Conserv Genet
DOI 10.1007/s10592-007-9358-3

RESEARCH ARTICLE

High genetic diversity in the blue-listed British Columbia population of the purple martin maintained by multiple sources of immigrants

Allan J. Baker · Annette D. Greenslade · Laura M. Darling · J. Cam Finlay

Received: 8 February 2007 / Accepted: 5 June 2007

Abstract To assess genetic diversity in the blue-listed purple martin (*Progne subis*) population in British Columbia, we analysed mitochondrial control region sequences of 93 individuals from British Columbia and 121 individuals collected from seven localities of the western and eastern North American subspecies *P. s. arboricola* and *P. s. subis*, respectively. Of the 47 haplotypes we detected, 34 were found exclusively in western populations, and 12 were found only in eastern populations. The most common eastern haplotype (25) was also found in three nestlings in British Columbia and one in Washington. Another British Columbia nestling had a haplotype (35) that differed by a C to T transition from haplotype 25. Coalescent analysis indicated that these five nestlings are probably descendents of recent immigrants dispersing from east to the west, because populations were estimated to have diverged about 200,000–400,000 ybp, making ancestral polymorphism a less likely explanation. Maximum likelihood estimates of gene flow among all populations detected asymmetrical gene flow into British Columbia not only of rare migrants from the eastern subspecies in Alberta but also a substantial number of migrants from the adjacent Washington population, and progressively lower numbers from Oregon in an isolation-by distance pattern. The influx of migrants from different populations is consistent with the migrant-pool model of recolonization which has maintained high genetic diversity in the small recovering population in British Columbia. Thus, the risk to this population is not from genetic erosion or inbreeding following a severe population crash, but from demographic stochasticity and extinction in small populations.

Keywords Genetic diversity · Migrant-pool recolonization · Blue-listed population · Purple martin · Nest box recovery

A. J. Baker (✉) · A. D. Greenslade
Department of Natural History, Royal Ontario Museum, Toronto, ON, Canada M5S 2C6
e-mail: allanb@rom.on.ca

L. M. Darling
B.C. Ministry of Environment, STN PROV GOVT, P. O. Box 9398, Victoria, BC, Canada V8W 9M9

J. C. Finlay
270 Trevlac Place, Victoria, BC, Canada V9E 2C4

Introduction

Populations at an ecologically limiting edge of a species range are prone to extinction-recolonization dynamics, and thus may be composed of spatially structured local populations linked by gene flow approximating a metapopulation. During recolonization or local bottleneck effects the reduction in effective population size would be expected to lead to a reduction in genetic diversity in newly recovered populations (Nei et al. 1975; Ardern et al. 1997; Tarr et al. 1998) and increased genetic differentiation among populations (Barton and Whitlock 1997; Tarr et al. 1998; Clegg et al. 2002). The consequences of local extinctions or near extinctions on genetic structure can vary depending on rates and patterns of migration. If colonizers originate from a single source population (the propagule-pool model) genetic differentiation is expected to increase, but not if the number of colonizers is high and they come from multiple populations (the migrant-pool model) (Slatkin 1977; Whitlock and McCauley 1990). The effects of extinction-recolonization events are therefore an important component of population recovery assessments in peripheral

populations because migration might be limited, and environmental and spatial factors could be important as well (Giles and Goudet 1997).

The purple martin (*Progne subis*) is the largest swallow found in North America. Three subspecies are currently recognized, each with supposedly non-overlapping breeding distributions. The nominate subspecies, *P. s. subis*, is known to breed from east of the Rockies to New Brunswick and throughout the eastern USA. The western subspecies, *P. s. arboricola*, was described from populations in the US Great basin mountain ranges of Utah and Colorado (Behle 1968). This designation has also been applied to taxonomically little studied populations breeding in central Rocky Mountain ranges and along the western seaboard from the southwest seaboard of British Columbia through Washington, Oregon, California, and into Mexico (Brown 1997; Pridgeon 1997). A third subspecies, *P. s. hesperia*, breeds in the Sonoran desert areas of the south-western United States and Mexico (Brown 1997). Populations in British Columbia represent the northwest limit of the species range. While there have been reports of the eastern subspecies in British Columbia, these individuals are assumed to be vagrants because they were rare non-breeding birds east of the Rocky Mountains near the northwest Alberta border (Campbell et al. 1997); only the western subspecies is known to breed in British Columbia (Brown 1997; Copely et al. 1999; Fraser et al. 2000). Purple martins were once distributed on southern and eastern Vancouver Island, and in the mainland in the Greater Vancouver area and lower Fraser Valley. They nested in buildings in Vancouver until 1948 (Campbell et al. 1997) and disappeared from the city in 1950 (Fraser et al. 2000); the last known pair to nest in the mainland was in a piling at Pitt Meadows in 1972 (Plath 1994). The breeding distribution of this population has contracted because of severe range-wide population declines in the mid-1900s attributed to loss of cavity nesting sites and strong competition for these cavities from two introduced bird species, European starlings (*Sturnus vulgaris*) and house sparrows (*Passer domesticus*). From the early 1970s to mid-1980s, there is very little documentation of breeding in the province and at best a small breeding population persisted on southeast Vancouver Island.

Recovery of the species on both Vancouver Island and the lower mainland has occurred, increasing from five known breeding pairs nesting in offshore pilings at 2–3 sites within British Columbia in 1985 to over 600 pairs at 45 occupied colony sites in 2006 (B. Cousens, unpublished data). Population recovery is entirely a result of nest boxes built, erected, monitored and maintained primarily by volunteers at ~50 marine coastal locations (Cousens et al. 2005a). A similar recovery of the Washington population began in 1975 when a nest box program was initiated. By 2004 the Washington population had expanded to about 700 pairs from a few remnant pairs in the mid-1970s (Cousens et al. 2005a). Nevertheless, population numbers remain small, and the western subspecies of the purple martin was Red-listed until 2006 when it was upgraded to Blue-listed (Vulnerable) in British Columbia, a Candidate Species for listing in Washington and a Species of Special Concern in Oregon and California. Given the population fluctuations on record, the British Columbia population might be expected to exhibit low genetic diversity from bottlenecking and possibly inbreeding, making it vulnerable to extinction from both demographic and genetic perspectives.

A fundamental requirement in any population recovery plan is an assessment of the level of genetic variation that currently exists in the British Columbia populations of purple martin, as the severe reduction in population size before 1985 could have produced a genetic bottleneck. Additionally, it is important to estimate the level of immigration per generation into the population from elsewhere, and to identify the source of the immigrants. If the recent recovery of the population was from a small propagule that persisted in British Columbia, the current population might have differentiated from other western populations due to enhanced genetic drift in the propagule, as predicted under the propagule-pool model. Conversely, if recolonization was from multiple source populations, the British Columbia population should not be differentiated from them as expected under the migrant-pool model. To test these models we focus on the rapidly evolving Domain I of the mitochondrial DNA control region (Baker and Marshall 1997) because it is hypervariable in the purple martin and thus provides a rich source of genetic markers for assessing genetic diversity and levels of immigration in the British Columbia population. Although concerns have been expressed about the use of mtDNA to infer population history if it is not selectively neutral (e.g. Ballard and Whitlock 2004), evidence for selection operating on this molecule within bird species is presently lacking or unlikely to affect phylogeographic inference (Zink 2005; Zink et al. 2006). We therefore assayed variation and tested for neutrality in the rapidly evolving control region of mitochondrial DNA in a large representative sample of birds from the British Columbia population and from three other populations in western USA that are potential source populations. Because the nominate subspecies *P. s. subis* east of the Rockies is a potential source of immigrants as well, we also assayed genetic variation in samples collected from Alberta to Pennsylvania.

Methods

Sample collection

Blood samples were collected from western populations at six localities in southwest British Columbia (Fig. 1), three

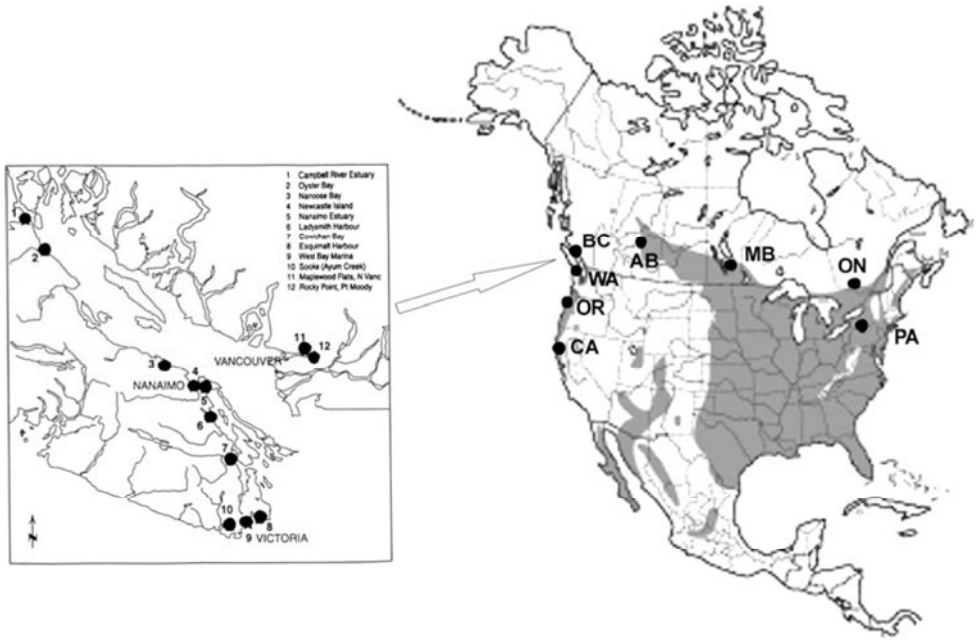

Fig. 1 Breeding distributions and localities in North America where purple martins were sampled in this study (adapted in part from Brown 1997). The inset shows details of where nest sites were recorded in British Columbia (from Fraser et al. 2000). DNA samples were obtained only from 1, 4, 6, 7, 8, 11

in Washington, five in Oregon, and one in California. Samples of blood were also collected from eastern populations at two localities in Manitoba, five in Alberta, and one each in Ontario and Pennsylvania (Table 1). Blood was taken from apparently unrelated adults and subadults in the sample from Sacramento, California, but in all other locales samples were from only one nestling per brood to sample matrilines as randomly as possible. A total of 214 samples were analyzed, of which 93 were obtained from the threatened British Columbia population. Samples in each state or province were pooled into regional populations for analysis of population differentiation and migration.

DNA extraction, PCR and sequencing

Genomic DNA was extracted from blood samples using standard procedures (Sambrook et al. 1989). Samples were homogenized in a solution of STE buffer (100 mM NaCl; 10 mM Tris–HCl, pH 8.0; 1 mM EDTA, pH 8.0), 0.1% SDS, and 10 µg/ml proteinase K, and incubated overnight at 55°C. Dilutions of these crude DNA extractions were used as templates for amplification of the mitochondrial DNA control region via the Polymerase Chain Reaction (PCR) in a total reaction volume of 12 µl, consisting of 10 mM Tris–HCl, pH 8.3; 50 mM KCl; 2.5 mM $MgCl_2$; 0.01% gelatine, 160 mg per ml BSA; 50 µM each dNTP; 0.4 µM each primer; and 1 U Taq DNA polymerase (Boehringer Mannheim). The primer combination used was a passerine-universal primer, H417 (Tarr 1995) and a martin-specific primer, PMb4rev (O. Haddrath, pers. comm.). These primers amplify an approximately 900 bp fragment from which 403 bp of the 5′ end of the control region (Domain I) was sequenced. The following thermal cycle profile was carried out using a Perkin Elmer DNA 480 Thermal Cycler: 94°C for 45 s, 50°C for 45 s; and 72°C for 1 min 30 s, for 35 cycles. An initial denaturation step of 94°C for 5 min and a final extension step of 72°C for 7 min were used.

Following amplification, the total volume of PCR product was subjected to electrophoresis through a 1.5% agarose gel in 1 × TA buffer, and visualized using ethidium bromide and UV illumination. The band containing the PCR product was excised from the gel and purified via filter pipet-tip centrifugation (Dean and Greenwald 1995). DNA sequencing reactions were carried out using a ThermoSequenase sequencing kit (Amersham) as per manufacturer's instructions. Sequencing reactions were run though polyacrylamide gels, dried, and exposed to X-ray film.

Statistical analysis

DNA sequences were read manually and aligned by eye in the computer program XESEE 3.2 (Cabot 1997), and are deposited in Genbank (accession numbers EF641579-EF641792). The program Modeltest 3.6 (Posada and Crandall 1998) was used to determine which model of sequence evolution best fitted the control region sequences. For the purple martin, the control region sequences were modeled best by the TrN (Tamura and Nei 1993) model + gamma. The program Arlequin 2.000 (Schneider et al. 2000) was used to calculate haplotype diversity (Nei 1987). To show relationships among haplotypes, a median-joining haplotype network was constructed from the aligned

Table 1 Purple martin samples analyzed in this study

Locality	Subspecies	Abbreviation	Number analyzed
British Columbia	*P. s. arboricola*	BC	93
Esquimalt Harbour			18
Ladysmith Harbour			16
Newcastle Island			17
Cowichan Bay			24
Maplewood Flats			13
Campbell River			5
Washington		WA	26
Vashon Island			12
Woodard			12
Hylebos Waterway			2
Oregon		OR	25
Roseburg			13
Coyote Creek			5
Long Tom River			3
Signal Island			4
Sacramento, California		CA	20
Alberta	*P. s. subis*	AB	21
Sherwood Park #1			5
Looma			3
Cooking Lake			4
Sampson Lake			5
Sherwood Park #2			4
Manitoba		MB	6
Plessis			4
Park Grove			2
Ontario		ON	9
Pennsylvania		PA	14

sequence data using the program Network 4.1 (Bandelt et al. 1999). Estimates of population differentiation were calculated using Arlequin. Population genetic structure was also assessed with hierarchical analysis of molecular variance (AMOVA) using Arlequin with two levels of population structure: among subspecies and among populations.

To test whether the control region sequences conform with neutral expectations we used DNAsp (Rozas et al. 2003) to compute Tajima's (1989) *D*-values, Fu and Li's (1993) D^* and F^* test statistics, and Fu's (1997) F_s statistic. The latter is more powerful for detecting an excess of young mutations (rare alleles) arising from the effects of population growth on neutral sequences or genetic hitchhiking on selected genes (Fu 1997), whereas D^* and F^* are best for detecting background selection against deleterious alleles (Fu 1997). Significance levels of the statistics were determined using coalescent simulations of 10,000 random samples based on the population parameter values of F_s and θ to obtain critical values.

Because population structure in the purple martin is likely to have been impacted in the late Pleistocene, populations are probably not in genetic equilibrium between the forces of mutation, genetic drift and migration. We therefore conducted coalescent analyses to estimate population divergence times between the two subspecies using the program MDIV (Nielsen and Wakeley 2001) because it potentially can distinguish between isolation and gene flow as determinants of differentiation in two non-equilibrium populations. To correct for multiple substitutions at sites, the HKY model of sequence evolution in MDIV was used as it is the only available option. MDIV is based on a coalescent model that jointly estimates the divergence time and migration rates among pairs of populations using DNA sequence data. Input for the program is DNA sequence and output is four estimated parameters: θ, M, T and TMRCA. These parameters are measures of mutation, migration and divergence time, and can be used to calculate effective population size of females (N_{ef}), the number of migrants exchanged between populations per generation (N_{ef}m), and population divergence time (T) and time to the most recent common ancestor (TMRCA) using the following equations for mitochondrial DNA sequences: $\theta = 2\ N_{ef}\mu$, $M = N_{ef}$m and T (or TMRCA) = t/N_{ef}. Markov Chain Monte Carlo simulations generate posterior probability distributions whose modes represent the parameter estimates. Credibility intervals for the estimates are taken as the interval that contains 95% of the posterior distribution (Nielsen and Wakeley 2001). Parameters were estimated using three runs of a 5,000,000 generation MCMC with an additional 2,000,000 generation burn-in period.

To convert coalescent times to years before present (ybp), a mutation rate (μ) for Domain I of the purple martin sequences is required. However, this rate is unknown for the species and is variable across birds, so instead we followed Brito (2005) in employing a wide range of rates from 5% to 20%/site/million years estimated for the portion of the control region we sequenced. MDIV requires a per locus estimate (μ) so we multiplied the mutation rate per site for Domain I by the sequence length and by generation time (g) to obtain μ. Generation time (g) in the purple martin was estimated from the equation given in (Sæther et al. 2005): g = α + (s/(1–s), where α is the age of first breeding (1 year) and s is annual survival (0.7) of adults (Brown 1997). Thus, generation time is 3.3 years in the purple martin. Using as an example a rate of 10%/site/million years, the rate converts to 0.1 substitutions/site/Myr, or 10^{-7} substitutions/site/yr. The required mutation rate/locus/generation (μ) is calculated as 3.3 year × 403 bp × 10^{-7} = 1.323 × 10^{-4}.

For a broader analysis of levels of gene flow among the eight regional populations we employed a maximum likelihood method using the coalescent as implemented in Migrate (Beerli and Felsenstein 1999, 2001)). This analysis does not assume the equilibrium island model in which

gene flow is symmetric among populations, but instead allows asymmetrical gene flow among populations. The following search parameters were used: 10 short chains of 50,000 steps followed by three long chains of 500,000 steps. Each chain was sampled every 100 steps heating with the following initial relative temperatures {1; 1.1; 1.3; 2} to avoid fatal attraction to $M = 0$. Nucleotide frequencies were estimated from the data, and initial estimates of theta and gene flow were obtained using F_{st} (Beerli 2004). The analysis was run three times with different random numbers and results were averaged because they gave similar estimates of the parameters.

Results

Sequence variation

The average base composition in the sequences from the 214 individuals was 30.1% A, 31.7% C, 12.2% G, and 26.0% T. These values are in accordance with the average base composition reported for Domain I of the control region in birds (Baker and Marshall 1997). Approximately 13% of surveyed sites (51 out of 403 base pairs) were found to be variable. Of these 51 variable sites, 39 were transitions and 12 were transversions, resulting in a transition/transversion ratio of 3.25.

A total of 47 haplotypes was identified among the 214 individuals (Fig. 2). Thirty-four of these were found exclusively in western populations, and 12 were found only in eastern populations. The only haplotype (25) that was shared between eastern and western populations is also the most common eastern haplotype. However, it is at much lower frequency in western populations, where it was found in three individuals from British Columbia and one individual from Washington (Table 2). Additionally, haplotype 35 was detected only in one bird in Washington, but it is part of the eastern assemblage of haplotypes. The British Columbia sample shared seven haplotypes with the neighbouring Washington sample, five with the Oregon sample and only three with the California population. This suggests that gene flow is consistent with an isolation-by-distance model. Another feature of the geographic distribution of haplotypes is that the two common haplotypes in western populations (1 and 11) are absent in eastern populations. Private alleles occur in each population, and this is evident in the large sample from the British Columbia population as well.

Genetic diversity

Both haplotype and nucleotide diversities are higher in western populations than in eastern populations (Table 3). The threatened British Columbia population has very high

```
          111 1111111111 1111111112 2222222222 2222233333 3
   3345566333 3333444555 5566788890 0111224445 5666612246 7
   5663803014 6789024126 8902125933 4149283590 1068911391 5

(56)  TCCGGGCTTG TGGAGGATTA AGTACTTGGC ATGTTTAAAT CCCCAAGCAT G
(6)   .......... .A........ ......A... .......... .......... .
(10)  .......... .......... .......... ......T... .......G.. .
(3)   .......... .......... .......A.. .......... .......G.. .
(1)   .......... .......... .......... .......... .......G.C .
(1)   .......... .......... .....C.... .......... .......... .
(1)   ..T....... .......... G......... .......... .......G.. .
(3)   .......... .......... G......... ......T... .......G.. .
(1)   .......... .......... GA........ .......... .......... .
(4)   .......... C......... .......... .......... .......... .
(15)  .......... .......... .......... .......... .......GGC .
(8)   .......... .......... .......... ....C..... .......GGC .
(1)   .......... .......... .......... .......... ........GC .
(13)  .......... .....A.... .......... .......... .......... A
(3)   .......... .......... ....T..... .......... .......G.. .
(1)   .......... .......... .......... .......... ..GG...GGC .
(1)   .......... .......... .......... ......T... ..GG...... .
(1)   .......... .......... G......... ......T... ..GG...G.. .
(1)   .....A.... ......G... .......... .......... .G.....G.. .
(1)   .......... .A........ ......A... ......T... .......... .
(1)   .......... .......... .......... .........G .GGG...GGC .
(1)   .......... .......... ........A. .......... ....G..G.. .
(3)   .......... .......... .......... ......T... .......... .
(1)   .......... .......... .......... .......... .........C .
(32)  CT..A..... ..AG...GAG ..G.T..... .CA....GG. T...G..G.. .
(1)   CT..A..... ..AG...GAG ..G.T....A CCA....GG. T...G..G.. .
(2)   CTT.A..... ..AG...GAG ..G.T..... .CA....GG. T...G..G.. .
(4)   CT..A..... ...G...GAG ..G.T..... .CA....GG. T...G..G.. .
(1)   CTT.A..C.. ..AG...GAG ..G.T..... .CA....GG. T...G..G.. .
(5)   CT..A..C.. ...GA..GAG ..G.T..... .CA....GG. T...G..G.. .
(1)   CT..A..... ..AG...GAG ..G....... .CA....GG. T...G..G.. .
(1)   CT..A....A ..AG...GAG ..G.T..... .CA....GG. T...G..G.. .
(1)   CT..A..... ..AG...GAG ..G.T..... .CA..C.GG. T...G..G.. .
(2)   CT.AA..... ..AG...GAG ..G.T..... .CA....GG. T...G..G.. .
(1)   CT..A..... ..AG...GAG ..G.T..... .CA....GG. T...G..G.C .
(6)   ......T... .......... .......... ....C.T... .......G.. .
(1)   ..T....... .......... .......... .......... .......... .
(1)   ..T....... .......... G......... .......... .........C .
(5)   .....A.... ......G... .......... .......... .......G.. .
(1)   .......... .......... .......... .......... .......GG. .
(1)   .......... .......... .......... ......TG.. .......G.. .
(1)   CT..A...C. ..AG...GAG ..G.T..... .CA.C..GG. T...G..G.. .
(2)   CT..A..... ..AG...GAG ..G.T..... .CA....GG. T...G.TG.. .
(1)   CT..A..... ..AG...GAG ..GGT..... .CA....GG. T...GG.G.. .
(1)   .......... .A........ .......... .......... .......... .
(5)   .......... .......... .......... ...C..TG.. .......G.. .
(1)   ..T....... .......... .......... ......TG.. .......G.. .
```

Fig. 2 Variable nucleotide sites within the 403 bp segment of the control region sequenced for this study. The frequency of each haplotype is indicated in parentheses, and dots indicate where each haplotype matches the reference sequence. Numbers above each site refer to the position within the aligned sequences. Transversions are indicated by asterisks (*)

genetic diversity in both of these measures, and this is not biased by the large sample size we analyzed for this locality because random draws of 25 individuals exhibited similarly high values (data not shown).

The greater genetic diversity in western populations, specifically in British Columbia, relative to eastern populations, is illustrated graphically by a median-joining network connecting the 47 haplotypes (Fig. 3). The greater diversity of haplotypes in western samples is apparent from the larger number of haplotypes and mutational connections in that part of the network. By contrast, the eastern population samples approximate a star phylogeny with most haplotypes differing from the common central one by one or two mutations. There are 14 mutational steps connecting the western to the eastern haplotype assemblage, suggestive of a relatively ancient split of these assemblages. Mean Tamura-Nei + G distances (0.01178 + 0.00190) between western populations were significantly larger (Kolmolgorov-Smirnov test, $P < 0.005$) than distances (0.00372 + 0.00046) among eastern populations.

Table 2 Control region haplotypes observed in purple martin samples

Haplotype	BC	WA	OR	CA	AB	MB	ON	PA	Total
1	30	14	5	7	–	–	–	–	56
2	5	–	1	–	–	–	–	–	6
3	4	–	5	1	–	–	–	–	10
4	2	1	–	–	–	–	–	–	3
5	1	–	–	–	–	–	–	–	1
6	1	–	–	–	–	–	–	–	1
7	1	–	–	–	–	–	–	–	1
8	3	–	–	–	–	–	–	–	3
9	1	–	–	–	–	–	–	–	1
10	4	–	–	–	–	–	–	–	4
11	10	3	2	–	–	–	–	–	15
12	6	2	–	–	–	–	–	–	8
13	1	–	–	–	–	–	–	–	1
14	7	1	5	–	–	–	–	–	13
15	2	1	–	–	–	–	–	–	3
16	1	–	–	–	–	–	–	–	1
17	1	–	–	–	–	–	–	–	1
18	1	–	–	–	–	–	–	–	1
19	–	1	–	–	–	–	–	–	1
20	1	–	–	–	–	–	–	–	1
21	1	–	–	–	–	–	–	–	1
22	1	–	–	–	–	–	–	–	1
23	3	–	–	–	–	–	–	–	3
24	1	–	–	–	–	–	–	–	1
25	3	1	–	–	12	4	6	6	32
26	–	–	–	–	–	1	–	–	1
27	–	–	–	–	2	–	–	–	2
28	–	–	–	–	2	–	–	2	4
29	–	–	–	–	1	–	–	–	1
30	–	–	–	–	1	–	2	2	5
31	–	–	–	–	1	–	–	–	1
32	–	–	–	–	1	–	–	–	1
33	–	–	–	–	–	1	–	–	1
34	–	–	–	–	1	–	–	1	2
35	–	1	–	–	–	–	–	–	1
36	1	–	–	5	–	–	–	–	6
37	1	–	–	–	–	–	–	–	1
38	–	1	–	–	–	–	–	–	1
39	–	–	5	–	–	–	–	–	5
40	–	–	1	–	–	–	–	–	1
41	–	–	1	–	–	–	–	–	1
42	–	–	–	–	–	–	1	–	1
43	–	–	–	–	–	–	–	2	2
44	–	–	–	–	–	–	–	1	1
45	–	–	–	1	–	–	–	–	1
46	–	–	–	5	–	–	–	–	5
47	–	–	–	1	–	–	–	–	1
Total	93	26	25	20	21	6	9	14	214

Table 3 Haplotype (H) and nucleotide (π_n) diversity in purple martin samples

Population	H	π_n
BC	0.8717 ± 0.0274	0.011767 ± 0.006434
WA	0.7077 ± 0.0949	0.015535 ± 0.008488
OR	0.8633 ± 0.0315	0.007879 ± 0.004698
CA	0.7842 ± 0.0529	0.008011 ± 0.004816
AB	0.6762 ± 0.1110	0.002938 ± 0.002190
MB	0.6000 ± 0.2152	0.002638 ± 0.002318
ON	0.5556 ± 0.1653	0.004532 ± 0.003253
PA	0.8022 ± 0.0905	0.004499 ± 0.003092

Population structure and gene flow

Analysis of population structure with F_{st} assuming the sampled populations are in genetic equilibrium indicated that the eastern and western subspecies are significantly differentiated from one another. Within each subspecies the sampled populations are significantly differentiated from each other, with the exception that the British Columbia and adjacent Washington and Oregon populations are effectively panmictic (Table 4). An exact test (Raymond and Rousset 1995) that does not assume the populations are in genetic equilibrium also showed the same pattern of significant differentiation among regional populations corresponding to the eastern and western subspecies, and confirming that within subspecies only the British Columbia, Washington and Oregon populations are not significantly differentiated (Table 5). Hierarchical analysis of molecular variance (AMOVA) using the program Arlequin (Schneider et al. 2000) apportioned 86% of the total genetic variance among populations in the eastern versus western populations (between subspecies), and only 14% of the variance was partitioned within populations.

Gene flow among sample localities estimated in Migrate was highly asymmetric (Table 6). Although a low level of migration ($Nm = 0.3$) was detected between the two subspecies it was only from a source population in Alberta into the Washington population, and not vice-versa. Conversely, the source of immigrants into the threatened British Columbia population was only from the neighboring states of Washington ($Nm = 50.1$) and Oregon ($Nm = 32.8$), both extremely high levels of gene flow. Populations of *P. s.*

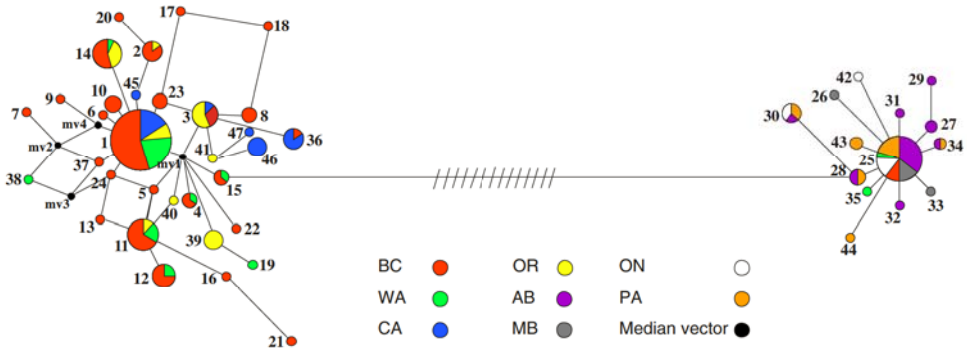

Fig. 3 Median-joining network connecting haplotypes in western (left) and eastern (right) North American populations of the purple martin. Individual haplotypes are represented by circles, the size of which is proportional to the total number of individuals found to possess that haplotype, and are colour-coded by sample locality. Median vectors that represent unobserved hypothetical haplotypes are depicted in black (Bandelt et al. 1999). The 14 mutational steps connecting the western to the eastern haplotype assemblages are depicted with dash marks

Table 4 Pairwise estimates of F_{ST} among regional populations of purple martin

Sample	BC	WA	OR	CA	AB	MB	PA	ON
BC	–							
WA	0.0005	–						
OR	0.0266	0.0559	–					
CA	0.1123*	0.1462*	0.1312*	–				
AB	0.8621*	0.8515*	0.9249*	0.9230*	–			
MB	0.8511*	0.8113*	0.9077*	0.9077*	0.0312	–		
ON	0.8521*	0.8172*	0.9060*	0.9066*	0.0115	0.0490	–	
PA	0.8543*	0.8288*	0.9100*	0.9098*	0.0049	0.0396	0.0494	–

Values with an asterisk are Bonferroni-corrected significant differences from zero at $P < 0.0018$

Table 5 Exact test of genetic differentiation among regional populations of the purple martin

Region	BC	WA	OR	CA	AB	MB	ON
WA	–						
OR	–	+					
CA	+	+	+				
AB	+	+	+	+			
MB	+	+	+	+	–		
ON	+	+	+	+	–	–	
PA	+	+	+	+	–	–	–

A + sign indicates Bonferroni-corrected statistically significant differences among compared regions ($P < 0.0018$)

arboricola are connected by moderate levels of gene flow in western North America, including from British Columbia to Washington, Oregon and California in an isolation-by-distance pattern. In the much more sparsely sampled range of *P. s. subis* in eastern North America migration was more asymmetric (Table 6), with Alberta receiving immigrants from Pennsylvania ($Nm = 10.7$), Ontario from Alberta and Manitoba ($Nm = 4.8$ and 7.4, respectively), and Pennsylvania from Ontario ($Nm = 14.5$).

Tests of neutrality

The large sample from the British Columbia population did not have significant values ($P > 0.10$) of Tajima's *D*-values or Fu and Li's D^* and F^* statistics, as expected under selective neutrality of the control region sequences. However, Fu's F_s statistic was significant and negative in 10,000 coalescent simulations of samples of 93 sequences using empirically derived values of θ and F_s ($F_s = -10.179$, $P = 0.013$). This indicates an excess of rare alleles over neutral expectations, which in theory could arise from either population growth or genetic hitchhiking on selected genes. In contrast, none of the tests in the other three western samples from California, Oregon and Washington or in all samples pooled were significantly different from neutral expectations ($P > 0.10$). Given the recent recolonization of British Columbia from multiple sources including descendents of immigrants (via Washington) from the eastern subspecies which likely contributed alleles that are rare in British Columbia, we conclude that the sequences are effectively neutral, and it is therefore appropriate to proceed with coalescent modeling of divergence time of the western and eastern subspecies.

Table 6 Estimates of gene flow among the eight regional populations of purple martins calculated with Migrate

Receiving population	Source of immigrants							
	BC	WA	OR	CA	AB	MB	ON	PA
BC	–	50.1	32.8	0	0	0	0	0
WA	5.4	–	0	4.1	0.3	0	0	0
OR	2.7	1.4	–	1.1	0	0	0	0
CA	3.3	0.3	1.4	–	0	0	0	0
AB	0	0	0	0	–	0	0	10.7
MB	0	0	0	0	0	–	0	0
ON	0	0	0	0	4.8	7.4	–	0
PA	0	0	0	0	0	0	14.5	–

Coalescent analysis of divergence times

Coalescent modeling of the control region sequences in the two subspecies using 5 million generations and a burn in of 2 million generations provided good estimates of the parameters θ, M, T and TMRCA. The modal value of θ in the Bayesian posterior distribution was 5.945 for eastern versus western populations of the purple martin. The coalescent estimate of gene flow was low ($M = 0.13$). Population divergence time ($T = 2.72$) was estimated at 60,795 generations or 200,625 years before present (ybp) based on a generation time of 3.3 years and a mutation rate per locus of 1.323×10^{-4} substitutions/locus/generation (Fig. 4). Using the 5% and 20%/Myr lower and upper bounds of mutation rates of the control region in birds, this equates to population divergence times between approximately 100,000 and 400,000 ybp. However, given the faster rates of molecular evolution of mtDNA in passerines (Pereira and Baker 2006) a more realistic time frame for population divergence is likely 200,000–400,000 ybp. Thus, the split of the eastern and western subspecies of purple martin is likely to have been in the second half of the Pleistocene, but clearly predates the last glacial maximum 18,000–22,000 ybp.

Discussion

Subspecies divergence and dispersal

Mitochondrial control region sequences indicate that the western subspecies of purple martin, *P. s. arboricola*, is genetically distinct from the eastern subspecies, *P. s. subis*. Only three birds (3.2%) from British Columbia and two (7.7%) from Washington have eastern haplotypes, and four of these have haplotype 25 which is the commonest haplotype in the eastern subspecies. The remaining bird (from Washington) had a unique eastern haplotype (35) which differed from haplotype 25 by a C to T transition. This suggests either that these birds are of the eastern race, or they have retained a haplotype found in the ancestors of both subspecies. Given the large genetic distance (6.6–7.5%) between eastern and western haplotypes, and the long time

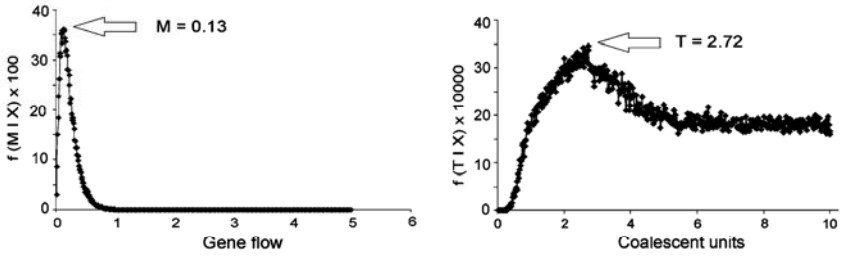

Fig. 4 Posterior distributions of M (gene flow) and T (population divergence time) from MDIV. Modal values of the parameters are indicated by arrows

(~200,000–400,000 year) that has elapsed since subspecies diverged, the sharing of an ancestral haplotype in both subspecies is an unlikely explanation unless long-term effective population is very high. Instead, these five individuals are much more likely to be descendents of recent immigrants from populations of the eastern subspecies because they were nestlings and had haplotypes that were either identical to the common haplotype in the east or differed by one mutation. While there have not been any previous reports of *P. s. subis* breeding in British Columbia, the subspecies does breed in Alberta, and *P. s. subis* has been reported east of the Rockies in northern British Columbia far removed from the breeding sites in southwest British Columbia (Copley et al. 1999; Fraser et al. 2000, 1997).

The large number of nucleotide substitutions between the control region sequences of the two subspecies nevertheless converts to a relatively short time of divergence, comparable to divergence times of other subspecies or very closely related species of North American songbirds such as tufted and black-crested titmice (*Parus b. bicolor* and *P. b. atricristatus*) and boat-tailed and great-tailed grackles (*Quiscalus major* and *Q. mexicanus*) (Klicka and Zink 1997). The genetic divergence of *P. s. subis* and *P. s. arboricola* also accords with their morphometric divergence, with the western birds being markedly larger than the eastern birds (Behle 1968; Brown 1997).

Genetic diversity and gene flow into and from the British Columbia population

The British Columbia population of purple martin is not genetically impoverished, despite being at the edge of the species range, and being of relatively small size. The highest haplotype diversity was found in British Columbia ($H = 0.8717$), as well as the second-highest nucleotide diversity ($\pi_n = 0.01177$). The maximum likelihood method based on the coalescent as implemented in Migrate detected asymmetrical gene flow not only of rare migrants from the eastern subspecies in Alberta but also substantial numbers of migrants from the adjacent Washington population, and progressively lower numbers of migrants from Oregon and California in an isolation-by distance pattern. Lower level migration was also detected from British Columbia to the other western subspecies populations (Table 6). The British Columbia and Washington populations are effectively one panmictic population. This population genetic inference is supported by resightings of banded birds in the two populations. A total of 50 numbered colour bands from Washington have been recorded throughout the British Columbia breeding range since 2002, mainly from Puget Sound where banding began in 2001. Conversely, 25 British Columbia colour-banded birds have been reported in Washington, mainly in Puget Sound, in the same period (Cousens et al. 2005b). Collectively, these findings are consistent with the migrant-pool model of recolonization (Slatkin 1977; Whitlock and McCauley 1990) with the proviso that more migrants are contributed by geographically closer populations.

Recolonization sourced by migrants from throughout the sampled range of the western subspecies therefore explains why genetic variation has not been lost in the population restoration programme using increased numbers of nest boxes, despite extreme reductions in population size in the very recent past. In the absence of an influx of new colonists from other source populations, such severe fluctuations in population size can enhance genetic drift and lead to reductions in haplotype and nucleotide diversities. Population declines due to increased competition for a reduced number of suitable nest sites, as well as weather-related mortality, may simply reflect the fact that populations at the edges of species ranges are more prone to local extinction as a result of marginal conditions for survival. Areas where weather-related mortality is more evident are typically reoccupied within several years (Brown 1997). Range-wide recruitment of migrants from western North America has instead increased genetic variation in the small British Columbia population relative to all other populations. Observed patterns of population decline and recovery are not unique to British Columbia, but are perhaps typical of the species in general; similar trends have been reported from Washington and Oregon (Fraser et al. 2000).

There is little information on migratory routes used by the species (Brown 1997), and until our study the source of colonizing birds into the British Columbian breeding populations (Fraser et al. 1997) and the degree of philopatry were largely unknown (Brown 1997). The isolation-by-distance pattern emerging from our analysis of western populations over a north–south latitudinal gradient implies that the species is not highly philopatric to natal breeding

populations. Furthermore, this suggests that gene flow might be mediated by loss of ephemeral breeding sites and long distance movements of breeding recruits typical of extinction-recolonization dynamics, as evidenced in the nest box inspired recoveries of the Washington and British Columbia populations following recent crashes. The two subspecies are assumed to mix in their over-wintering range in the Amazon Basin, but there little evidence of this from banded birds as yet; one band applied in Oregon in 1975 was recovered in the state of Sao Paulo in Brazil (Hill and Dellinger 1997). Mixing raises the possibility that occasional interchange of migrants from the two subspecies could occur by birds mistakenly switching flyways.

In summary, the potential loss of genetic diversity during the recent severe reduction in population size has been counteracted by migrant-pool immigration into the purple martin population in British Columbia. Thus, the real risks of extinction to this population are ecological, notably from a lack of suitable nest sites. Human intervention in the form of establishing and managing artificial nest boxes has been demonstrated to be highly effective and even essential to prevent local extinction and for population restoration in British Columbia (Pridgeon 1997; Fraser et al. 2000; Cousens et al. 2005a). Although the western and eastern subspecies are well differentiated genetically their inferred low level of interbreeding during the recolonization of British Columbia suggests they are not reproductively isolated.

Acknowledgements The research was made possible by grants and sponsorship from NSERC, the Rocky Point Bird Observatory, British Columbia Ministry of Environment, Lands and Parks, Canada Department of National Defense, Washington Department of Fish and Wildlife, British Columbia Habitat Conservation Trust Fund, Toronto Dominion Friends of the Environment Foundation, and the Purple Martin Conservation Association. We gratefully acknowledge our many volunteer partners who assisted collecting blood samples: California: Dan Airola, Thomas Leeman, Stan Kostka, Stan Wright; Oregon: Bruce Campbell, Kat Beal, Terry Farrell, Marnie Albritten, Wayne Burns; Washington: Michelle Tirhi, Kevin Li; Pennsylvania: James R. Hill III, Eugene Morton, Bridget Stutchbury, Tony Salvadori; Manitoba: Ernie Didur; Ontario, Bill Petrie; Alberta: Hardy Pletz, Pamela Gordy; British Columbia: Tom Gillespie, Michael Setterington. Laboratory work was conducted by O. Haddrath (OH) and A. Greenslade (ADG), and statistical analysis of the DNA data was performed by ADG and AJB. O. Haddrath, Royal Ontario Museum, designed a martin-specific primer and assisted with data collection and experiment design. We thank Bruce Cousens who provided essential input to the data collection and natural history aspects of this paper, and carefully checked the manuscript.

References

Ardern SL, Lambert DM, Rodrigo AG, McLean IG (1997) The effects of population bottlenecks on multilocus DNA variation in robins. J Heredity 88:79–186

Baker AJ, Marshall HD (1997) Mitochondrial control region sequences as tools for understanding evolution. In: Mindell DP (ed) Avian molecular evolution and systematics. Academic Press, San Diego

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. Mol Ecol 13:729–744

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Barton NH, Whitlock MC (1997) The evolution of metapopulations. In: Hanski IA, Gilpin ME (eds) Metapopulation biology: ecology, genetics, and evolution. Academic Press, San Diego

Beerli P (2004) Migrate: documentation and program, part of LAMARC. Version 2.0. Revised December 23, 2004. Distributed over the Internet, http://evolution.gs.washington.edu/lamarc.html

Beerli P, Felsenstein J (1999) Maximum likelihood estimation of migration rates and population numbers of 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. Proc Natl Acad Sci USA 98:4563–4568

Behle WH (1968) A new race of the purple martin from Utah. Condor 70:166–169

Brito PH (2005) The influence of Pleistocene glacial refugia on tawny owl genetic diversity and phylogeography in western Europe. Mol Ecol 14:3077–3094

Brown CR (1997) Purple martin (*Progne subis*). In: Poole A, Gill F (eds) The birds of North America, No. 287. Academy of natural sciences, Philadelphia, PA, and American Ornithologists Union, Washington, DC

Cabot E (1997) XESEE 3.1, Eyeball Sequencer Editor

Campbell RW, Dawe NK, McTaggart-Cowan I, Cooper JM, Kaiser GW, McNall MC, Smith GE (1997) The birds of British Columbia. Vol.3. Passerines: flycatchers through vireos. UBC Press, Vancouver, British Columbia

Clegg SM, Degnan SD, Kikkawa J, Moritz C, Estoup A, Owens IPF (2002) Genetic consequences of sequential founder events by an island-colonizing bird. Proc Natl Acad Sci USA 99:8127–8132

Copley D, Fraser D, Finlay JC (1999) Purple martin, *Progne subis*: a British Columbia success story. Can Field Nat 113:226–229

Cousens B, Lee JC, Darling LM, Finlay JC, Gillespie TW (2005a) Two Decades of Purple martin stewardship and recovery in British Columbia—successes and challenges. Proceedings of the 2005 Puget Sound Georgia Basin research conference, 29–31 March 2005. Seattle, Washington

Cousens B, Lee JC, Kostka S, Darling LM, Tirhi M, Finlay JC, Gillespie TW (2005b) Recovery of the Western Purple martin bordering the 'Salish Sea'—the Georgia Basin of British Columbia and Puget Sound, Washington. Proceedings of the 2005 Puget Sound Georgia Basin research conference, 29–31 March 2005. Seattle, Washington

Dean AD, Greenwald JE (1995) Use of filtered pipet tips to elute DNA from agarose gels. Biotechniques 18:980

Fraser DF, Copley D, Finlay JC (2000) The return of the purple martin in British Columbia. In: Darling LM (ed) Proceedings of a conference on the biology and management of species and habitats at risk, vol 1, Kamloops BC 15–19 Feb (1999), B. C. Ministry of Environment, Lands and Parks, Victoria, B. C. and University College of the Cariboo, Kamloops, B. C

Fraser DF, Siddle C, Copley D, Walters E (1997) Status of the purple martin in British Columbia. Wildlife Status Report Number WR-89. B. C. Ministry of Environment, Lands and Parks. Wildlife Branch

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–925

Fu Y-X, W-H Li (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Giles BE, Goudet J (1997) A case study of genetic structure in a plant metapopulation. In: Hanski IA, Gilpin ME (eds) Metapopulation

biology: ecology, genetics, and evolution. Academic Press, San Diego

Hill JR III, Dellinger TB (1997) Purple martin recoveries south of the border. Purple Martin Update 8:28–29

Klicka J, RM Zink (1997) The importance of recent ice ages in speciation: a failed paradigm. Science 277:1666–1669

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, NY, USA

Nei M, Maruyama T, Chakraborthy R (1975) The bottleneck effect and genetic variability in populations. Evolution 29:1–10

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. Genetics 158:885–896

Pereira SL, AJ Baker (2006) A mitogenomics timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Mol Biol Evol 23:1731–1740

Plath T (1994) Purple martin (*Progne subis*) nest box program—first recorded nesting in 22 years for BC Mainland. Vancouver Natl History Soc Discover 23:143–145

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818

Pridgeon S (1997) Wildlife at risk in British Columbia: purple martin. B. C. Ministry of Environment, Lands and Parks, Victoria, B.C

Raymond M, Rousset F (1995) An exact test for population differentiation. Evolution 49:1280–1283

Rozas J, Sanchez-DelBarrio JC, Messequer X, Rozas R (2003) DNAsp, DNA polymorhism analysis by the coalescent and other methods. Bioinformatics 19:2496–2497

Sæther B, Lande R, Engen S, Weimerskirch H, Lillegård M, Altwegg R, Becker PH, Bregnoballe T, Brommer JE, McCleery RH, Merilä J, Nyholm E, Rendell W, Robertson RR, Tryjanowski P, Visser ME (2005) Generation time and temporal scaling of bird population dynamics. Nature 436:99–102

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Schneider S, Roessli D, Excoffier L (2000) Arlequin ver. 2.000 : a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland

Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. Theoret Pop Biol 12:253–262

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Tarr CL (1995) Primers for amplification and determination of mitochondrial control-region sequences in oscine passerines. Mol Ecol 4:527–529

Tarr CL, Conant S, Fleischer RC (1998) Founder events and variation at microsatellite loci in an insular passerine bird, the Laysan finch (*Telespiza cantans*). Mol Ecol 7:719–731

Whitlock MC, McCauley DE (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. Evolution 44:1717–1724

Zink RM (2005) Natural selection on mitochondrial DNA in *Parus* and its relevance for phylogeographic studies. Proc Roy Soc B 272:71–78

Zink RM, Drovetski SV, Rohwer S (2006) Selective neutrality of mitochondrial ND2 sequences, phylogeography and species limits in *Sitta europea*. Mol Phylogenet Evol 40:679–686