for θ_{crecca} and $\theta_{carolinensis}$, as for the analysis of nuDNA. In addition, the posterior distribution for t contained a distinct peak but a flat tail; following Peters *et al.* (2007), we set an upper prior for t based on the 95% confidence interval (CI) for time to most recent common ancestor (*TMRCA*) estimated during the preliminary runs. By doing so, we assumed that the divergence between populations could not be older than the deepest coalescent. For nuclear DNA, we ran a cold chain and 39 heated chains with a geometric heating scheme for a burn-in of 500 000 steps followed by 10 000 000 steps with a sampling interval of 25 steps. For mtDNA, we ran a cold chain and 11 heated chains for the same number of steps. Each run was replicated with a different random number seed which yielded consistent results.

Because estimates of demographic parameters are not directly comparable between mtDNA and nuDNA, we scaled the estimates of migration rates to the effective population size. At each recorded step during the MCMC chain, we calculated $2N_em$ as $\theta M/2$ for each population. Assuming N_e is the same between marker types (after adjusting for differences in modes of inheritance by defining an inheritance scalar of 0.25 for mtDNA in the IM input), these conversions allow for a direct comparison between markers.

Coalescent simulations

To test for among-locus heterogeneity in the genetic signatures of population differentiation, we used the program MS (Hudson 2002) to simulate genetic data under the IM nuclear history. To account for uncertainty in population-level parameters, we randomly subsampled 1000 values from the posterior distributions of each parameter estimated in IM (see Peters *et al.* 2012a). We also incorporated uncertainty in locus-specific recombination and mutation rates into the simulations. To obtain a posterior distribution of recombination rates, we jointly estimated Θ_{crecca} and recombination rates ($r = \rho/\mu$; where ρ is the per-site recombination rate and μ is the per-site mutation rate) in LAMARC ver. 2.1.6 (Kuhner 2006). LAMARC was run for 1 000 000 burn-in steps followed by 10 000 000 steps sampling parameters every 1000 steps. We then subsampled 1000 values of r for each locus. For MS simulations, r was converted to ρ as $\Theta_{crecca} \times r \times (l - 1)$, where l is the locus-specific fragment length (number of base pairs). We obtained relative substitution rates (μ_R) among the nuclear loci from Peters *et al.* (2012a), which were estimated by comparing sequence divergence among eight deeply divergent anseriform taxa (long-term evolutionary rates). Because that study compared 22 loci, we re-scaled μ_R to the average rate (mean rate = 1.0) among the eight loci sequenced for teal. For our simulations, we converted IM's estimate of θ to locus-specific values as $\theta \times \mu_R \times l_R$, where l_R is the locus-specific length relative to the geometric mean of all fragment lengths used in IM. We adjusted θ for the Z-linked locus CHD1Z by a factor of 0.75.

We then simulated genetic diversity for 1000 data sets, each consisting of seven autosomal loci and one Z-linked locus. We calculated Φ_{ST} from the simulated data using MS.OUTPUT (Peters *et al.* 2012a), constructed posterior predictive distributions of Φ_{ST} for each locus and compared the empirical data to those distributions (see Peters *et al.* 2012a). We determined an empirical locus to be a significant outlier if it fell within the 2.5% upper or lower tails (two-tailed test) of the posterior predictive distributions of its replicated simulations (Meng 1994; Becquet & Przeworski 2007).

Results

mtDNA differentiation

Among the 144 teal sampled, we found 96 different mtDNA haplotypes that clustered into two distinct clades that were 6.9% divergent (Fig. 2). One clade (Old World, OW) consisted mostly of haplotypes sampled from *crecca*, whereas the other clade (New World, NW) consisted mostly of haplotypes from *carolinensis*. Only one (1.7%) Eurasian individual (from north-east Russia) had a NW haplotype, and two (2.3%) North American individuals (from Ontario and the Alaska Peninsula) had OW haplotypes. Power analysis shows that 58 *crecca* gives a 95% chance of detecting all haplotypes occurring with a frequency >9% and that 86 *carolinensis* gives a 95% chance of detecting all haplotypes occurring at a frequency of 6–7% or higher (Gregorius 1980). Overall, 89.2% of the total mtDNA genetic variation was partitioned between continents, and 93.4% was partitioned between the two clades.

nuDNA differentiation

In contrast to mtDNA, nuDNA haplotypes did not cluster into distinct OW and NW groups; rather, many haplotypes were shared and phylogenetically intermixed (Fig. 2). However, a significant proportion of the total genetic variation was partitioned between continents for four of the eight loci (Fig. 2). Overall,

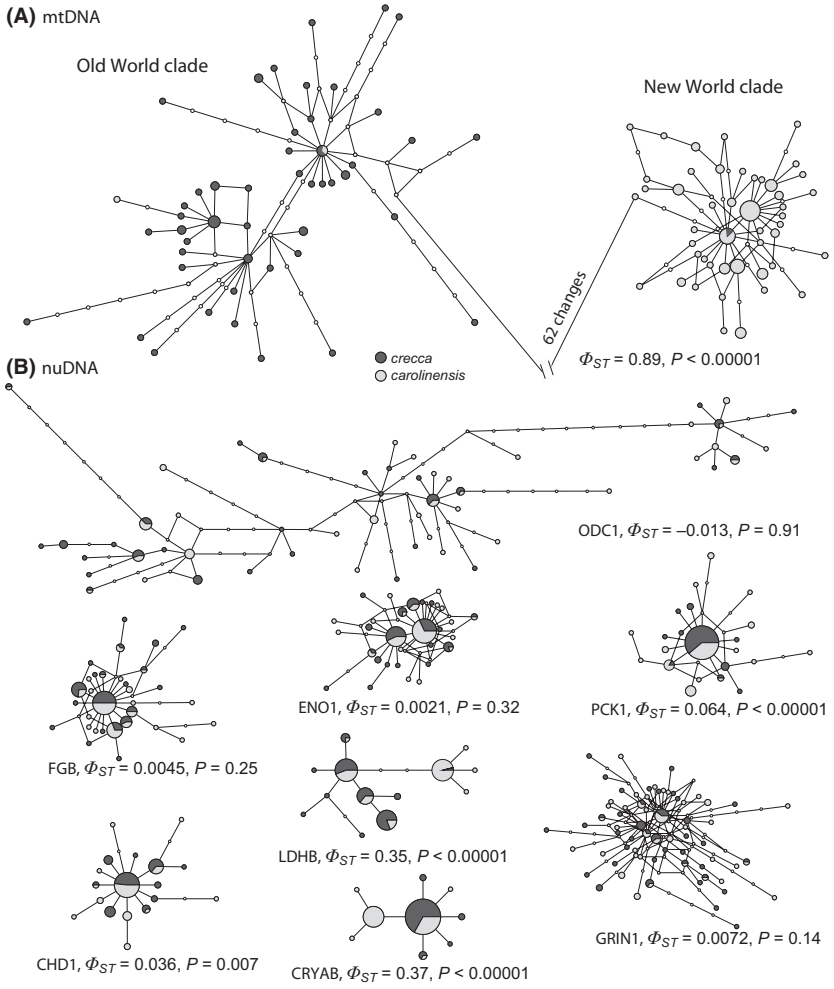

Fig. 2 Haplotype networks for Old World (OW) (dark) and New World (NW) (light) teal for (A) mtDNA (990 bp) and (B) eight nuclear loci (fragment sizes = 310–534 bp). Each circle is proportional to the number of haplotype copies sampled; small open circles indicate unsampled intermediate haplotypes, and each branch separated by a node indicates a single mutation. For each locus, Φ_{ST} values indicated the proportion of the total genetic variation partitioned between OW and NW teal.

Φ_{ST} ranged between −0.013 and 0.38 (mean Φ_{ST} = 0.10 ± 0.16 SD). Genetic differentiation was especially strong for LDHB and CRYAB ($\Phi_{ST} \geq 0.35$; Fig. 2). Among the five loci available for the closely related *Anas flavirostris*, Φ_{ST} for *crecca–carolinensis* ranged between −0.013 and 0.064 (mean Φ_{ST} = 0.013), whereas Φ_{ST} for *carolinensis–flavirostris* ranged between 0.098 and 0.41 (mean Φ_{ST} = 0.22).

On the basis of STRUCTURE results, the most likely number of populations (*K*) was two. A total of 23 OW individuals (92.0%) were assigned to population 1 with a mean assignment probability (*Q*[1]) of 94.6% (±6.6% SD), and two individuals (including the Eurasian individual with NW mtDNA) were assigned to population 2 with assignment probabilities (*Q*[2]) of 69.3% and 71.9% (Fig. 3). Similarly, only two NW individuals were assigned to population 1 with a *Q*[1] of 70.0% and 81.4%, and 23 individuals (92.0%) were assigned to population 2 with a mean *Q*[2] of 95.7% (±6.0% SD; Fig. 3). Of the four individuals with the lowest assignment probabilities (*Q*[1] = 0.20–0.80), which should include those with the most admixed genomes (including three of the four misassigned individuals), two were female and two were male. Our subsampling scheme was blind to mtDNA haplotype, and the two NW individuals with OW mtDNA haplotypes were not sequenced for nuDNA.

Coalescent analyses of population demography and gene flow

The IM analyses of both mtDNA and nuDNA suggested that OW and NW teal had similar effective population sizes, with 95% highest posterior densities broadly overlapping (Fig. 4A). Whereas mtDNA did not provide information about the ancestral population size, nuDNA supported an ancestral size that was substantially smaller than the current sizes (Fig. 4A). Coalescent analyses of both mtDNA and nuDNA supported intercontinental gene flow (Fig. 4B). Both markers showed evidence of prolonged divergence; *t* peaked at 30.1 (95% HPD, 18.4–39.1) for mtDNA and 0.32 (95%

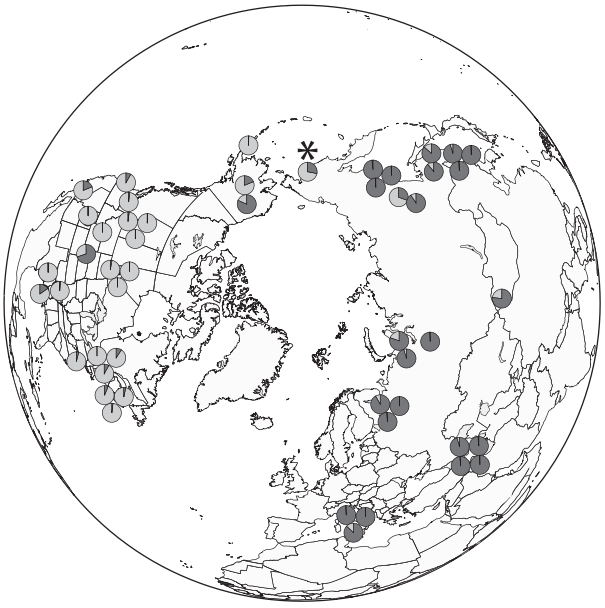

Fig. 3 Population assignment probabilities for 50 individuals of teal based on eight nuclear introns; population 1 = dark slices; population 2 = light slices. The asterisk (*) in north-east Russia identifies the OW teal that had a NW mtDNA haplotype.

HPD, 0.20–0.48) for nuDNA (Fig. 4C). Assuming a mean substitution rate of approximately $0.6–2.2 \times 10^{-9}$ substitutions/site/year for nuDNA (the average lower and upper bound on estimates for five loci; Peters *et al.* 2008), we estimated that OW and NW teal diverged approximately 0.4–3.0 Ma.

For mtDNA, the effective number of migrants ($2N_em$) was estimated at ~1 individual per generation in each direction (95% HPD, 0.1–4.8 and 0.2–5.1 immigrants for OW and NW, respectively; Fig. 5). In contrast, nuDNA suggested asymmetrical gene flow. The inferred immigration rate from the NW into the OW was consistent with no gene flow—the posterior distribution peaked near the bin containing the lowest values (0.25; 95% HPD, 0–4.9 migrants per generation; Fig. 5A; although this value is statistically consistent with zero, our result showing non-zero mtDNA gene flow requires some level of nuDNA gene flow). However, the estimate of immigration from the OW into the NW was ~20 individuals per generation, and the 95% HPD (7.7–48.5 migrants; Fig. 5B) did not overlap the estimate for the opposite direction. In addition, these analyses suggest that biparentally inherited nuDNA has introgressed into the NW significantly more often than maternally inherited mtDNA (Fig. 5B).

Simulations of genetic diversity

Among the nuclear loci, μ_R ranged between 0.56 and 1.3 times the mean rate (Peters *et al.* 2012a). Recombination rates, estimated from empirical data in LAMARC, also varied among loci; FGB, ENO1 and GRIN1 had high recombination rates ($r > 2.0$); CRYAB, CHD1Z, LDHB and PCK1 had low rates consistent with no recombination ($r < 1.0$) and ODC1 had a low, but non-zero, recombination rate (Table S2, Supporting information). These differences among loci were incorporated into coalescent simulations.

Simulating genetic diversity under selective neutrality and the inferred isolation–migration model, we found

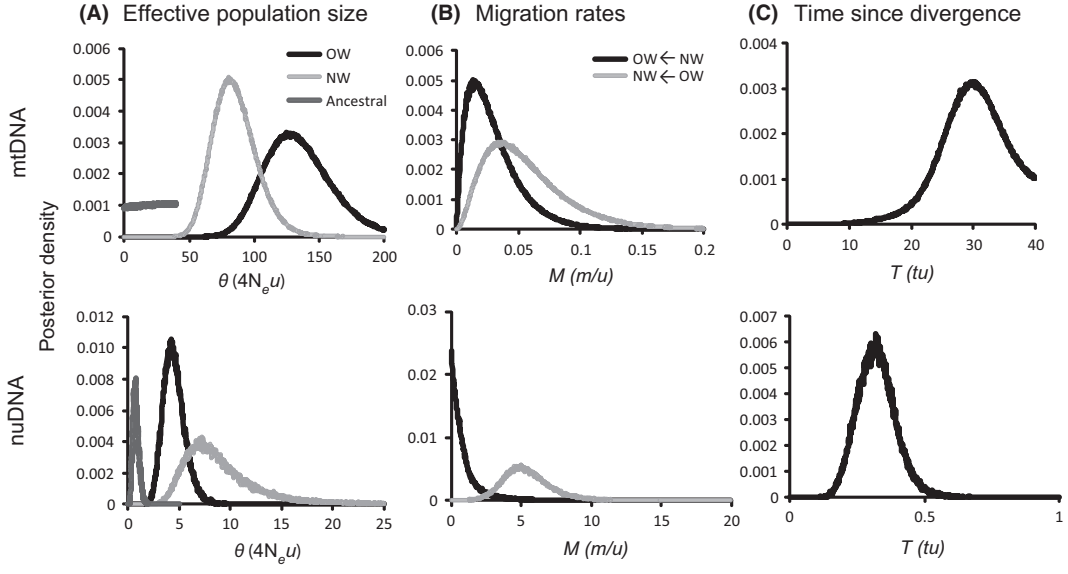

Fig. 4 Posterior distributions of (A) effective population sizes, (B) migration rates and (C) time since divergence, scaled to the mutation rate, estimated using an isolation–migration model of divergence for mtDNA (top panel) and eight nuclear loci (bottom panel).

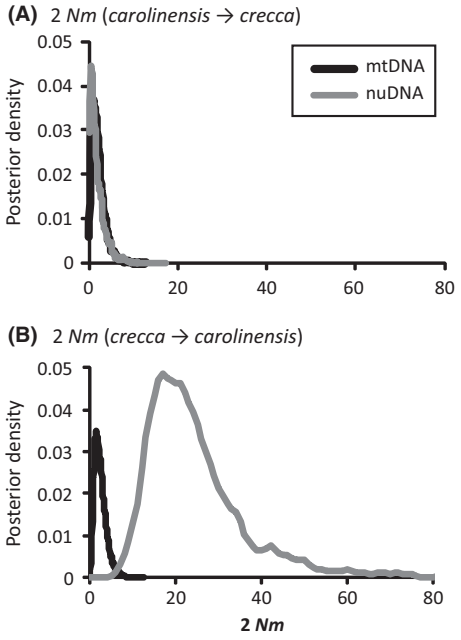

Fig. 5 The number of effective immigrants ($2N_em = \theta M/2$) into (A) OW common teal and (B) NW green-winged teal. Dark lines show estimates from mtDNA, whereas light curves show estimates from eight nuclear loci. Note that by using an inheritance scalar of 0.25 for mtDNA, the results are on the same scale as and directly comparable to the nuDNA results.

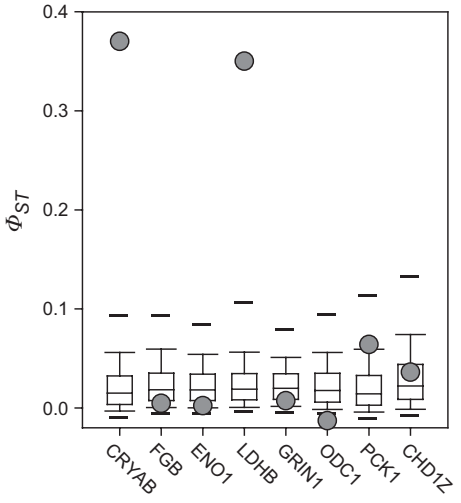

Fig. 6 The per cent of genetic diversity proportioned between *crecca* and *carolinensis* estimated from empirical data and data simulated under the inferred isolation–migration model of population history. The box plots show the 50 (box) and 90 (error bars) percentiles of the simulated values, and the horizontal lines show the 95 percentile; dots indicate the empirical values for each locus. CRYAB and LDHB are more strongly structured than predicted from the model.

that CRYAB and LDHB were significantly more structured than expected (Fig. 6). In contrast, ODC1 had a slightly lower Φ_{ST} than simulated, and the remaining

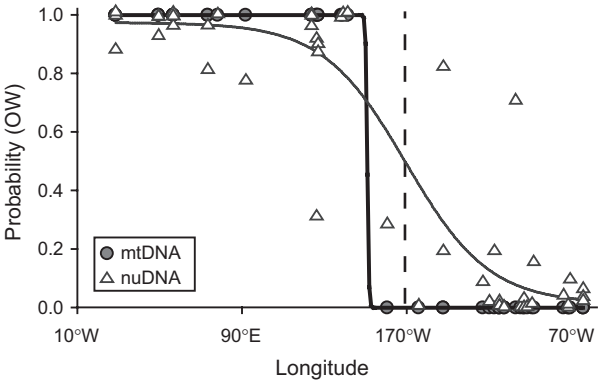

Fig. 7 Assignment probabilities for being from the Old World for the 50 individuals sampled in this study, based on eight nuclear loci (open triangles) and mtDNA clade assignment (closed circles). The lines are three-parameter sigmoidal curves fit to each data type (light and heavy lines, respectively).

five loci were within the 95% CI of the simulated data. Overall, our mean value for Φ_{ST} was 0.10 for the empirical nuclear data, which was significantly higher than the mean values simulated for an eight locus data set (mean $\Phi_{ST} = 0.025$; 95% CI, 0.007–0.052). Removing CRYAB and LDHB resulted in a mean Φ_{ST} of 0.017, which was well within the CIs of the simulated data. Collectively, these results indicate among-locus heterogeneity in population structure.

Clines

The geography of mtDNA haplotype groups across the contact zone of the two taxa in Beringia showed a sharp transition (width = −0.5°) centred around 163°E (Fig. 7), which is largely congruent with that of male breeding plumage phenotype (not shown). In contrast, nuDNA genotypes yielded a different pattern, with a right-shifted centre near 171° W and a broader, shallower slope (width = −25.2°) relative to mtDNA (Fig. 7).

Discussion

Our study of the *Anas crecca–carolinensis* complex revealed the following four major findings: (i) Continental populations are genetically differentiated between Eurasia and North America but not reciprocally monophyletic in either mtDNA or nuDNA. (ii) Gene flow between continents has occurred symmetrically in mtDNA at the rate of ~1 individual per generation, but asymmetrically in nuDNA, being low (nearly zero) into the OW population and ~20 individuals per generation into the NW population. (iii) Nuclear introns were heterogeneous in their allelic differences between populations. (iv) The transition in mtDNA haplotypes matches that of male breeding plumage (a sharp cline in

Beringia), but the effects of male-mediated gene flow into the NW caused a more gradual shift for nuDNA genotypes.

Genetic differentiation

Both mtDNA and nuDNA indicate that *crecca* and *carolinensis* are genetically differentiated. Similar to previous studies with limited geographical coverage and sample sizes (Zink *et al.* 1995; Johnson & Sorenson 1999; Humphries & Winker 2011), our comprehensive sampling revealed two deeply divergent mtDNA haplogroups (6.9% divergent) that are each predominantly found in one population. Based on power analyses, our sample sizes are sufficient to reject the hypothesis of rampant intermixing of mtDNA lineages between populations. Furthermore, the depth of divergence and the limited intermixing of haplotypes between populations (2.1% in our data set) suggest that mtDNA has sorted to reciprocal monophyly, but that dispersal results in some mixing of haplotypes between continents. This pattern is similar to that of the mallard, which also has a Holarctic distribution (Kulikova *et al.* 2005; Kraus *et al.* 2011). In contrast to mtDNA, the nuclear loci showed no evidence of reciprocal monophyly between continents, although CRYAB and LDHB were strongly differentiated in allelic frequencies. The higher effective population size of nuDNA relative to mtDNA, male-biased gene flow or a combination of both could have contributed to this lack of sorting (Zink & Barrowclough 2008; Peters *et al.* 2012b). Regardless, the eight nuclear loci examined in this study are fairly diagnostic for taxon identification and demonstrate that *crecca* and *carolinensis* are genetically differentiated. Furthermore, coalescent analysis of nuDNA indicates a genomic divergence of 0.4–3.0 Ma, which encompasses the estimate of 2.6 Ma from mtDNA (Johnson & Sorenson 1999). Thus, these teal populations have achieved a considerable degree of distinctiveness in male plumage, mtDNA and nuDNA and thus represent substantially differentiated populations.

As noted previously, mtDNA haplotypes of *carolinensis* are more similar to those of the South American *Anas flavirostris* than to *crecca*, rendering mtDNA from *A. crecca* (sensu lato) paraphyletic (Johnson & Sorenson 1999). However, nuDNA differentiation between *carolinensis* and *crecca* is weaker than between *carolinensis* and *flavirostris*. This apparent conflict between mtDNA and nuDNA could be explained by the smaller population size of *flavirostris* (Delany & Scott 2002), which causes genetic drift to be more influential in altering allele frequencies. Alternatively, male-mediated gene flow between *crecca* and *carolinensis* inhibits nuclear differentiation between these populations (Peters *et al.* 2012b), whereas gene flow between the northern and southern forms is probably nonexistent or rare (strict allopatry).

Gene flow

Despite significant genetic differentiation between *crecca* and *carolinensis*, the two populations are connected in a complex manner by gene flow. First, one OW and two NW individuals had mtDNA haplotypes grouping with individuals from the opposite population. Given the depth of divergence between the two haplogroups and the nearly complete sorting between taxa coupled with large effective population sizes, it is unlikely that incomplete sorting of ancestral polymorphisms could explain this sharing (see Drovetski *et al.* 2005; Peters *et al.* 2007; Kraus *et al.* 2012). Although we cannot be certain whether some individuals were vagrants, our results illustrate intercontinental movements and opportunities for gene flow. Second, the IM analysis of nuDNA also rejected a scenario of no gene flow. Incomplete lineage sorting probably explains much of the nuclear allele sharing between these teal, but IM accommodates sharing of ancestral polymorphisms by including a divergence time parameter while estimating migration rates. Furthermore, higher rates of gene flow for nuDNA fit theoretical predictions, because ducks have male-biased dispersal (Rohwer & Anderson 1988; Anderson *et al.* 1992; Peters *et al.* 2012b). Third, on the basis of nuDNA genotypes, four individuals were misassigned (Fig. 3). These individuals received relatively low assignment probabilities, suggesting that they might have admixed genomes. Indeed, one of these (UWBM 43947) was a male collected in north-east Russia during the breeding season and had a NW mtDNA haplotype. Although the genetic data are consistent with this individual being descendent from a recent hybrid mating ($Q[1] = 0.287$), a post hoc examination of plumage phenotype was inconclusive regarding subspecies (this individual was in partial moult, and the diagnostic plumage characteristics were not obvious). Finally, hybrids are known from the wild (Sangster *et al.* 2001; Gibson & Byrd 2007). These lines of evidence demonstrate that these teal populations interbreed and that alleles have a propensity for introgressing between populations.

There are three possible ways in which gene flow has occurred in this system: recent secondary contact, repeated events of secondary contact during cycles of Beringian climate change or effectively continuous gene flow, though perhaps fluctuating in frequency. Given the deep mtDNA clades coupled with shallow relationships between introgressed haplotypes and the remainder of their respective clades, the mtDNA topology seems consistent with secondary contact. However, if

gene flow occurred prior to complete lineage sorting within these clades, then introgressed alleles can only have one of two fates, extinction or fixation. Either way, the signature of ancient gene flow will be lost, and therefore, the mtDNA topology is inconclusive regarding the timing of gene flow. Given the biology of these teal in Beringia today (i.e. strong seasonal migrants and the ability to use coastal habitats), coupled with fossil and palynological evidence of suitable habitats through the Quaternary and the species' presence during the Wisconsinan glaciation (Fitzgerald 1991; Elias 2001; Guthrie 2001; Abbott *et al.* 2010), the latter two hypotheses seem more likely. Thus, these populations have probably been connected by gene flow, at least intermittently, throughout much of their history.

The asymmetric gene flow from *crecca* into *carolinensis* fits a much wider pattern of more Asian taxa breeding in Alaska than vice-versa (Flint *et al.* 1984; Winker & Gibson 2010). Two congeners, the mallard and the gadwall (*Anas strepera*), also exhibit this OW → NW bias in the direction of gene flow (Kulikova *et al.* 2005; Peters *et al.* 2008; Kraus *et al.* 2011). However, we found that the estimate of nuDNA gene flow into the OW was close to zero. It is theoretically possible that this imbalance reflects a 'genetic wake', in which a 'retreating' taxon introgresses into a postglacially advancing one (Rohwer *et al.* 2001; Krosby & Rohwer 2009; Maley & Winker 2010). However, such a phenomenon should also affect mtDNA, for which we found no evidence. Another theoretical possibility would be if populations of *crecca* were much larger than those of *carolinensis*, but again we found no evidence for that (e.g. Fig. 4 and Fig. S1, Supporting information), and continental populations of both taxa number in the millions (Delany & Scott 2002). The evidence suggests that higher dispersal rates occur into the NW, fitting the broader, among-species Asia-to-America movement pattern found in this region (Winker & Gibson 2010). Kraus *et al.* (2011) attributed similar asymmetries in mallards to the direction of the prevailing winds across the Bering Sea (Eurasia → North America; the 'drifting by wind' hypothesis). As a future test for asymmetric gene flow, we would predict that *crecca* males are more common in North America than *carolinensis* males in Eurasia. An alternative hypothesis is that females of both *carolinensis* and *crecca* preferentially mate with *crecca* males, which could also cause *crecca* DNA to asymmetrically introgress into the NW population.

Heterogeneous among-locus differentiation

Models of speciation with gene flow predict that differentiation among unlinked loci will be heterogeneous, because selection inhibits the introgression of some regions of the genome but not others (Coyne & Orr 1998; Wu 2001; Hey 2006; Nosil & Feder 2012). Indeed, we found significant heterogeneity in values of Φ_{ST} among nuclear loci in these teal, with some loci strongly structured (mtDNA, CRYAB & LDHB), some weak-to-moderately structured (CHD1Z & PCK1) and some having no detectable structure (ENO1, FGB, GRIN1 & ODC1). Our coalescent analyses suggest that male-mediated gene flow and female philopatry can explain the differences between mtDNA and nuDNA as has been suggested for other ducks (e.g. Scribner *et al.* 2001; Pearce *et al.* 2005; Sonsthagen *et al.* 2011; Peters *et al.* 2012b). However, the heterogeneity in nuDNA cannot be explained by sex-biased dispersal and suggests the influence of selection in generating differential introgression among loci (e.g. Borge *et al.* 2005; Carling & Brumfield 2008; McCracken *et al.* 2009a,b). For example, polymorphisms linked to loci that are beneficial in one habitat or genome but disadvantageous in other settings or genetic backgrounds will introgress less readily than neutral loci, resulting in elevated Φ_{ST} values (Lewontin & Krakauer 1973; Maynard Smith & Haigh 1974; Nosil *et al.* 2009; Nosil & Feder 2012). Selection could also have contributed to the high differentiation observed for mtDNA (Galtier *et al.* 2000; Ballard & Whitlock 2004; Bazin *et al.* 2006). Although we cannot rule out the possibility that a more complex neutral history contributed to this heterogeneity (Becquet & Przeworski 2009; Strasburg & Rieseberg 2011; Peters *et al.* 2012a), the prominent differences among loci are consistent with a hypothesis of speciation with gene flow.

Cline asymmetries

A key attribute of parapatric speciation theory is that it does not favour congruent clines between selected and neutral loci (Coyne & Orr 2004), such congruency being more an attribute of secondary contact. Whereas nuclear genotypes seem to transition over a gradient, the sharp transition between the two mtDNA clades coincides geographically with the distribution of the different male plumage types, both of which are probably associated with female behaviour (philopatry and mate choice).

Periods of aridity in Beringia (promoting lower densities), rather than continental ice sheets and sea-level changes (promoting range separation), may have limited gene flow between the two continental populations, even when the continents were connected (Guthrie 2001; Abbott *et al.* 2010). During the LGM, there was a mesic zone, centred on the Bering Sea, in a belt of more arid steppe upon which some water bodies went dry (Guthrie 2001; Abbott *et al.* 2010). Such geographical features likely served as a 'density trough', features that

often characterize the centres of tension zones between diverging populations (Hewitt 1989). Thus, even without the permeable interglacial vicariant barrier of the Bering Sea (such as we see today), climatic reconstructions suggest reasons for the mtDNA cline to remain centred upon it during glacials and interglacials.

Parapatric speciation

Our results are consistent with a hypothesis of divergence with gene flow, suggesting that these teal represent a case of parapatric speciation (in the classic, geographical sense). Gene flow holds populations together in evolutionary time if it occurs at sufficient levels to overcome divergent selection (Hartl & Clark 1989), and even low levels can prevent speciation and local adaptation (Rice & Hostert 1993; Hostert 1997; Postma & van Noordwijk 2005). The low levels of gene flow that we found in teal mtDNA would be sufficient to enable these two populations to complete speciation under even relatively weak divergent selection (Hartl & Clark 1989; Rice & Hostert 1993), although alleles of high selective advantage could spread between them (Morjan & Rieseberg 2004). However, the higher levels of nuclear gene flow from *crecca* into *carolinensis* (~20 individuals per generation) are almost certainly too high to overcome even strong divergent selection (Hartl & Clark 1989; Rice & Hostert 1993; Hostert 1997). Thus, our data indicate that these two populations remain linked in evolutionary time, stalled short of complete speciation, despite deeply divergent mtDNA.

These teal represent two large continental populations connected in a relatively narrow zone of (perhaps cyclic) parapatry, and in this respect, they are a good example of a 'dumbbell' model. Female mate choice and migratory behaviour both likely contribute to limited gene flow (i.e. a narrow handle) and perhaps divergent selection. Although the two teal are ecological replacements in Eurasia and North America, sexual selection likely contributes to differences in male plumage and in frequency differences in male displays (Laurie-Ahlberg & McKinney 1979). Migratory direction might also be a source of divergent selection: after breeding in Beringia, birds from the two different continental populations go in different directions to their wintering grounds. However, they migrate in flocks, their movements are loose in timing and they lack pronounced winter site fidelity (Johnson 1995), suggesting that migratory direction might not have a strong genetic component. Furthermore, males of the other subspecies occasionally occur on each others' wintering grounds on both sides of the Atlantic and Pacific oceans (Johnson 1995). Because ducks form pair bonds on their wintering grounds (Rohwer & Anderson 1988), these transoceanic movements to wintering areas provide opportunities for gene flow.

In summarizing decades of laboratory research on speciation with gene flow, Rice & Hostert (1993) emphasized the importance of restricted gene flow coupled with strong, discontinuous and multifarious divergent selection. Nosil *et al.* (2009) also emphasized that multiple traits are important for the completion of speciation when there is gene flow. Divergent selection may be occurring at the molecular level in two of our nuclear loci (Fig. 6) and perhaps also in mtDNA. However, the strength of divergent selection and/or the number of traits undergoing such selection appear to fall short of that required for completion of speciation given the estimated levels of nuclear gene flow.

Considering our data as a snapshot of a stage in speciation, there are several ways in which these teal might progress to full species. Each requires that hybrids are less fit. The first is an accelerated divergence via pleiotropy/hitchhiking as more genetic variation is incorporated into the divergence process, expanding its base (Rice & Hostert 1993). The second is through increased incompatibilities between mtDNA and nuDNA in male crosses. Males represent evolutionary dead ends for the maternally inherited mtDNA, which decreases the effectiveness of selection to purge deleterious mutations, causing a male-specific, asymmetric mutation load to accumulate (Innocenti *et al.* 2011). Finally, the genomes of the two taxa may diverge sufficiently for Haldane's rule to apply, in which case the heterogametic sex of hybrids, females in birds, would be sterile, rare or absent. Given that half (2/4) of the individuals with the lowest nuclear assignment probabilities were female, this last situation may be the least likely. Furthermore, evidence suggests that although Haldane's rule applies to ducks, its effects may be relatively weak (Tubaro & Lijtmaer 2002; Kirby *et al.* 2004).

Species limits

It has been suggested that small portions of mtDNA can be used for species delimitation (see discussions in DeSalle *et al.* 2005; Wiens 2007; Winker 2009). This case in teal shows that such a simplistic approach to determining species limits can be inadequate for understanding the relatedness of populations as they diverge towards full species, gaining sufficient isolating mechanisms to preclude reticulation. Our data demonstrate that a multilocus approach can be essential to illuminate the complexities of speciation (e.g. Oyler-McCance *et al.* 2010; Hausdorf *et al.* 2011). In particular, the mtDNA divergence in these teal does not reflect genomic levels of interpopulation gene flow. Petit & Excoffier (2009) proposed that species delimitation focus on the use of assignment tests on multiple unlinked

genetic markers that have high rates of gene flow within populations (i.e. loci representing the sex with the higher dispersal rate). By such criteria, these two taxa are not yet full species.

The *crecca–carolinensis* complex appears to have been in the speciation process for a long time. Despite two deeply divergent lineages in mtDNA, analysis of nuDNA reveals heterogeneous patterns of differentiation, with the overall pattern suggesting that these teal are far from being reproductively isolated. Although speciation in the face of gene flow and introgression might be a common feature of *Anas* ducks (Kraus *et al.* 2012), the level of gene flow between these teal populations likely prevents the completion of speciation. Considering the differences we found between mtDNA and nuDNA, the philopatric behaviour of female teal might be analogous to the host specificity found in the brood parasitic greater honeyguide (*Indicator indicator*; Spottiswoode *et al.* 2011); both behaviours, coupled with male-mediated gene flow, result in strong mtDNA differentiation but weak nuDNA differentiation. However, the *crecca–carolinensis* complex might be closer to full speciation than the honeyguide, as we found some nuclear loci for which gene flow is likely inhibited. In contrast, these teal do not seem to be as far along in the speciation process as two *Ficedula* flycatchers (Sætre & Sæther 2010). Selection seems more effective in preventing introgression in that system (Borge *et al.* 2005; Backström *et al.* 2010), despite a shallower mtDNA divergence (Sætre *et al.* 1997) than observed in these teal. Thus, *crecca* and *carolinensis* appear to be at an intermediate phase of speciation, despite what is likely to be millions of years of divergence in semi-isolation. This long divergence illustrates that species delimitation using a single marker oversimplifies the complexity of the speciation process and suggests that even with divergent selection, moderate levels of gene flow may stall speciation short of completion.

Acknowledgements

This work was supported by the University of Alaska Museum, Alaska EPSCoR (NSF EPS-0346770), the National Science Foundation (DEB-9981915, DEB-0444748, DEB-0746365, DEB-0926162), the United States Department of Agriculture (SCA 58-6612-2-217) and the American Museum of Natural History. We thank K Millam and KA Bolender for assistance with data collection, S Houston and the UAF Life Science Informatics for computer resources and three anonymous reviewers for helpful comments on earlier versions of the manuscript.

References

Abbott MB, Edwards ME, Finney BP (2010) A 40,000-yr record of environmental change from Burial Lake in northwest Alaska. *Quaternary Research*, **74**, 156–165.

American Ornithologists' Union (AOU) (1998) *Check-list of North American Birds*, 7th edn. American Ornithologists' Union, Washington, DC.

Anderson MG, Rhymer JM, Rohwer FC (1992) Philopatry, dispersal, and the genetic structure of waterfowl populations. In: *Ecology and Management of Breeding Waterfowl* (eds Batt BDJ, Afton AD, Anderson MG *et al.*), pp. 365–395. University of Minnesota Press, Minneapolis, USA.

Backström N, Lindell J, Zhang Y *et al.* (2010) A high-density scan of the Z chromosome in *Ficedula* flycatchers reveals candidate loci for diversifying selection. *Evolution*, **64**, 3461–3475.

Ballard JW, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular ecology*, **13**, 729–744.

Bandelt H, Forster P, Roehl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

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

Becquet C, Przeworski M (2007) A new approach to estimate parameters of speciation models with application to apes. *Genome Research*, **17**, 1505–1519.

Becquet C, Przeworski M (2009) Learning about modes of speciation by computational approaches. *Evolution*, **63**, 2547–2562.

Borge T, Webster MT, Andersson G, Saetre G (2005) Contrasting patterns of polymorphism and divergence on the Z chromosome and autosomes in two *Ficedula* flycatcher species. *Genetics*, **171**, 1861–1873.

Carling MD, Brumfield RT (2008) Integrating phylogenetic and population genetic analyses of multiple loci to test species divergence hypotheses in *Passerina* buntings. *Genetics*, **178**, 363–377.

Coyne JA, Orr HA (1998) The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **353**, 287–305.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Inc., Sunderland, Massachusetts.

Delany S, Scott D (2002) *Waterbird Population Estimates*, 3rd edn. Wetlands International, Wageningen, The Netherlands.

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1905–1916.

Drovetski SV, Pearson SF, Rohwer S (2005) Streaked horned lark *Eremophila alpestris strigata* has distinct mitochondrial DNA. *Conservation Genetics*, **6**, 875–883.

Earl D, vonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Elias SA (2001) Beringian paleoecology: results from the 1997 workshop. *Quaternary Science Reviews*, **20**, 7–13.

Endler JA (1982) Problems in distinguishing historical from ecological factors in biogeography. *American Zoologist*, **22**, 441–452.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Fitzgerald GR (1991) Pleistocene ducks of the Old Crow Basin, Yukon Territory, Canada. *Canadian Journal of Earth Sciences*, **28**, 1561–1571.

Flint VE, Boehme RL, Kostin YV, Kuznetsov AA (1984) *Birds of the USSR*. Princeton University Press, Princeton, New Jersey.

Flot J (2010) Seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments. *Molecular Ecology Resources*, **10**, 162–166.

Galtier N, Depaulis F, Barton NH (2000) Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics*, **155**, 981–987.

Gavrilets S (2004) *Fitness Landscapes and the Origin of Species*. Princeton Monographs in Population Biology No. 41.

Gibson DD, Byrd GV (2007) *The Birds of the Aleutian Islands, Alaska*. Nuttall Ornithological Club, Cambridge, Massachusetts.

Gregorius HR (1980) The probability of losing an allele when diploid genotypes are sampled. *Biometrics*, **36**, 643–652.

Guthrie RD (2001) Origin and causes of the mammoth steppe: a story of cloud cover, wooly mammal tooth pits, buckles, and inside-out Beringia. *Quaternary Science Reviews*, **20**, 549–574.

Haffer J (2007) *Ornithology, Evolution, and Philosophy: The Life and Science of Ernst Mayr 1904–2005*. Springer, Berlin.

Harrington CR, ed. (2003) *Annotated Bibliography of Quaternary Vertebrates of Northern North America*. University of Toronto Press, Toronto, Ontario, Canada.

Hartl DL, Clark AG (1989) *Principles of Population Genetics*, 2nd edn. Sinauer Associates, Inc., Sunderland, Massachusetts.

Hausdorf B, Wilkens H, Strecker U (2011) Population genetic patterns revealed by microsatellite data challenge the mitochondrial DNA based taxonomy of *Astyanax* in Mexico (Characidae, Teleosti). *Molecular Phylogenetics and Evolution*, **60**, 89–97.

Hewitt G (1989) The subdivision of species by hybrid zones. In: *Speciation and its Consequences* (eds Otte D and Endler JA), pp. 85–110. Sinauer Associates, Inc., Sunderland, Massachusetts.

Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics & Development*, **16**, 592–596.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Hopkins DM, ed. (1967) *The Bering Land Bridge*. Stanford University Press, Stanford, California.

Hostert EE (1997) Reinforcement: a new perspective on an old controversy. *Evolution*, **51**, 697–702.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Humphries EM, Winker K (2011) Discord reigns among nuclear, mitochondrial and phenotypic estimates of divergence in nine lineages of trans-Beringian birds. *Molecular Ecology*, **20**, 573–583.

Innocenti P, Morrow EH, Dowling DK (2011) Experimental evidence supports a sex-specific selective sieve in mitochondrial genome evolution. *Science*, **332**, 845–848.

Johnson K (1995) Green-winged Teal (*Anas crecca*). In: *The Birds of North America Online* (ed Poole A). Cornell Lab of Ornithology, Ithaca.

Johnson K, Sorenson MD (1999) Phylogeny and biogeography of dabbling ducks (genus *Anas*): a comparison of molecular and morphological evidence. *Auk*, **116**, 195–208.

Kaufman DC, Manley WF (2004) Pleistocene maximum and Late Wisconsinan glacier extents across Alaska, U.S.A. In: *Quaternary Glaciations – Extent and Chronology, Part II* (eds Ehlers J and Gibbard PL), pp. 9–27. Elsevier, Amsterdam.

Kirby RE, Sargeant GA, Schluter D (2004) Haldane's rule and American black duck x mallard hybridization. *Canadian Journal of Zoology*, **82**, 1827–1831.

Kraus R, Zeddeman A, van Hooft P *et al.* (2011) Evolution and connectivity in the world-wide migration system of the mallard: inferences from mitochondrial DNA. *BMC Genetics*, **12**, 99.

Kraus R, Kerstens H, van Hooft P *et al.* (2012) Widespread horizontal genomic exchange does not erode species barriers among sympatric ducks. *BMC Evolutionary Biology*, **12**, 45.

Krosby M, Rohwer S (2009) A 2000 km yields evidence of northern glacial refugia and hybrid zone movement in a pair of songbirds. *Proceedings of the Royal Society of London B*, **276**, 615–621.

Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics*, **22**, 768–770.

Kulikova IV, Drovetski SV, Gibson DD *et al.* (2005) Phylogeography of the mallard (*Anas platyrhynchos*): hybridization, dispersal, and lineage sorting contribute to complex geographic structure. *Auk*, **122**, 949–965.

Laurie-Ahlberg CC, McKinney F (1979) The nod-swim display of male Green-winged Teal (*Anas crecca*). *Animal Behaviour*, **27**, 165–172.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

Livezey BC (1991) A phylogenetic analysis and classification of recent dabbling ducks (tribe Anatini) based on comparative morphology. *Auk*, **108**, 471–507.

Lötters S, van der Meijden A, Rödder D *et al.* (2010) Reinforcing and expanding the predictions of the disturbance vicariance hypothesis in Amazonian harlequin frogs: a molecular phylogenetic and climate envelope modeling approach. *Biodiversity Conservation*, **19**, 2125–2146.

Maley JM, Winker K (2010) Diversification at high latitudes: speciation of buntings in the genus *Plectrophenax* inferred from mitochondrial and nuclear markers. *Molecular Ecology*, **19**, 785–797.

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical Research*, **2**, 3–35.

Mayr E (1940) Speciation phenomena in birds. *American Naturalist*, **74**, 249–278.

Mayr E (1942) *Systematics and the Origin of Species*. Columbia University Press, New York.

Mayr E (1982) *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Belknap Press, Cambridge, Massachusetts.

McCracken KG, Barger CP, Bulgarella M *et al.* (2009a) Signatures of high-altitude adaptation in the major hemoglobin of five species of Andean dabbling ducks. *American Naturalist*, **174**, 631–650.

McCracken KG, Bulgarella M, Johnson KP *et al.* (2009b) Gene flow in the face of countervailing selection: adaptation to high-altitude hypoxia in the beta-A hemoglobin subunit of yellow-billed pintails in the Andes. *Molecular Biology and Evolution*, **26**, 815–827.

Meng X (1994) Posterior predictive p-values. *The Annals of Statistics*, **22**, 1142–1160.

Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, **13**, 1341–1356.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103–2106.

Nosil P, Feder JL (2012) Widespread yet heterogeneous genomic divergence. *Molecular Ecology*, **21**, 2829–2832.

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology and Evolution*, **24**, 145–156.

Oyler-McCance SJ, St. John J, Quinn TW (2010) Rapid evolution in lekking grouse: implications for taxonomic definitions. *Ornithological Monographs*, **67**, 114–122.

Pearce JM, Talbot SL, Petersen MR, Rearick JR (2005) Limited genetic differentiation among breeding, molting, and wintering groups of the threatened Steller's eider: the role of historic and contemporary factors. *Conservation Genetics*, **6**, 743–757.

Peters JL, Zhuravlev Y, Fefelov I *et al.* (2007) Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). *Evolution*, **61**, 1992–2006.

Peters JL, Zhuravlev YN, Fefelov I *et al.* (2008) Multilocus phylogeography of a Holarctic duck: colonization of North America from Eurasia by gadwall (*Anas strepera*). *Evolution*, **62**, 1469–1483.

Peters JL, Roberts TE, Winker K, McCracken KG (2012a) Heterogeneity in genetic diversity among non-coding loci fails to fit neutral coalescent models of population history. *PLoS One*, **7**, e31972.

Peters JL, Bolender KA, Pearce JM (2012b) Behavioral vs. molecular sources of conflict between nuclear and mitochondrial DNA: the role of male-biased dispersal in a Holarctic sea duck. *Molecular Ecology*, **21**, 3562–3575.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, **24**, 386–393.

Postma E, van Noordwijk AJ (2005) Gene flow maintains a large genetic difference in clutch size at a small spatial scale. *Nature*, **433**, 65–68.

Price T (2008) *Speciation in Birds*. Roberts and Company, Greenwood Village, Colorado.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Pyron RA, Burbrink FT (2010) Hard and soft allopatry: physicallyand ecologically mediated modes of geographic speciation. *Journal of Biogeography*, **37**, 2005–2015.

Rave DP, Baldassarre GA (1989) Activity budget of green-winged teal wintering in coastal wetlands of Louisiana. *The Journal of Wildlife Management*, **53**, 753–759.

Rice WR, Hostert EE (1993) Laboratory experiments on speciation: what have we learned in 40 years? *Evolution*, **47**, 1637–1653.

Rohwer FC, Anderson MG (1988) Female-biased philopatry, monogamy and the timing of pair formation in waterfowl. *Current Ornithology*, **5**, 187–221.

Rohwer SE, Bermingham E, Wood C (2001) Plumage and mitochondrial DNA haplotype variation across a moving hybrid zone. *Evolution*, **55**, 405–422.

Ruddiman WF, Raymo M, McIntyre A (1986) Matuyama 41,000-year cycles: North Atlantic Ocean and northern hemisphere ice sheets. *Earth and Planetary Science Letters*, **80**, 117–129.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.

Sætre G, Sæther SA (2010) Ecology and genetics of speciation in *Ficedula* flycatchers. *Molecular Ecology*, **19**, 1091–1106.

Sætre G, Moum T, Bureš S *et al.* (1997) A sexually selected character displacement in flycatchers reinforces premating isolation. *Nature*, **387**, 589–592.

Sangster G, Collinson M, Helbig AJ *et al.* (2001) The taxonomic status of green-winged teal *Anas carolinensis*. *British Birds*, **94**, 218–226.

Sangster G, Knox AG, Helbig AJ, Parkin DT (2002) Taxonomic recommendations for European birds. *Ibis*, **144**, 153–159.

Scribner KT, Petersen MR, Fields RL *et al.* (2001) Sex-biased gene flow in spectacled eiders (Anatidae): inferences from molecular markers with contrasting modes of inheritance. *Evolution*, **55**, 2105–2115.

Sonsthagen SA, Talbot SL, Scribner KT, McCracken KG (2011) Multilocus phylogeography and population structure of common eiders breeding in North America and Scandinavia. *Journal of Biogeography*, **38**, 1368–1380.

Sorenson MD, Fleischer RC (1996) Multiple independent transpositions of mitochondrial DNA control region sequences to the nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 15239–15243.

Sorenson MD, Ast JC, Dimcheff DE *et al.* (1999) Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Molecular Phylogenetics and Evolution*, **12**, 105–114.

Spottiswoode CN, Stryjewski KF, Quader S *et al.* (2011) Ancient host specificity within a single species of brood parasitic bird. *Proceedings of the National Academy of Sciences of the United States of America*, **43**, 17738–17742.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Strasburg JL, Rieseberg LH (2011) Interpreting the estimated timing of migration events between hybridizing species. *Molecular Ecology*, **20**, 2353–2366.

Tubaro PL, Lijtmaer DA (2002) Hybridization patterns and the evolution of reproductive isolation in ducks. *Biological Journal of the Linnean Society*, **77**, 193–200.

Weir JT, Schluter D (2008) Calibrating the avian molecular clock. *Molecular Ecology*, **17**, 2321–2328.

White MJD (1978) *Modes of Speciation*, W. H. Freeman and Company, San Francisco.
Wiens J (2007) Species delimitation: new approaches for discovering diversity. *Systematic Biology*, **56**, 875–878.
Winker K (2009) Reuniting phenotype and genotype in biodiversity research. *BioScience*, **59**, 657–665.
Winker K (2010) On the origin of species through heteropatric differentiation: a review and a model of speciation in migratory animals. *Ornithological Monographs*, **69**, 1–30.
Winker K, Gibson DD (2010) The Asia-to-America influx of avian influenza wild bird hosts is large. *Avian Diseases*, **54**, 477–482.
Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, **23**, 1851–1853.
Wu C (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.
Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107–2121.
Zink RM, Rohwer S, Andreev A, Dittmann D (1995) Trans-Beringia comparisons of mitochondrial DNA differentiation in birds. *Condor*, **97**, 639–649.

J.L.P. studies the genomic signatures of population history and selection in ducks. K.G.M. studies population genetics and adaptation, especially in waterfowl. C.L.P. focuses on avian ecology and genetics. S.R. studies life histories with an emphasis on color and molt cycles. S.V.D. focuses in the ares of evolutionary ecology, genetics of speciation, and biogeography. Y.N.Z. has research interests in large-scale mechanisms of evolution. I.K. specializes in phylogeography and population genetics. D.D.G. is a recent retiree from the University of Alaska Museum (Bird Collection Manager). K.W. is an evolutionary biologist interested in the processes of organismal divergence and speciation.

Data accessibility

DNA sequences: GenBank accessions JX137688–JX138231.

Other data files (e.g. FASTA files; IM, LAMARC, STRUCTURE and ARLEQUIN input files; Table of GenBank accession numbers): Dryad accession doi: 10.6061/dryad.2f1k3.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Bayesian skyline plots for mtDNA, reconstructing historical demographics for the New World (n = 86; shaded plot) and Old World (n = 58; unshaded plot) clades.

Table S1 Specimen data for samples of *Anas crecca crecca* and *A. c. carolinensis* sequenced in this study.

Table S2 Joint estimates of Θ and recombination rates for nuclear loci.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.