MOLECULAR ECOLOGY

Molecular Ecology (2012) doi: 10.1111/mec.12098

# Global lack of flyway structure in a cosmopolitan bird revealed by a genome wide survey of single nucleotide polymorphisms

ROBERT H. S. KRAUS,*† PIM VAN HOOFT,* HENDRIK-JAN MEGENS,‡ ARSENY TSVEY,§ SERGEI Y. FOKIN,¶ RONALD C. YDENBERG*,** and HERBERT H. T. PRINS*

*Resource Ecology Group, Wageningen University, PO Box 47, 6700 AA, Wageningen, The Netherlands, †Conservation Genetics Group, Senckenberg Research Institute and Natural History Museum, Clamecystr. 12, D-63571, Gelnhausen, Germany, ‡Animal Breeding and Genomics Centre, Wageningen University, De Elst 1, Wageningen, 6708 WD, The Netherlands, §Biological Station Rybachy of the Zoological Institute RAS, 238535, Kaliningrad Region, Russia, ¶Hunting Ornithology Section, State Information-Analytical Centre of Game Management and Environment, Ministry of Environment of Russia, Moscow, Russia, **Centre for Wildlife Ecology, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6

## Abstract

**Knowledge about population structure and connectivity of waterfowl species, especially mallards (*Anas platyrhynchos*), is a priority because of recent outbreaks of avian influenza. Ringing studies that trace large-scale movement patterns have to date been unable to detect clearly delineated mallard populations. We employed 363 single nucleotide polymorphism markers in combination with population genetics and phylogeographical approaches to conduct a population genomic test of panmixia in 801 mallards from 45 locations worldwide. Basic population genetic and phylogenetic methods suggest no or very little population structure on continental scales. Nor could individual-based structuring algorithms discern geographical structuring. Model-based coalescent analyses for testing models of population structure pointed to strong genetic connectivity among the world's mallard population. These diverse approaches all support the conclusion that there is a lack of clear population structure, suggesting that the world's mallards, perhaps with minor exceptions, form a single large, mainly interbreeding population.**

*Keywords*: *Anas platyrhynchos*, ascertainment bias, coalescent analysis, conservation management, population genomics, single nucleotide polymorphisms

*Received 12 January 2012; revision received 19 September 2012; accepted 20 September 2012*

## Introduction

One of the applications of molecular ecology is the study of geographical genetic structure of species and populations: phylogeography (Avise 1987). Newly developed methods in statistical phylogeography (Knowles 2004; Carstens & Richards 2007; Nielsen & Beaumont 2009; Beaumont *et al.* 2010; Bloomquist *et al.* 2010; Templeton 2010) based on the (structured) coalescent (Kingman 1982a,b; Beerli & Felsenstein 2001) have good ability to explicitly model demographical quantities such as effective population sizes and between-population migration rates (e.g. immigration and emigration across generations in the population genetics sense—as opposed to seasonal migration). Model-based approaches towards the structured coalescent (Nielsen & Beaumont 2009; Beaumont *et al.* 2010) are among the most promising methods to infer complex and realistic phylogeographical and demographical scenarios on a population level.

For coalescent approaches, accuracy increases with increasing density of genetic markers (Kuhner 2006; Wang & Hey 2010). Single nucleotide polymorphisms (SNPs) are abundant and widespread in genomes, evolve

Correspondence: Robert H. S. Kraus, Fax: +0049 6051 61954 3118; E-mail: robert.kraus@senckenberg.de

in a manner that is well described by simple mutation models (Vignal *et al.* 2002) and therefore offer advantages as markers over mitochondrial DNA (mtDNA) and microsatellites (Estoup *et al.* 2002; Morin *et al.* 2004; Selkoe & Toonen 2006).

The term 'population genomics' was coined a decade ago to describe the use of many independent genetic markers across all regions of the genome (Black IV *et al.* 2001; Luikart *et al.* 2003). SNP sets containing more than 100 markers have been used in human (Altshuler *et al.* 2000) and other model organism studies for some time, but few studies using SNPs have been carried out on nonmodel organisms, and most used fewer than 100 loci (Kovach *et al.* 2010; Campbell & Narum 2011; Mesnick *et al.* 2011; Sacks *et al.* 2011). A recent boom in sequencing technology has developed assay panels with hundreds or even thousands of SNPs (Kerstens *et al.* 2009; Jonker *et al.* 2012; Kraus *et al.* 2011a) and has enabled molecular ecological and conservation studies on nonmodel organisms to complement traditional approaches (Santure *et al.* 2010; Willing *et al.* 2010; Jonker *et al.* 2011a; Williams & Oleksiak 2011).

The migration systems of waterfowl have been extensively studied by ringing, telemetry, morphometrics, radar tracking and isotope analysis. In general, seasonal migration routes run between northern breeding and southern nonbreeding areas, and both in North America and in Eurasia, migration routes have for management purposes been divided into geographically distinct 'flyways' (Anon http://www.flyways.us/flyways/info#flyways-bio; Scott & Rose 1996; Miyabayashi & Mundkur 1999). Many 'irregularities' in migration routes have been described, especially in duck species. For example, individuals often switch migratory routes, a phenomenon termed 'abmigration' (Thompson 1931) or 'flyway permeability' (Guillemain *et al.* 2005).

The mallard (*Anas platyrhynchos* L.; Anseriformes: Anatidae) has a Holarctic distribution and is the most numerous waterfowl species. Extensive analyses using traditional, that is nongenetic, approaches show no clear population structure (Scott & Rose 1996). Northern breeding birds are mostly migratory, wintering south, while birds breeding in temperate regions, especially in parts of Western Europe, can be resident (Scott & Rose 1996). Migratory mallards may travel thousands of kilometres between breeding and nonbreeding locales (Kulikova *et al.* 2005). On a continental scale, so-called clade A mtDNA haplotypes predominate in Eurasia and clade B haplotypes in North America (Avise *et al.* 1990; Kulikova *et al.* 2005; Kraus *et al.* 2011b). There are as yet no large-scale data on nuclear markers, but within continents, no mitochondrial genetic structuring has been detected (Kraus *et al.* 2011b) because mallards disperse easily across predefined flyway boundaries (Gunnarsson *et al.* 2012). These results lead to the suggestion that, at least on the continental scale, mallards constitute single large panmictic populations.

In this study, we employ a SNP marker set comprising hundreds of loci to conduct a population genomic test for population genetic structure in mallards. Diverse approaches, including coalescent analyses, Bayesian frameworks and model selection procedures, were used to scrutinize aspects of proposed mallard migration models.

## Methods

### *Sampling*

Mallard blood from 801 individuals from 45 localities throughout all of the mallard's native range on three continents was collected on FTA cards (Smith & Burgoyne 2004). Most samples were contributed by hunters, but in the Faroe Islands, France, Greenland, Iran, Norway, Portugal, Sweden and Alaska mallards were trapped, blood drawn from the wing or foot vein and released. All sampling procedures were approved by the animal ethical committee of Wageningen University, as well as the appropriate local authorities. Sampling localities are abbreviated here by a four letter code: letters one and two represent the ISO code of the country (ISO3166 2007), and letters three and four represent the locality (e.g. DEWU: 'DE' for Germany and 'WU' for 'Wunsiedel'). Full details on sampling localities and samples can be found in Fig. 1 and Table 1 and the Appendix S1 (Supporting information), 'sample-details.xls'.

### *DNA isolation, SNP genotyping and descriptive statistics*

We isolated DNA and genotyped SNPs using the Illumina GoldenGate Genotyping assay on the Illumina BeadXpress (Kraus *et al.* 2011a). For each mallard, we screened SNP genotypes across 384 SNPs (accession numbers ss263068950–ss263069333 in dbSNP; Sherry *et al.* 2001). Raw data were analysed in GenomeStudio (Illumina Inc.), in which 363 SNPs were scored as polymorphic. The SNP set contained SNPs on nearly all chromosomes of the mallard as inferred from their mapping positions in the chicken genome (Kraus *et al.* 2011a). SNPs did not show significant departures from neutrality or linkage disequilibrium (Kraus *et al.* 2012). Further, we performed individual-based isolation-by-distance analyses for each continent in GENALEX (Peakall & Smouse 2006), version 6.41, as per manual instructions and with 9999 permutations in the Mantel tests.


**Fig. 1** Geographical overview of sampling localities (triangles). Some localities are too close to be accurately labelled, and these include locality NOSS in northern Europe (flyway EU-NW) and EETA in the Baltic as well as three Russian localities (RUNO, RUTV and RUYA) near Moscow; the latter four localities being part of the hypothesized European west Mediterranean flyway (EU-WM). For more details on flyways hypotheses, please compare with Fig. 2 and Table 1.

### *Phylogenetic analysis*

We did not expect the data to conform to a tree-like configuration because with recombination (Posada & Crandall 2001), each of the many independent nuclear genetic markers could have a unique phylogenetic history. A programme to take this into account is NEIGHBOUR-NET (Bryant & Moulton 2004), implemented in SPLITSTREE (Huson 1998; Huson & Bryant 2006), version 4. Willing *et al.* (2010) successfully adopted this method for use with genomewide SNP data, and we used their settings. For each individual, the genotype at each SNP was collapsed into a single base character and concatenated to a sequence of 363 nucleotides. Heterozygote genotypes were represented by IUPAC nucleotide ambiguity codes and missing data denoted 'N'.

### *Population assignments*

One of the most widely used programmes to determine the number of genetic clusters and assign individuals to them is STRUCTURE (Pritchard *et al.* 2000). To avoid determination of spurious genetic clusters because of close relatives, we first identified closely related individuals in our data set. Separately for each sampling locality, and only using genotype data from these individual localities, we assessed pairwise relatedness ($r$) with COANCESTRY (Wang 2011), version 1. As an estimator for $r$ we chose the dyadic maximum-likelihood estimator of Milligan (2003) because it produced the best correlations with known $r$ values in a simulation study with bi-allelic SNPs in a similarly sized SNP set (unpublished data). Ninety-five per cent confidence intervals (CIs) were calculated by 1000 bootstraps. If the lower bound of the 95% CI was >0.2, one of the two individuals of the tested pair from the Structure analysis was excluded because of potential half-sib relationship. The most likely number of genetic clusters (value of $K$) was determined according to the study by Evanno *et al.* (2005), using STRUCTURE version 2.3.3 (10 replicates for all values of $K$ from 1 to 20, that is, twice the number of flyways in our population model, for 1 000 000 steps of which the first 200 000 were discarded as burn-in).

Discriminant analysis of principal components (DAPC; Jombart *et al.* 2010) from *adegenet* (Jombart 2008) version 1.2.8 in R (R Development Core Team 2009) was used in addition to STRUCTURE to detect the number of genetic clusters and assignment of individuals. In contrast to STRUCTURE, DAPC does not suffer from the assumption of unrelatedness, and therefore,

**Table 1** Sampling localities (also see main text for explanation and Appendix S1, Supporting information for more info), sample sizes (*n*), pooling strategy of localities into hypothesized flyways and genetic marker performance

| Flyway | *n* | Mono. Loci* | Locality | *n* | Mono. Loci* | Lat† | Long† |
|---|---|---|---|---|---|---|---|
| NA-Pacific | 22 | 14 | USMF | 22 | 14 | 64.9 | −148.9 |
| NA-Central | 22 | 17 | CARM | 20 | 18 | 50.628 | −101.159 |
| | | | CASL | 2 | — | 49.542† | −112.056† |
| NA-Atlantic | 16 | 13 | CACO | 4 | — | 45.58† | −63.845† |
| | | | CAEK | 4 | — | 44.736 | −75.969 |
| | | | CAJC | 1 | — | 42.324 | −82.314 |
| | | | CALM | 7 | — | 43.962 | −80.4 |
| Greenland | 29 | 20 | GLIS | 9 | — | 67.1 | −50 |
| | | | GLNU | 20 | 75 | 64.19 | −51.708 |
| EU-NW | 209 | 1 | FIOU | 19 | 10 | 65.057† | 25.197† |
| | | | FOTO | 24 | 8 | 62.02 | −6.78 |
| | | | GBAB | 20 | 6 | 57.433 | −2.393 |
| | | | GBFE | 11 | 20 | 55.901 | −3.061 |
| | | | GBNM | 20 | 8 | 51.712† | −1.433† |
| | | | ISHV | 4 | — | 63.748 | −20.239 |
| | | | NLFR | 32 | 5 | 53.035 | 5.574 |
| | | | NOBE | 32 | 10 | 60.35† | 5.323† |
| | | | NOSS | 16 | 8 | 58.856† | 7.332† |
| | | | SEOB | 31 | 13 | 56.2 | 16.4 |
| EU-WM | 360 | 1 | ATHO | 25 | 9 | 48.615† | 24.625† |
| | | | DEWU | 27 | 4 | 50.042† | 11.78† |
| | | | EETA | 22 | 18 | 58.324† | 27.178† |
| | | | FRAL | 10 | 27 | 48.789 | 8.019 |
| | | | FRMV | 32 | 7 | 43.55 | 4.733 |
| | | | LTVE | 17 | 8 | 55.342 | 21.192 |
| | | | PTDJ | 32 | 8 | 40.664 | −8.732 |
| | | | RUIV | 27 | 8 | 56.47† | 41.37† |
| | | | RULE | 31 | 14 | 59.65 | 28.35 |
| | | | RUNO | 8 | — | 58.167 | 31.517 |
| | | | RUTV | 19 | 15 | 57.81† | 36.528† |
| | | | RUVL | 29 | 9 | 55.884† | 39.218† |
| | | | RUVO | 31 | 4 | 59.498† | 37.511† |
| | | | RUYA | 25 | 16 | 56.319† | 39.041† |
| | | | SILJ | 19 | 8 | 46.17 | 14.69 |
| | | | UADU | 3 | — | 51.565 | 26.573 |
| | | | UALV | 3 | — | 49.825† | 23.573† |
| EU-BS/EM | 21 | 13 | CYLA | 5 | — | 34.883 | 33.622 |
| | | | GREV | 16 | 18 | 40.86 | 25.89 |
| Asia-SW | 15 | 14 | IRUK | 15 | 14 | 36 | 51 |
| Asia-Central | 51 | 9 | PKHA | 17 | — | 34.001 | 72.934 |
| | | | RUOM | 12 | 41 | 55.961† | 73.311† |
| | | | RUTO | 32 | 13 | 56.526† | 83.349† |
| Asia-East | 12 | 28 | CNLI | 5 | — | 28.563 | 115.943 |
| | | | RUKH | 7 | — | 52.937† | 138.941† |

NA, North America; EU-NW, north-western Europe; EU-WM, western Mediterranean Europe; EU-BS/EM, Black Sea/eastern Mediterranean Europe; Asia-SW, south-western Asia.

Lat/Long are decimal GPS coordinates of the sampling localities.

*Genetic marker performance (localities with <10 sampled individuals not counted): amount of SNP loci that were monomorphic in all sampled individuals of the flyway (first occurrence in table)/locality (second occurrence); for example, only one allele appeared in the pool of individuals genotypes.

†Some coordinates are averages of several near-by places, where ducks have been sampled and combined into one sampling locality.

potentially closely related individuals can be included in this analysis. Using the function *find.clusters*, we determined the most likely number of genetic clusters in the data, using all available principal components (PCs). To calculate the probability of assignment of individuals to each of these clusters using DAPC, we

determined the optimal number of PCs. As advised in the manual, to avoid unstable assignments of individuals to clusters, we retained only 242 PCs (sample size divided by three), but used all discriminant functions, in a preliminary DAPC run. The results were then reiterated by the *optim.a.score* function with 25 simulations to determine the optimal number of PCs, and a final DAPC was subsequently carried out with the optimal number of PCs.

### *Migration modelling*

We used the coalescent-based program MIGRATE-N (Beerli & Felsenstein 1999, 2001) to estimate population parameters of our hypothesized flyways. MIGRATE-N calculates the marginal likelihoods for each model (Kass & Raftery 1995; Beerli & Palczewski 2010). These can be used to evaluate multiple models when based on the same data using Bayes factors (Bloomquist *et al.* 2010). Information from all loci was combined into a global estimate by Bezier approximation of the thermodynamic scores. The probability of a certain model is then retrieved by dividing the exponentiated (on the base of *e*) log likelihoods by the sum of all exponentiated log likelihoods (Beerli 2010).

We developed a number of models (Fig. 2) to investigate the global extent of population genetic structure in mallards and to investigate possible restrictions in the directionality of gene flow. Sampling localities were aggregated in various ways based on alleged mallard-specific flyways in Europe (Scott & Rose 1996), Asia (Miyabayashi & Mundkur 1999) and North America (Anon, http://www.flyways.us/flyways/info#flyways-bio; Table 1). The two localities from Greenland were classified as a separate flyway/population unit (Scott & Rose 1996).

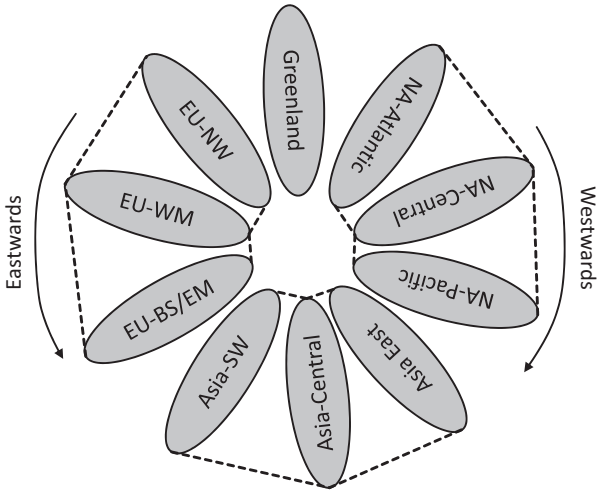


**Fig. 2** Hypothesized population structuring scheme, with flyways as basal units. As viewed from above the North Pole, dashed lines join several neighbouring flyways into 'land masses'. Abbreviations of flyways: NA, North Atlantic; EU-BS/EM, European Black Sea/eastern Mediterranean flyway; EU-WM, European west Mediterranean flyway; EU-NW, European north-western flyway.

In model group 1 (models 1A–C), each flyway is regarded as a potential population unit. In model 1A, gene flow between all possible pairs of population units is permitted, while in model 1B, only geographically neighbouring flyways are connected by gene flow, and in model 1C, this is possible only in an eastward direction, because of a hypothesized influence of the westerly winds on bird flight (Lamb 1975; Liechti 2006; Kraus *et al.* 2011b). Model 2 is defined by the complete absence of global population structure: all sampled individuals are hypothesized to belong to the same population. In model group 3, existence of an Eurasian, a North American and a Greenland population is hypothesized. Model 3A allows gene flow between all possible pairs, model 3B allows gene flow only in an eastward direction, and model 3C isolates Greenland. Model group 4 encompasses situations in which Greenland mallards are either part of a North American population (models 4A1 and A2) or a Eurasian one (4B1 and B2). The difference between those two possibilities (4A1/4A2 and 4B1/4B2) lies in the fact that models 4A1 and 4B1 are full migration models (migration may occur in both directions), and models 4A2 and 4B2 only allow eastward gene flow from Europe to North America. Finally, we also set up model 5, which comprises population assignments as inferred by DAPC (see above), and all gene flow directionalities were permitted.

For the MIGRATE-N analysis of each model, we used Bayesian inference in version 3.2.14. The data type was specified as SNP. Starting values for $\Theta$ and $M$ were calculated from Wright's $F_{ST}$ as implemented in MIGRATE-N. The input data were defined as finite sites nucleotide data, and we calculated the transition/transversion ratio (2.76) as well as nucleotide frequencies (A: 0.445355, C: 0.104690, G: 0.416658 and T: 0.033297) from the data and supplied them to MIGRATE-N as constants. Mutation rates were set to be constant among all loci. The priors for $\Theta$ and $M$ were uniform 0–0.1 and 0–15 000, respectively. These settings performed best in preliminary runs. Along the Markov chain, the slice sampler option was used. After a burn-in of 2 000 000 steps, we sampled 25 000 states from a single Markov chain, one every 20 steps. Four chains were run in parallel, with heating terms '1', '1.5', '3' and '10 000'. The estimated mutation scaled migration parameter $M$ was translated into the effective number of immigrants per generation ($Nm$) by multiplying with $\Theta$ and dividing by four (the SNPs are diploid and biparentally inherited): $Nm = \Theta_i \times M_{j \to i}/4$.

## Results

### *Genotyping and basic statistics*

Virtually all genotyped SNPs were polymorphic within the flyways from which we obtained good sample sizes (European north-west flyway and European western Mediterranean flyway). On flyways with smaller sample sizes, a few loci were monomorphic (a maximum of 28 in the East Asian flyway). The same applied within sampling localities. Localities with less than ten sampled individuals were excluded. Table 1 lists all details in the columns 'mono. loci'.

Genetic differentiation on the continental scale was very low. $F_{ST}$ between Eurasian and North American samples was only 0.006, but significant ($P < 0.001$), whereas $F_{ST}$ between Greenland and these former two geographical units was an order of magnitude greater ($F_{ST} = 0.073$ with Eurasia, $F_{ST} = 0.079$ with North America; both statistically significant, $P < 0.001$). Also on a flyway level, the Greenland population stood out as being most differentiated among all flyways with significant $F_{ST}$ values around 0.1 with all other flyways. Most other flyway comparisons did not display significant differentiation, and those that did (mainly involving Asian flyways) were much lower in magnitude (Table 2). Within land masses, only north-western Europe was significantly differentiated from other flyways. Differentiation between all other flyways within Eurasia or North America was insignificant.

No isolation-by-distance pattern was detectable. The slope of the regression line in the Mantel test (correlation of geographical and genetic distance) was basically flat, no matter which data subset we chose, and not significantly different from zero: Eurasia: $P = 0.17$, $R^2 = 0.0008$; and individually, Asia: $P = 0.21$, $R^2 = 0.0045$, Europe: $P = 0.24$, $R^2 = 0.0002$ and North America: $P = 0.06$, $R^2 = 0.0049$.

### *Phylogenetic network*

If populations were differentiated from each other, the NEIGHBOUR-NET algorithm would display reticulate relationships more densely within less differentiated groups and less densely in more differentiated groups (cf. Fig. 2 of Willing *et al.* 2010). In contrast, the network obtained from our data did not indicate any population genetic structure; it resembled a bush rather than an unrooted tree, with complex reticulations remaining unresolved up towards the tips (Fig. 3) and no grouping could be distinguished. The sole irregularities were two discernable spikes: (i) one at the bottom in Fig. 3 and (ii) a smaller one in the top right corner. Both groups represented individuals from apparently unrelated regions: Russia, Portugal, Ukraine, Norway, Alaska, Estonia, Iran and Canada for (i), and the Faroe Islands, Russia, Slovenia, Iran and Great Britain for (ii).

### *Population genetic clustering*

We excluded 135 individuals from the STRUCTURE analysis, because of suspected close familial relatedness. We analysed models with 1–20 genetic clusters. $\Delta K$, the statistic to detect the most likely value of $K$, clearly peaks at $K = 2$. Hence, the best supported model according to the study by Evanno *et al.* (2005) was a model with two genetic clusters (Fig. 4). Unfortunately, the $\Delta K$ method cannot evaluate a model of full panmixia in which $K$ is 1.

The posterior probability of assignment of individuals to the two inferred clusters was intermediate for the great majority of the individuals. No individual in the whole data set could be assigned to one of the two

**Table 2** Pairwise $F_{ST}$ values for all flyways

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| US-Pacific | — | | | * | * | | * | * | | * |
| US-Central | 0.006 | — | | * | * | | | * | | * |
| US-Atlantic | 0.011 | 0.008 | — | * | | | | | | |
| Greenland | **0.086** | **0.091** | **0.080** | — | * | * | * | * | * | * |
| EU-NW | **0.011** | **0.012** | 0.008 | **0.076** | — | * | * | * | * | * |
| EU-WM | 0.008 | 0.009 | 0.007 | **0.073** | **0.003** | — | | | | |
| EU-BS/EM | **0.013** | 0.012 | 0.005 | **0.085** | **0.009** | 0.005 | — | | | |
| Asia-SW | **0.043** | **0.046** | 0.026 | **0.112** | **0.031** | 0.025 | 0.024 | — | | |
| Asia-Central | 0.001 | 0.012 | 0.009 | **0.082** | **0.006** | 0.004 | 0.007 | 0.027 | — | |
| Asia-East | **0.029** | **0.028** | 0.016 | **0.097** | **0.019** | 0.017 | 0.015 | 0.030 | 0.016 | — |

EU-NW, north-western Europe; EU-WM, western Mediterranean Europe; EU-BS/EM, Black Sea/eastern Mediterranean Europe; Asia-SW, south-western Asia.

Above the diagonal, the statistical significance ($P < 0.05$) after Bonferroni correction is indicated; below the diagonal, $F_{ST}$ values are shown, with statically significant values printed boldface for clarity.

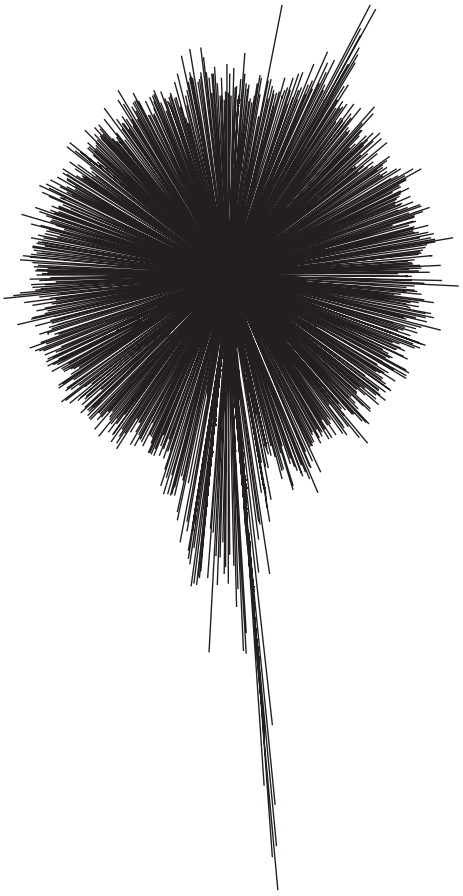

**Fig. 3** Phylogenetic network generated in SPLITSTREE (Huson 1998; Huson & Bryant 2006). Node labels are omitted for clarity. No genetic groups can be detected (cf. Fig. 2 in Willing *et al*. 2010, for an example of clear grouping).

clusters with more than 86% posterior probability. Moreover, most individuals (521 of 666) were assigned to their genetic cluster with 60% or less posterior probability. When $K = 3$ (as inferred by DAPC, see below), the situation was essentially the same, but individuals from Greenland formed a separate cluster. The highest values for $L(K)$ were observed for $K = 9$ and $K = 10$ (not significantly different from each other, $P = 0.5$, $t$-test). In both these runs (and all other values of $K$), the Greenland individuals always formed the only cluster in which individuals were not admixed by more or less equal proportions from all other clusters.

The DAPC gave most support to a structure with three clusters (Fig. 5). Figure 6 shows a plot of the first two PCs calculated from the data. Most individuals were assigned to their genetic cluster with high probability. Only 29 individuals had assignment probabilities of <0.9 to their respective cluster. The most separate of the three clusters was composed entirely of Greenland individuals. In fact, only a single sample collected in Greenland (GLIS002) was not assigned with high posterior probability to the Greenland cluster. The other two clusters lay close to each other. Just 69 individuals from northern Europe constituted the first of these clusters, while the second contained 659 individuals from all flyways except Greenland. When Greenland was excluded from the analysis, essentially the same clustering was obtained.

### *Migration model selection*

The evaluation of all tested models clearly gave most support to model 1A, in which each flyway constituted a population and all flyways exchange migrants (i.e. there is gene flow; Table 3). The long burn-in period used resulted in good convergence and narrow posterior density peaks for migration parameters (Table 4). Estimates of $\Theta$ were less precise. The 95% posterior densities spanned an order of magnitude in all flyways except for north-western Europe, western Mediterranean Europe and Central Asia (i.e. those flyways with largest sample sizes, $n = 209$, 360 and 51, respectively). In the flyway with the smallest sample size, East Asia ($n = 12$), it even spanned three orders of magnitude (see Appendix S2, Supporting information, 'mallard-flyways-model-1A.pdf'). However, the mode and mean of all estimates of $\Theta$ were nearly identical, and the density distributions were symmetrical. For calculating $Nm$ (Table 4), we thus used the modes of the $\Theta$ distributions. Migration rates among flyways were mostly even among pairwise comparisons. Only emigration from north-western Europe and western Mediterranean Europe to all other flyways was higher, as well as immigration into these two flyways from Central Asia.

## Discussion

### *Absence of strong population structure*

In this study, we employed basic population genetic techniques, individual-based genetic clustering algorithms and coalescent-based demographical modelling with subsequent model selection based on a data set with 363 SNP markers distributed across the entire genome of the mallard. Samples from throughout the native range of the mallard were analysed to test the hypothesis of panmixia, which has been proposed previously (Kulikova *et al*. 2005; Kraus *et al*. 2011b), on several geographical scales in the mallard. *F*-statistics indicated hardly any genetic structure within continents. Only Greenland is genetically differentiated from the remaining mallard population, and to a lesser extent, slight differentiation is observed between flyways in Europe and North America. However, significant $F_{ST}$ may not necessarily be a good indicator for real population

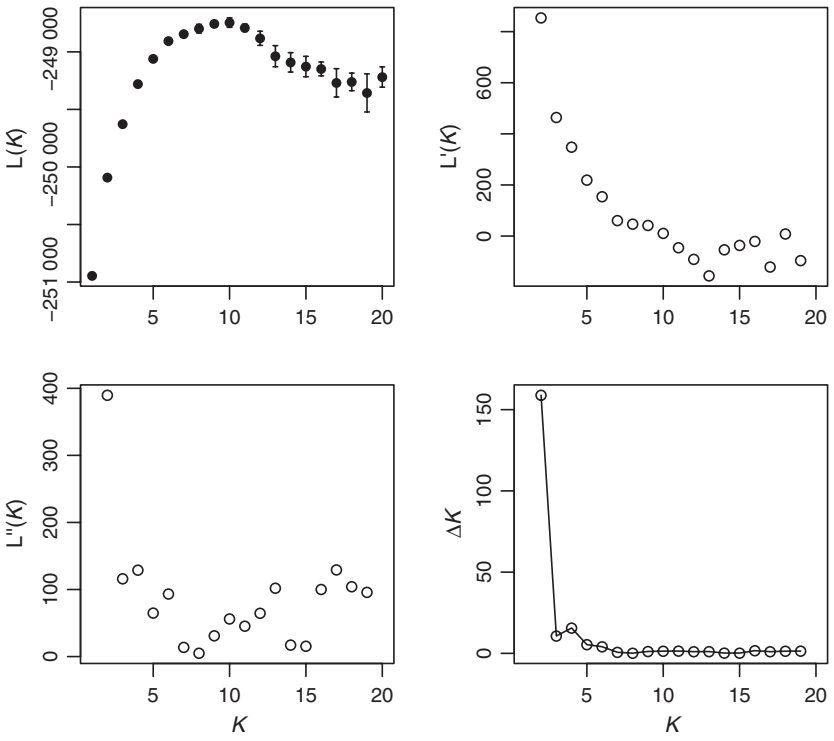


**Fig. 4** Posterior likelihood [$L(K)$] values from several Structure runs with different $K$. Top left panel shows $L(K)$ means from 10 independent runs (error bars are SD). $\Delta K$ (bottom right panel) is based on the first- and second-order rates of change $L'(K)$ and $L''(K)$, based on the Evanno method (Evanno *et al.* 2005). The highest level hierarchical structure in the data suggests two genetic clusters (see bottom right panel). For details, see Methods section.

structure when low, as was found in a recent study of another duck species, the common pochard (*Aythya ferina*; Liu *et al.* 2011). Testing for isolation by distance revealed no relationship between genetic and geographical distances, which confirms previous findings of a lack of structure within continents (Kraus *et al.* 2011b). A phylogenetic network method (SPLITSTREE) was unable to resolve the complex reticulate structure of the worldwide mallard population structure. A model selection procedure to infer the number of genetic clusters using STRUCTURE resulted in the best support for the model in which the number of genetic clusters was set to two, but examination of the posterior assignment probabilities indicates that STRUCTURE was not able to assign the individuals into these clusters with high probability, a pattern also observed for models assuming more genetic clusters. Hence, this method also failed to detect significant substructuring of the global mallard gene pool, except that it consistently placed individuals from Greenland as a separate group.

Discriminant analysis of principal components (Jombart *et al.* 2010) also showed that Greenland mallards are most differentiated and further implied that some individuals from populations in northern Europe form a genetic cluster different from the main global population. This finding confirms that some European mallards may be genetically different from the remaining population (Kraus *et al.* 2011b), because of, for instance, resident lifestyle (Scott & Rose 1996) or the impact of releases of farmed mallards for hunting purposes (Laikre *et al.* 2006; Champagnon *et al.* 2009).

Finally, the model selection approach with MIGRATE-N gave most support to the full flyway model (model 1), over the global (model 2) and continental (model groups 3 and 4) panmixia models. This is surprising because not only our own findings from phylogenetic and population genetic analyses and individual-based clustering but also previous mtDNA studies indicate panmixia at least on the continental scale (Kulikova *et al.* 2005; Kraus *et al.* 2011b). One has to bear in mind, though, that the outcome of MIGRATE-N alone should be interpreted with some care because our data potentially violate the assumption of migration–drift equilibrium in MIGRATE-N's algorithm. This equilibrium is especially hard to reach in large populations, such as in mallards that effectively number in the millions (Kraus *et al.* 2011b, 2012). In addition, although migration rates between any two populations are low, summing the total number of immigrants coming from all populations would be quite high and consistent with strong connectivity across the Northern Hemisphere.

### *Is the global mallard population panmictic?*

The low population genetic migration rates between proposed flyways when inferred by MIGRATE-N formally contradict panmixia. However, except in Europe, all populations appear to exchange migrants, and in fairly

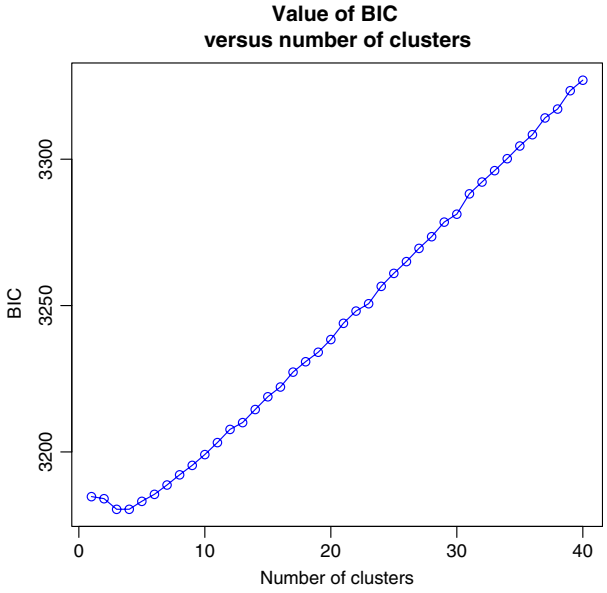


**Fig. 5** Inference of the number of genetic clusters by discriminant analysis of principal components (DAPC). The Bayesian information criterion (BIC) as calculated during the *find.clusters* function of the DAPC package infers a most likely number of clusters when it is minimized. The lowest BIC values are found with three and four clusters. When two numbers of clusters have equal BICs, the smaller one (here, number of clusters = 3) is usually the correct one.

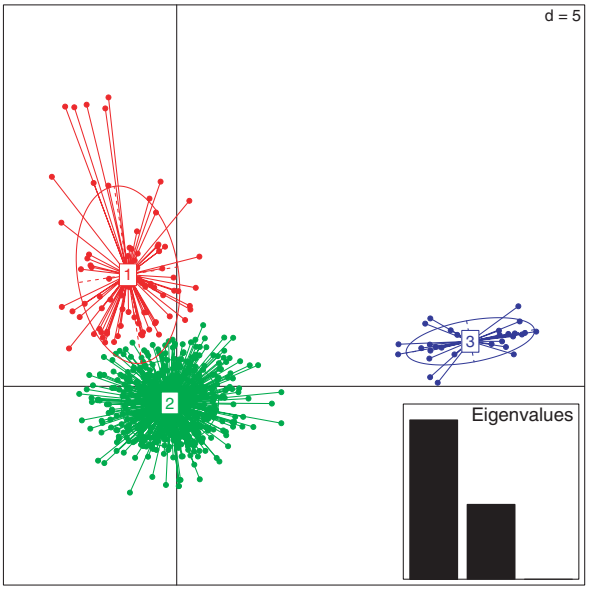


**Fig. 6** Principal component (PC) scatter plot. Samples are assigned to their genetic cluster by discriminant analysis of PCs analysis. The bar graph inset displays the variance explained by the two discriminant eigenvalues used for plotting. The 67% inertia ellipses are drawn for each cluster, representing the variance of both PCs. Cluster *Three* is composed of samples from Greenland, cluster *Two* of some samples from the flyways EU-NW and EU-WM, and cluster *One* contains samples from all flyways.

**Table 3** Details of model selection procedure in MIGRATE-N

| Model | Marginal likelihood | Delta | $e^{\text{delta}}$ | Probability* |
|---|---|---|---|---|
| 1A | −205825.02 | 0 | 1 | 1 |
| 1B | −211028.08 | −5203.06 | 0 | 0 |
| 1C | −211568.54 | −5743.52 | 0 | 0 |
| 5 | −231609.00 | −25783.98 | 0 | 0 |
| 3A | −232656.77 | −26831.75 | 0 | 0 |
| 3B | −232745.53 | −26920.51 | 0 | 0 |
| 4A2 | −233959.12 | −28134.10 | 0 | 0 |
| 4A1 | −234534.44 | −28709.42 | 0 | 0 |
| 4B2 | −236911.44 | −31086.42 | 0 | 0 |
| 4B1 | −237231.09 | −31406.07 | 0 | 0 |
| 2 | −242982.20 | −37157.18 | 0 | 0 |
| 3C | −245189.81 | −39364.79 | 0 | 0 |

Models are ranked by their marginal likelihoods as obtained by Bezier approximation. Differences between each alternative model and model with highest rank (1A) are in column delta. Exponentiated model differences (column $e^{\text{delta}}$) are not presented with full precision because the values are so small that they are essentially zero (e.g. $2.2 \times 10^{-2260}$ for the second best model, 1B).

*Model probability calculated by dividing $e^{\text{delta}}$ by the sum of all $e^{\text{delta}}$.

equal magnitude, so geographical proximity has no relation to genetic proximity (also see isolation-by-distance analysis). A recent study of Mediterranean flamingo colonies comes to the same conclusion using MIGRATE-N when testing for panmixia in their system as well, with migration estimates between their defined populations quantitatively very similar to our values for mallard (Geraci *et al.* 2012).

Selecting a full flyway model over reduced flyway models in which migration is only possible between neighbours further indicates strong gene pool connectivity. This is not quite the same as panmixia because mating is not random (still locally biased), but gene flow between all locations is high enough to swamp most genetic structuring that might emerge. Flyways may have been very dynamic through the ice ages, but locally biased mating within flyways shaped the gene pool. This is probably the pattern picked-up by MIGRATE-N, which is thus compatible with a highly interconnected and dynamic population structure on the global geographical scale. Focussing on connectivity between Eurasia and North America via Beringia, our results are in line with those of other Holarctic birds. Northern pintails (*Anas acuta*), for instance, are known to exhibit low levels of nuclear genetic differentiation between Japan and California (Flint *et al.* 2009), and gadwalls (*Anas strepera*) as well as teal (*Anas crecca*) from Asia and North America share nuclear haplotypes (Peters *et al.* 2007, 2008, 2012).

**Table 4** Migration matrix for model 1A

| | NA-Pacific | NA-Central | NA-Atlantic | Greenland | EU-NW | EU-WM | EU-BS/EM | Asia-SW | Asia-Central | Asia-East |
|---|---|---|---|---|---|---|---|---|---|---|
| NA-Pacific | — | 0.44 (0.31–0.57) | 0.41 (0.27–0.55) | 0.59 (0.44–0.73) | 1.60 (1.44–1.75) | 3.14 (2.92–3.30) | 0.38 (0.26–0.50) | 0.30 (0.18–0.42) | 0.69 (0.58–0.81) | 0.44 (0.31–0.56) |
| NA-Central | 0.41 (0.28–0.54) | — | 0.42 (0.27–0.56) | 0.62 (0.49–0.75) | 1.97 (1.83–2.11) | 3.14 (2.96–3.32) | 0.54 (0.41–0.67) | 0.45 (0.30–0.59) | 0.67 (0.54–0.80) | 0.32 (0.17–0.47) |
| NA-Atlantic | 0.45 (0.33–0.57) | 0.49 (0.38–0.61) | — | 0.52 (0.39–0.77) | 1.71 (1.39–1.86) | 2.92 (2.79–3.05) | 0.53 (0.41–0.65) | 0.53 (0.42–0.65) | 0.83 (0.70–0.96) | 0.48 (0.32–0.62) |
| Greenland | 0.54 (0.38–0.70) | 0.51 (0.34–0.67) | 0.37 (0.20–0.55) | — | 2.32 (2.11–2.53) | 3.71 (3.54–3.89) | 0.56 (0.40–0.71) | 0.44 (0.24–0.63) | 0.80 (0.63–0.98) | 0.48 (0.33–0.63) |
| EU-NW | 0.85 (0.10–1.59) | 0.89 (0.14–1.63) | 0.80 (0.06–1.52) | 0.94 (0.18–1.69) | — | 13.8 (13.0–14.6) | 0.89 (0.14–1.63) | 0.78 (0.05–1.49) | 1.45 0.70–2.21) | 0.70 (0.00–1.39) |
| EU-WM | 1.03 (0.00–2.20) | 0.84 (0.00–2.03) | 0.91 (0.00–2.09) | 1.16 (0.00–2.33) | 6.97 (5.67–8.26) | — | 0.98 (0.00–2.17) | 0.81 (0.00–2.00) | 1.69 (0.39–2.98) | 0.61 (0.00–1.81) |
| EU-BS/EM | 0.48 (0.35–0.61) | 0.49 (0.36–0.62) | 0.48 (0.34–0.62) | 0.61 (0.48–0.74) | 2.03 (1.82–2.22) | 3.25 (3.08–3.41) | — | 0.42 (0.28–0.57) | 0.70 (0.55–0.85) | 0.40 (0.26–0.53) |
| Asia-SW | 0.58 (0.45–0.71) | 0.51 (0.40–0.63) | 0.42 (0.29–0.60) | 0.70 (0.53–0.84) | 1.63 (1.34–1.76) | 3.15 (3.01–3.28) | 0.49 (0.36–0.62) | — | 0.68 (0.55–0.81) | 0.33 (0.21–0.44) |
| Asia-Central | 0.52 (0.28–0.75) | 0.50 (0.26–0.75) | 0.42 (0.19–0.65) | 0.54 (0.30–0.77) | 3.10 (2.84–3.36) | 5.83 (5.54–6.11) | 0.45 (0.21–0.68) | 0.38 (0.13–0.61) | — | 0.39 (0.15–0.62) |
| Asia-East | 0.58 (0.47–0.69) | 0.59 (0.44–0.74) | 0.61 (0.50–0.71) | 0.69 (0.57–0.80) | 1.84 (1.69–1.96) | 2.72 (2.55–2.88) | 0.47 (0.36–0.59) | 0.49 (0.38–0.60) | 0.70 (0.58–0.81) | — |

NA, North America; EU-NW, north-western Europe; EU-WM, western Mediterranean Europe; EU-BS/EM, Black Sea/eastern Mediterranean Europe; Asia-SW, south-western Asia.

Immigration rates from 'column' into 'row' are given as effective numbers of immigrants (*Nm*) per generation as the mode of their posterior density function, and their low and high 95% posterior density bounds in brackets.

Few previous studies have investigated the large-scale mallard migration systems with molecular tools. A study with allozymes indicates that flyway structure resembles true population structure in North America (Rhodes *et al.* 1995). In contrast, two studies on mtDNA do not support currently delineated mallard flyways in Asia (Kulikova *et al.* 2005) or globally (Kraus *et al.* 2011b), although a clear separation between old world and new world mallard mtDNA clades exists. This difference in the largest possible scale is most likely evidence for sex-biased dispersal. However, also other possibilities, such as selection or lineage sorting, have been discussed to explain discord in multiple trans-Beringian bird species (Humphries & Winker 2011). Eventually, it is a realistic scenario to assume gene flow across the Bering Strait. Although we are only aware of a single cross-continental mallard recapture, this is evidence that movements are generally possible: in 1959, a male mallard ringed in central Canada, Saskatchewan, was found in Far East Russia, Chukchi Peninsula (Konstantin E. Litvin, Russian Ringing Centre, personal communication). Genetic data are often 'strikingly at odds with data derived by direct observational techniques' (Koenig *et al.* 1996). For example, male-biased gene flow in Siberian jays (*Perisoreus infaustus*) was proven genetically, while all observational evidence suggested otherwise (Li & Merilä 2010). Similarly, no tufted duck (*Aythya fuligula*) ringed in India was ever encountered in Europe; however, genetic data indicate high levels of genetic connectivity between these regions (Liu *et al.* 2012).

Migration model analysis formally rejects panmixia, but individual-based genetic clustering does not resolve these flyways, nor delineates alternative ones. However, the diverse approaches followed in this study support each other in their basic conclusions: the mallard population in its indigenous range, the Northern Hemisphere, is highly connected over huge distances. Therefore, perhaps with the minor exception of Greenland—as expected based on their slightly differing morphology (Scott & Rose 1996)—it forms a single large, mainly interbreeding population. While not displaying formal panmixia throughout the mallard's full range, we at least confirm the expectation (Kraus *et al.* 2011b) of continent wide lack of population structure. Finally, we wish to stress the cautionary note that our results should not be treated as a final answer to the question whether there is flyway structure. The fact that we were not able to sample only breeding mallards and thereby to capture the breeding population structure directly needs to be taken into account when designing future studies following the methods developed in this study.

Developing clearly panmictic population structure on continental or global scales is hardly possible. Geographical structure is omnipresent in nature and forms the basis of the field of phylogeography (Avise 1987). Almost always larger-scale panmixia is rejected as the null model against which data are tested. A few examples where this is not the case are known from microorganisms (Rypien *et al.* 2008), but in higher organisms, the literature is equivocal. For instance, in some marine species, panmixia was proposed. Analysis of nuclear genetic markers of white shark pointed towards panmixia (Pardini *et al.* 2001) but was questioned later (Jorgensen *et al.* 2010). Eel populations also were candidates (Han *et al.* 2010; Als *et al.* 2011). But unlike our example of a cosmopolitan duck species in these examples, panmixia seems to be achieved by natal philopatry combined with aggregated mating in only one area for the whole species. Like marine species, birds usually have good dispersal abilities, yet panmixia is hardly ever observed (Friesen *et al.* 2007; Reudink *et al.* 2011).

### *Implications for conservation and management*

Mallards are abundant across the whole world, in some places—outside their native range—even considered an invasive pest species threatening the genetic integrity of indigenous ducks through introgressive hybridization (Rhymer 2006). According to the IUCN (2011), the mallard is a species of least concern. However, recently, strong population declines have been reported locally (Eaton *et al.* 2009). Ducks are important components of wetlands, and mallards are an abundant species in this community (Elmberg 2009). The preservation of wetlands for ecosystem services relies on the functionality of this community. Our finding that mallard populations are highly genetically connected implies that local declines in mallard numbers and genetic diversity can be buffered by the continental populations, but some alterations in local conditions (e.g. introduction of farmed mallards and massive hunting) might have cascading effects.

Management of wetlands does not only have profound importance from a nature conservation point-of-view. In the last few years, the spread of zoonotic diseases such as avian influenza, via wild birds, has gained considerable attention (Olsen *et al.* 2006). Aquatic birds are the natural reservoirs of this zoonotic virus (Webster *et al.* 1992), which is transmitted via the faecal-oral route, especially among birds that live and feed on water. Avian influenza viruses have been shown to remain infectious in open water for several days, depending on environmental circumstances (Ito *et al.* 1995; Stallknecht *et al.* 2010; Lebarbenchon *et al.* 2011), and were reported from ice cores in Siberia (Zhang *et al.* 2006). Our current study suggests that dispersion of avian influenza in the wild could occur very rapidly

even between distant flyways (at least within Eurasia or North America). Therefore, management and research on wetlands is necessary to increase our ability to monitor for potential routes of avian influenza outbreaks in humans (Si *et al.* 2009).

## Acknowledgements

The following people contributed samples (in alphabetical order of the respective sample IDs): Ernst Niedermayer, Hans Jörg Damm (Stiftung Fürst Liechtenstein, Austria), Darren Hasson, Garnet Baker, Steven Evans, Thomas Kondratowicz, David Lamble, Garry Grigg, Aaron Everingham, Andrew Iwaniuk (Canada), Yan-Ling Son (Key Laboratory of Animal Ecology and Conservation Biology, Chinese Academy of Sciences), Nicolaos Kassinis (Game Fund Cyprus), Severin Wejbora (Lehr- und Forschungsrevier des Landesjagdverbandes Bayern, Germany), Urmas Võro (Estonia), Antti Paasivaara (Finnish Game and Fisheries Research Institute, Oulu, Finland), Jens Kjeld Jensen, Tróndur Leivsson (The environment agency, Faroe Islands), Mathieu Boos (NATURACONST@, Research Agency in Applied Ecology, Wilshausen, France), Matthieu Guillemain (Office National de la Chasse et de la Faune Sauvage, Arles, France), Anne Zeddeman (Laboratory for Infectious Diseases and Screening (LIS), National Institute for Public Health and the Environment (RIVM), the Netherlands), Andy Richardson (Safari in Scotland, Scotland), T. Cameron Manson (Scotland), Charles Bull (Northmore, Britain), Ruth Cromie (Wildfowl and Wetlands Trust, Britain), Apostolos Tsiompanoudis (Greece), Sasan Fereidouni (Friedrich-Loeffler-Institut, Germany), Bjorn Birgisson (The Icelandic Hunting Club), Ricardas Patapavicius, Julius Morkunas (Lithuania), Herman Postma (the Netherlands), Jan Bokdam (Nature Conservation and Plant Ecology, Wageningen University, the Netherlands), Alf Tore Mjøs (Museum Stavanger, Norway), Shah Nawaz Khan, Muhammad Hashim, Ahmed Khan (Pakistan Wetlands Programme, Islamabad), David Rodrigues (Escola Superior Agrária de Coimbra, Portugal), A. A. Samoilov, A. V. Karelov, A. Y. Volkov, M. V. Gavrilickev, G. V. Gonokhin, A. N. Orlov, N. D. Poyarkov, V.N. Stepanov, O. Tutenkov, V.I. Zalogin, V. N. Stepanov, Y. Konstantinov, V. S. Galtsov, Valery Buzun (Russia), Dmitry A. Sartakov (Ecological Watch Siberia, Omsk, Russia), Sergei A. Soloviev (Omsk State University, Russia), Anna Palmé (Stockholm University, Sweden), David Schonberg Alm, Jonas Waldenström (Ottenby Bird Observatory, Sweden), Mitja Kersnik (Slovenia), O. V. Koshyn, I. V. Shydlovkyy, D. O. Klymysmyn, Ihor V. Shydlovkyy, Oksana Zakala (Ivan Franko National University of Lviv, Ukraine), Brandt Meixell, Danielle Mondloch and Jonathan Runstadler (University of Alaska Fairbanks, USA). Alyn Walsh and Dominic Berridge (Wexford Wildfowl Reserve, Ireland) helped in sampling set-up and coordination in Great Britain. Holly Middleton helped in coordinating sampling efforts in Canada. We thank the staff of the Greenland Institute of Natural Resources for their support during our Greenland expedition, especially Aili Lage Labansen for organizing our stay, and Carsten Egevang for hosting us in his laboratory. Hans Geisler supported our trapping activities at the sampling site in Nuuk. The Animal Breeding and Genomics Group (Wageningen, the Netherlands) generously hosted us in their molecular laboratory. Sylvia Kinders, Tineke Veenendaal and Bert Dibbits are thanked for helping with laboratory work. Peter Beerli provided outstanding support for MIGRATE-N, and Rudy Jonker helped in exploring DAPC and MIGRATE-N, and Mathias Kuemmerlen created Fig. 1. Conny Smith assisted in scrutinizing North American ringing data sets. This work was financially supported by the KNJV (Royal Netherlands Hunters Association), the Dutch Ministry of Agriculture, the Faunafonds, the Stichting de Eik Trust and the Wageningen LEB Trust (all in the Netherlands).

## References

Als TD, Hansen MM, Maes GE *et al.* (2011) All roads lead to home: panmixia of European eel in the Sargasso Sea. *Molecular Ecology*, **20**, 1333–1346.

Altshuler D, Pollara VJ, Cowles CR *et al.* (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, **407**, 513–516.

Anon. U.S. Fish and Wildlife Service. Available from http://www.flyways.us/flyways/info#flyways-bio.

Avise JC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.

Avise JC, Ankney CD, Nelson WS (1990) Mitochondrial gene trees and the evolutionary relationship of Mallard and Black Ducks. *Evolution*, **44**, 1109–1119.

Beaumont MA, Nielsen R, Robert C *et al.* (2010) In defence of model-based inference in phylogeography. *Molecular Ecology*, **19**, 436–446.

Beerli P (2010) Tutorial: Comparison of gene flow models using Bayes Factors. Available from Web Tutorial http://popgen.sc.fsu.edu/Migrate/Tutorials/Entries/2010/7/12_Day_of_longboarding.html.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763–773.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4563–4568.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, **185**, 313–326.

Black WC IV, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: genome-wide sampling of insect populations. *Annual Review Of Entomology*, **46**, 441–469.

Bloomquist EW, Lemey P, Suchard MA (2010) Three roads diverged? Routes to phylogeographic inference. *Trends in Ecology and Evolution*, **25**, 626–632.

Bryant D, Moulton V (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, **21**, 255–265.

Campbell NR, Narum SR (2011) Development of 54 novel single-nucleotide polymorphism (SNP) assays for sockeye and coho salmon and assessment of available SNPs to differentiate stocks within the Columbia River. *Molecular Ecology Resources*, **11**, 20–30.

Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**, 1439–1454.

Champagnon J, Guillemain M, Gauthier-Clerc M, Lebreton JD, Elmberg J (2009) Consequences of massive bird releases for hunting purposes: mallard *Anas platyrhynchos* in the Camargue, southern France. *Wildfow* Special issue, **2**, 184–191.

Eaton MA, Brown AF, Noble DG *et al.* (2009) Birds of conservation concern 3: the population status of birds in the United Kingdom, Channel Islands and Isle of Man. *British Birds*, **102**, 296–341.

Elmberg J (2009) Are dabbling ducks major players or merely noise in freshwater ecosystems? A European perspective, with references to population limitation and density dependence. *Wildfow* Special issue, **2**, 9–23.

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591–1604.

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.

Flint PL, Ozaki K, Pearce JM *et al.* (2009) Breeding-season sympatry facilitates genetic exchange among allopatric wintering populations of Northern Pintails in Japan and California. *Condor*, **111**, 591–598.

Friesen VL, Burg TM, McCoy KD (2007) Mechanisms of population differentiation in seabirds. *Molecular Ecology*, **16**, 1765–1785.

Geraci J, Béchet A, Cézilly F *et al.* (2012) Greater flamingo colonies around the Mediterranean form a single interbreeding population and share a common history. *Journal of Avian Biology*, **43**, 341–354.

Guillemain M, Sadoul N, Simon G (2005) European flyway permeability and abmigration in Teal *Anas crecca*, an analysis based on ringing recoveries. *Ibis*, **147**, 688–696.

Gunnarsson G, Latorre-Margalef N, Hobson KA *et al.* (2012) Disease dynamics and bird migration-linking mallards *Anas platyrhynchos* and subtype diversity of the influenza a virus in time and space. *PLoS ONE*, **7**, e35679.

Han YS, Hung CL, Liao YF, Tzeng WN (2010) Population genetic structure of the Japanese eel *Anguilla japonica*: panmixia at spatial and temporal scales. *Marine Ecology Progress Series*, **401**, 221–232.

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.

Huson DH (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*, **14**, 68–73.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

ISO3166 (2007) ISO 3166 Maintenance agency. International Organization for Standardization.

Ito T, Okazaki K, Kawaoka Y *et al.* (1995) Perpetuation of influenza A viruses in Alaskan waterfowl reservoirs. *Archives of Virology*, **140**, 1163–1172.

IUCN (2011) IUCN Red List of Threatened Species. Downloaded on 13 July 2011, Version 2011.1.

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jonker RM, Kraus RHS, Zhang Q *et al.* (2011a) Genetic consequences of breaking migratory traditions in barnacle geese. In: Revolutionary Non-Migratory Migrants (ed. Jonker RM), pp. 71–83 PhD Thesis. Wageningen University. Wageningen, The Netherlands.

Jonker RM, Zhang Q, Van Hooft P *et al.* (2012) The development of a genome wide SNP set for the Barnacle Goose *Branta leucopsis*. *PLoS ONE*, **7**, e38412.

Jorgensen SJ, Reeb CA, Chapple TK *et al.* (2010) Philopatry and migration of Pacific white sharks. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **277**, 679–688.

Kass RE, Raftery AE (1995) Bayes Factors. *Journal of the American Statistical Association*, **90**, 773–795.

Kerstens HHD, Crooijmans RPMA, Veenendaal A *et al.* (2009) Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: applied to Turkey. *BMC Genomics*, **10**, 479.

Kingman JFC (1982a) The coalescent. *Stochastic Processes and Their Applications*, **13**, 235–248.

Kingman JFC (1982b) On the genealogy of large populations. *Journal of Applied Probability*, **19**, 27–43.

Knowles LL (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, **17**, 1–10.

Koenig WD, Van Vuren D, Hooge PN (1996) Detectability, philopatry, and the distribution of dispersal distances in vertebrates. *Trends in Ecology and Evolution*, **11**, 514–517.

Kovach AI, Breton TS, Berlinsky DL, Maceda L, Wirgin I (2010) Fine-scale spatial and temporal genetic structure of Atlantic cod off the Atlantic coast of the USA. *Marine Ecology Progress Series*, **410**, 177–195.

Kraus RHS, Kerstens HHD, van Hooft P *et al.* (2011a) Genome wide SNP discovery, analysis and evaluation in mallard (*Anas platyrhynchos*). *BMC Genomics*, **12**, 150.

Kraus RHS, Zeddeman A, Sartakov D, Soloviev SA, Ydenberg RC, Prins HHT (2011b) Evolution and connectivity in the world-wide migration system of the mallard: inferences from mitochondrial DNA. *BMC Genetics*, **12**, 99.

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

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.

Laikre L, Palmé A, Josefsson M, Utter F, Ryman N (2006) Release of alien populations in Sweden. *Ambio*, **35**, 255–261.

Lamb HH (1975) Our understanding of global wind circulation and climatic variations. *Bird Study*, **22**, 121–141.

Lebarbenchon C, Yang M, Keeler SP *et al.* (2011) Viral replication, persistence in water and genetic characterization of two influenza a viruses isolated from surface lake water. *PLoS ONE*, **6**, e26566.

Li MH, Merilä J (2010) Genetic evidence for male-biased dispersal in the Siberian jay (*Perisoreus infaustus*) based on autosomal and Z-chromosomal markers. *Molecular Ecology*, **19**, 5281–5295.

Liechti F (2006) Birds: Blowin' by the wind? *Journal of Ornithology*, **147**, 202–211.

Liu Y, Keller I, Heckel G (2011) Range-wide genetic population structure of common pochard (*Aythya ferina*): a potentially important vector of highly pathogenic avian influenza viruses. *Ecology and Evolution*, **1**, 529–545.

Liu Y, Keller I, Heckel G (2012) Breeding site fidelity and winter admixture in a long-distance migrant, the tufted duck (*Aythya fuligula*). *Heredity*, **109**, 108–116.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Mesnick SL, Taylor BL, Archer FI *et al.* (2011) Sperm whale population structure in the eastern and central North Pacific inferred by the use of single-nucleotide polymorphisms, microsatellites and mitochondrial DNA. *Molecular Ecology Resources*, **11**, 278–298.

Milligan BG (2003) Maximum-likelihood estimation of relatedness. *Genetics*, **163**, 1153–1167.

Miyabayashi Y, Mundkur T (1999) Atlas of Key Sites for Anatidae in the East Asian Flyway. Wetlands International—Japan, Tokyo, and Wetlands International—Asia Pacific, Kuala Lumpur.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, **19**, 208–216.

Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Molecular Ecology*, **18**, 1034–1047.

Olsen B, Munster VJ, Wallensten A, Waldenström J, Osterhaus ADME, Fouchier RAM (2006) Global patterns of influenza A virus in wild birds. *Science*, **312**, 384–388.

Pardini AT, Jones CS, Noble LR *et al.* (2001) Sex-biased dispersal of great white sharks. *Nature*, **412**, 139–140.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Peters JL, Zhuravlev Y, Fefelov I, Logie A, Omland KE (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, Humphries EM, Omland KE (2008) Multilocus phylogeography of a Holarctic duck: colonization of North America from Eurasia by gadwall (*Anas strepera*). *Evolution*, **62**, 1469–1483.

Peters JL, McCracken KG, Pruett CL *et al.* (2012) A parapatric propensity for breeding precludes the completion of speciation in common teal (*Anas crecca*, sensu lato). *Molecular Ecology*, **21**, 4563–4577.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, **16**, 37–45.

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

R Development Core Team (2009) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org.

Reudink MW, Kyle CJ, Nocera JJ, Oomen RA, Green MC, Somers CM (2011) Panmixia on a continental scale in a widely distributed colonial waterbird. *Biological Journal of the Linnean Society*, **102**, 583–592.

Rhodes OE Jr, Smith LM, Chesser RK (1995) Apportionment of genetic variance in migrating and wintering Mallards. *Canadian Journal of Zoology*, **73**, 1182–1185.

Rhymer JM (2006) Extinction by hybridization and introgression in anatine ducks. *Acta Zoologica Sinica*, **52**(Suppl), 583–585.

Rypien KL, Andras JP, Harvell CD (2008) Globally panmictic population structure in the opportunistic fungal pathogen *Aspergillus sydowii*. *Molecular Ecology*, **17**, 4068–4078.

Sacks BN, Moore M, Statham MJ, Wittmer HU (2011) A restricted hybrid zone between native and introduced red fox (*Vulpes vulpes*) populations suggests reproductive barriers and competitive exclusion. *Molecular Ecology*, **20**, 326–341.

Santure AW, Stapley J, Ball AD, Birkhead TR, Burke T, Slate J (2010) On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology*, **19**, 1439–1451.

Scott DA, Rose PM (1996) Atlas of Anatidae Populations in Africa and Western Eurasia. Wetlands International Publication No. 41, Wetlands International, Wageningen, The Netherlands.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615–629.

Sherry ST, Ward MH, Kholodov M *et al.* (2001) DbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, **29**, 308–311.

Si Y, Skidmore AK, Wang T *et al.* (2009) Spatio-temporal dynamics of global H5N1 outbreaks match bird migration patterns. *Geospatial Health*, **4**, 65–78.

Smith LM, Burgoyne LA (2004) Collecting, archiving and processing DNA from wildlife samples using FTA® databasing paper. *BMC Ecology*, **4**, 4.

Stallknecht DE, Goekjian VH, Wilcox BR, Poulson RL, Brown JD (2010) Avian influenza virus in aquatic habitats: what do we need to learn? *Avian Diseases*, **54**, 461–465.

Templeton AR (2010) Coalescent-based, maximum likelihood inference in phylogeography. *Molecular Ecology*, **19**, 431–435.

Thompson AL (1931) On 'abmigration' among the ducks, an anomaly shown by the results of bird-marking. *Proceedings of the International Ornithological Congress*, **7**, 382–388.

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, **34**, 275–305.

Wang J (2011) Coancestry: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, **11**, 141–145.

Wang Y, Hey J (2010) Estimating divergence parameters with small samples from a large number of loci. *Genetics*, **184**, 363–379.

Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y (1992) Evolution and ecology of influenza A viruses. *Microbiological Reviews*, **56**, 152–179.

Williams LM, Oleksiak MF (2011) Ecologically and evolutionarily important SNPs identified in natural populations. *Molecular Biology and Evolution*, **28**, 1817–1826.

Willing EM, Bentzen P, Van Oosterhout C *et al.* (2010) Genome-wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. *Molecular Ecology*, **19**, 968–984.

Zhang G, Shoham D, Gilichinsky D, Davydov S, Castello JD, Rogers SO (2006) Evidence of influenza A virus RNA in Siberian lake ice. *Journal of Virology*, **80**, 12229–12235.

---

R.H.S.K. designed the study, coordinated sample collection, prepared DNA, analysed and interpreted data, and wrote the manuscript. P.V.H. analysed and interpreted the data, and revised the manuscript. H.-J.M. interpreted data and revised the manuscript. H.H.T.P. and R.C.Y. interpreted data, co-wrote the manuscript, and coordinated sample collection. A.T. and S.Y.F. coordinated sampling and revised the manuscript. All authors read and approved this paper.

---

## Data accessibility

All data presented in this study can be obtained from the corresponding author. Details on samples as well as the output of the MIGRATE-N analysis are available as online material on the journal website. SNP genotypes are deposited in the DRYAD repository under doi: 10.5061/dryad.1bq39.

## Supporting information

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

**Appendix S1** Full details on sampling localities and samples, including individual IDs, sampling dates and locations (country, region/city, longitude/latitude), sex, and collector names. File type MS Excel (.xls).

**Appendix S2** Analytical output from the best supported model in the MIGRATE-N analysis. File type Adobe portable document file (.pdf).