Cryptic genetic variation and paraphyly in ravens

Kevin E. Omland\(^1\), Cheryl L. Tarr\(^2\), William I. Boarman\(^3\), John M. Marzluff\(^4\) and Robert C. Fleischer\(^1\)

\(^1\)Molecular Genetics Laboratory, National Zoological Park, 3001 Connecticut Avenue NW, Washington, DC 20008, USA
\(^2\)Department of Biology, Pennsylvania State University, University Park, PA 16802, USA
\(^3\)US Geological Survey, Department of Biology, University of California, Riverside, CA 92521, USA
\(^4\)College of Forest Resources, University of Washington, Seattle, WA 98195, USA

Widespread species that are morphologically uniform may be likely to harbour cryptic genetic variation. Common ravens (\textit{Corvus corax}) have an extensive range covering nearly the entire Northern Hemisphere, but show little discrete phenotypic variation. We obtained tissue samples from throughout much of this range and collected mitochondrial sequence and nuclear microsatellite data. Our study revealed a deep genetic break between ravens from the western United States and ravens from throughout the rest of the world. These two groups, the 'California clade' and the 'Holarctic clade' are well supported and over 4\% divergent in mitochondrial coding sequence. Microsatellites also reveal significant differentiation between these two groups. Ravens from Minnesota, Maine and Alaska are more similar to ravens from Asia and Europe than they are to ravens from California. The two clades come in contact over a huge area of the western United States, with mixtures of the two mitochondrial groups present in Washington, Idaho and California. In addition, the restricted range Chihuahuan raven (\textit{Corvus cryptoleucus}) of the south-west United States and Mexico is genetically nested within the paraphyletic common raven. Our findings suggest that the common raven may have formerly consisted of two allopatric groups that may be in the process of remerging.

**Keywords:** paraphyletic species; speciation; phylogeography; mitochondrial DNA; microsatellites; \textit{Corvus corax}

1. INTRODUCTION

Understanding patterns of genetic variation within species is crucial to our knowledge of speciation, hybridization, adaptation and lineage diversification. Molecular studies in a wide range of taxa have found evidence of substantial intraspecific genetic differentiation (e.g. > 2\% uncorrected mitochondrial sequence divergence) (Avise 2000). Such patterns can result from a number of systematic practices and evolutionary processes including (i) insufficient taxonomic splitting of polytypic species (e.g. Omland \textit{et al.} 1999; Voelker 1999); (ii) remixing of two or more distinct genetic groups (e.g. Quinn 1992); or (iii) retention of genetic polymorphism within a phenotypically uniform species (Neigel \& Avise 1986; Avise 1994). Discrete genetic breaks within species can provide unique insights into species and population histories, and are an important focus of phylogeographic studies (e.g. Avise 2000).

Several factors increase the chances that species will exhibit major discrete genetic differentiation. Species within genera that are phenotypically uniform may be likely to hide cryptic genetic variation (e.g. chickadees; Gill \textit{et al.} 1993, 1999). Having widespread distributions may also make it more likely that species will show molecular divergence (e.g. Zink \textit{et al.} 1995). Furthermore, over evolutionary time, widespread species may bud off other species by several modes of speciation, especially peripheral isolates speciation (e.g. Mayr 1963). This process will result in widespread ancestral species that are paraphyletic with respect to the restricted-range derived species (e.g. Harrison 1991, 1998; Avise 1994).

The common raven (\textit{Corvus corax}) is an interesting species for a survey of genetic variation because several factors make it likely that they may exhibit cryptic genetic variation and paraphyletic patterns. Common ravens breed throughout most of the Northern Hemisphere (figure 1) and vary little in morphology (although many subspecies have been described based on slight clinal morphometric variation; e.g. Vaurie 1959). Furthermore, ravens are a member of a genus with low morphological variability, especially in plumage (over half the species in \textit{Corvus} are completely black; Madge \& Burn 1994). Several other large-bodied species of \textit{Corvus} are considered closely related to common ravens and are also referred to as ‘ravens’, including the Chihuahuan raven (\textit{Corvus cryptoleucus}) of south-western North America and five species of raven-like corvids in Africa (Madge \& Burn 1994). All of these species have relatively restricted ranges compared to the common raven. For example, the Chihuahuan raven is restricted to the south-western United States and northern Mexico. The Chihuahuan raven differs from the common raven in its smaller size, shorter bill, higher-pitched call and cryptic white bases of the neck feathers (Goodwin 1976).

Mitochondrial sequence and nuclear microsatellite data were used to assess genetic variation in the common raven. Populations were sampled from throughout the Old World and New World, especially the western United States. The principal goal of our study was to determine if the widespread and morphologically uniform common raven harboured extensive genetic variation (see Tarr &
A second goal was to determine if and how any variation was geographically structured. Finally, samples of the Chihuahuan raven were sequenced to establish how this restricted-range species is related to the common raven.

2. METHODS

(a) Laboratory procedures

Common raven samples of 72 individuals were sequenced from localities throughout much of their range (figure 1). Electronic Appendix A available on The Royal Society Web site lists the samples used, along with collection locality, voucher or band information and tissue type. (One published raven cytochrome b sequence from France was used; Gibois & Pasquet 1999.) Also, six Chihuahuan ravens from Texas and New Mexico were sequenced. Samples of outgroup taxa were obtained during the course of previous studies (e.g. Tarr & Fleischer 1999).

DNA was extracted using Qiaamp Tissue Extraction Kits (Qiagen, Valencia, CA, USA). Portions of two mitochondrial gene regions were sequenced, the control region and cytochrome b. Primers used for domain I of the control region were CorLGL-2 (Tarr & Fleischer 1999) and H417 (Tarr 1995), and for cytochrome b were B1 and B2 from Kocher et al. (1989). A typical amplification involved an initial cycle (3 min at 93 °C, 1 min at 50 °C, 45 s at 72 °C) followed by 35 cycles (1 min at 93 °C, 1 min at 50 °C, 45 s at 72 °C) and a final 10-min extension at 72 °C. The resulting PCR products were cleaned using QIAquick Kits (Qiagen). Each gene region was sequenced in both directions with the above primers, using cycle sequencing (10 s at 96 °C, 5 s at 50 °C, 4 min at 60 °C, 25 cycles). The sequencing products were cleaned using Centri-Sep columns (Princeton Separations, Adelphi, NJ, USA) and sequenced on an ABI 373 automated sequencer. Chromatograms were aligned and confirmed using Sequencher sequence analysis software (GeneCodes Corporation, Inc., Ann Arbor, MI, USA). We obtained 314 base pairs (bp) of control region sequence. For cytochrome b we sequenced a subset of the individuals (46 out of 72) and obtained 307 bp of sequence. These sequences have been deposited in GenBank (accession numbers AY003869–AY005980).

Three microsatellite loci designed for Corvus kubaryi by Tarr & Fleischer (1998) were used (Ck.1B6G, Ck.2A5A and Ck.4B6D) following their general protocols. These products were cleaned using Centri-Sep columns (Princeton Separations) and were run out on an ABI 373 automated sequencer with a known size standard. Allele sizes were scored using GeneScan software (ABI). We genotyped a subset of the Holarctic clade individuals in the mitochondrial DNA (mtDNA) study, and a larger sample of individuals from the California populations (see electronic Appendix A).

(b) Genetic analyses

Mitochondrial DNA sequences were imported into PAUP* (Swofford 1999) for phylogenetic analyses. Trees were rooted using several other Corvus as outgroups: American crow (Corvus brachyrhynchos), Mariana crow (Corvus kubaryi) and Hawaiian crow (Corvus hawaiiensis). Broader analyses of published corvid sequences show that these are appropriate outgroups (Fleischer & McIntosh 2000; K. E. Omland, unpublished analysis). Parsimony and neighbour-joining searches were conducted using PAUP* on the genes separately and combined. Bootstrap analyses were conducted with ‘maxtrees’ set at 100, three random addition replicates and a total of 500 bootstrap pseudoreplicates.

Microsatellite allele data were analysed in Arlequin (Schneider et al. 1999) to test for significant population structure at several spatial scales. The mtDNA results were used as a basis for defining major groups for the microsatellite analyses, but we also tested for significant microsatellite variation within mitochondrial groups among geographical regions. FST-values were calculated using AMOVA (Excoffier et al. 1992) and a permutation test implemented in Arlequin was used to test the significance of differentiation between geographical regions. Arlequin permutes genotypes between populations to obtain a null distribution of pairwise FST-values (Schneider et al. 1999). Sequence data were also imported into Arlequin to test for significant geographically structure in the data again using AMOVA and the permutation test.

3. RESULTS

(a) Two distinct clades of common raven

Common ravens form two distinct mitochondrial clades worldwide (figure 2). One group mainly consists of ravens from California and also includes 50% of the individuals from Washington and Idaho; we will refer to this

![Figure 1. Distribution of the common raven (shaded) and Chihuahuan raven (outline) (based on Madge & Burn 1994).](image-url)
Cryptic genetic variation and paraphyly in ravens

K. E. Omland and others

Figure 2. 50% bootstrap tree based on combined cytochrome b plus control region sequences for all individuals for which both regions were sequenced. Values above the branches are bootstrap support values that resulted from 500 pseudoreplicates based on an equally weighted parsimony analysis. Tree shows the two clades of common raven, with the Chihuahuan raven branching within the common raven (bootstrap values relevant to those groups are shown in bold).

Table 1. Percent divergence within and between clades of common raven for two gene regions

(Calculations are based on uncorrected distances among all unique haplotypes within each gene region.)

<table>
<thead>
<tr>
<th>Region</th>
<th>between clades</th>
<th>within California</th>
<th>within Holarctic</th>
</tr>
</thead>
<tbody>
<tr>
<td>cytochrome b</td>
<td>mean 4.04</td>
<td>0.33</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>range (3.26-4.89)</td>
<td>(0-0.33)</td>
<td>(0-1.95)</td>
</tr>
<tr>
<td>control region</td>
<td>mean 5.02</td>
<td>1.10</td>
<td>1.78</td>
</tr>
<tr>
<td></td>
<td>range (3.50-6.69)</td>
<td>(0-2.23)</td>
<td>(0-3.50)</td>
</tr>
</tbody>
</table>

California versus Holarctic allele frequency differences are evident in two out of the three individual loci (table 2). Locus Ck.IB6G shows one short allele that is only found in the California clade (12% of alleles) and one long allele that is restricted to the Holarctic clade (7% of alleles). At locus Ck.2A5A the most common alleles differ between clades, with 55% of the Holarctic alleles being 147 bp, whereas only 26% of the California
alleles are that length. Between clade differentiation is significant for this locus ($F_{ST} = 0.128; p < 0.0001$).

(b) Genetic variation within clades
There is also substantial mtDNA variation within the clades (figure 3). The majority of individuals have unique control region haplotypes (39 haplotypes out of 72 individuals), with up to 3.5% divergence within the Holarctic clade and up to 2.2% divergence within the California clade (table 1). However, this variation does not show strong geographical structuring within clades. Generally, individuals from the New and Old Worlds are found throughout much of the Holarctic clade. For example, notice that two Alaska individuals (collected in Fairbanks on the same day from the same locality) are in different parts of the clade. Indeed, the basal branches in the Holarctic clade include representatives of a wide assortment of Old and New World populations. However, there is limited geographical structure evident in the Holarctic group (figure 3). A large homogeneous clade from Alaska-954 to Maine-3 consists entirely of New World birds. Analysis of geographical structure in Arlequin does reveal evidence of significant differentiation between Old World and New World populations in the Holarctic clade ($F_{ST} = 0.226; p = < 0.0001$). There is no apparent geographical structure within the California clade.

c) Contact between the two clades
The two clades seem to meet over a large region in the western United States. The map in figure 4 shows the number of the two haplotype groups found in each locality. Twenty-seven out of the 28 individuals from California have ‘California-type’ DNA. All individuals from Alaska, Minnesota, Maine, New Mexico, Russia, Siberia, Mongolia and Europe have Holarctic type DNA. However, populations from several states, especially the Pacific North-West showed mixtures of the two mitochondrial types. Twelve individuals from Idaho were sequenced from the same county, of which seven are in the California clade and five are in the Holarctic clade. Similarly, in Washington State, populations from the Olympic Peninsula and inland Washington both show a random mixture of the two haplotype groups. Although only one of the individuals sequenced from California has Holarctic type DNA, both of the individuals from New Mexico are in the Holarctic clade.

(d) Position of the Chihuahuan raven
The Chihuahuan raven is genetically nested within the two clades of the ‘common raven’. The widespread common raven is paraphyletic with respect to the Chihuahuan raven. The Chihuahuan raven is sister to the California clade and branches with it in 73% of bootstrap replications. Chihuahuan raven samples from Texas to western New Mexico all show this relationship. However, the Chihuahuan sequences are very distinctive. The Chihuahuan raven sequences differ from each of the common raven haplotypes by at least 13 substitutions in the combined cyt b plus control region data. Considering cyt b alone, the average sequence divergence between the California clade haplotypes and the Chihuahuan ravens is 1.9% (range 1.6–2.3%). In contrast, the average cyt b divergence between members of the Holarctic clade and the Chihuahuan raven is 4.7% (range 3.9–5.9%).

4. DISCUSSION
(a) Deep split between California and Holarctic clades
Our findings based on mtDNA sequences and nuclear microsatellite alleles show a distinct split between two clades of the common raven. This divide generally separates populations in California from populations throughout the rest of the world; these two groups come into contact in the western United States. The two clades are strongly supported by phylogenetic analyses, and cyt b sequences reveal over 4% average sequence divergence between clades. Assuming mitochondrial coding sequence divergence of ca. 1.6% per million years (Myr) (e.g. Tarr & Fleischer 1993; Fleischer et al. 1998) this level of divergence suggests that the two clades split over 2 Myr ago. (See Hillis et al. (1996) for a discussion of the factors that affect confidence limits on molecular clock estimates.) Such timing may correspond roughly to a dramatic glacial period (Behrensmeyer et al. 1992) that could have pushed common ravens into southern refugia. Dispersal of founding individuals to or from southwestern North America during this time-period is also possible.

The period of apparent isolation and molecular divergence seems to have been accompanied by overall phenotypic stasis. No obvious discrete morphological or behavioral differences separate the two groups (e.g.
Willett 1941; Grinnell & Miller 1944; but see discussion of subspecies in § 4(d)). Although decoupling of molecular and morphological divergence is not the rule, such decoupling has been documented in many taxa (reviewed in Omland 1997b).

The two clades of common ravens now meet across large areas of the western United States; they are found together in the same populations in nearly equal numbers in Idaho and Washington. The finding of the two haplotype groups intermingled over large areas of the western United States suggests that this isolation has been followed by subsequent remixing of the two genetic groups. The two groups may interbreed because they are found together over such a large area of the west. Another possible explanation for our findings is that the two mitochondrial haplotypes could generally correspond to ecologically distinct groups, one predominantly a high-altitude/forest raven, and the other a low-altitude/desert raven (also see Boorman & Heinrich 1999). Future genetic and field research is needed to determine whether there is current gene flow between the two clades and whether the two clades are ecologically distinct.

Figure 3. Neighbour-joining tree based on uncorrected distances among control region sequences for all individuals in the study. Note the deep split between the Holarctic clade and California clade of the common raven.
(b) Comparisons with genetic breaks found in other species

Our findings of a distinct genetic break without apparent phenotypic differentiation in common ravens is similar to cryptic genetic variation documented in many other taxa (reviewed in Avise 2000). However, few previous studies have shown such a deep mitochondrial split that is geographically structured but has such a wide contact zone (but see Orti et al. 1994, sticklebacks; R. C. Fleischer, unpublished data, Asian elephants). For example, Carolina chickadee haplotypes completely shift from one mitochondrial type to the other across a relatively narrow 400 km east–west transect (Gill et al. 1993, 1999; also see Patton & da Silva 1998). Other research has revealed deep mitochondrial splits that show little or no geographical structure (e.g. Quinn 1992, snow geese; Kahn et al. 1999, large-bodied sage grouse). Many other studies have documented distinct intraspecific breaks in mitochondrial data (e.g. Cicero 1996; Omland et al. 1999), but these breaks correspond to clear differences in morphology, behaviour and/or nuclear markers.

The California clade of common ravens seems to be confined to the south-west coast region of North America. There are several other birds that may show congruent geographical breaks in mtDNA including titmice (Cicero 1996) and vireos (Cicero & Johnson 1998). (Also, yellow-billed magpies, Pica nuttalli, are restricted to this area and show distinctive coloration compared to black-billed magpies, Pica pica, e.g. Peterson 1990). Furthermore, several other very different taxa show evidence of phylogeographical breaks in the Pacific North-West, including several plant species (Soltis et al. 1997) and perhaps black bears (Cronin et al. 1991).

(c) Common raven paraphyly relative to the Chihuahuan raven

Our study shows that the restricted-range Chihuahuan raven is genetically nested within the common raven. In other words, the common raven is paraphyletic with respect to the Chihuahuan raven. This finding is supported by sequences from six Chihuahuan ravens, and confirmed by sequences from cytochrome b and the control region. What does common raven paraphyly suggest about speciation in the Chihuahuan raven? This pattern suggests that the Chihuahuan raven split off from the California clade of the common raven after the California–Holarctic split. If that is the case then the widespread common raven probably has the phenotypic characteristics of the ancestor of the Chihuahuan raven. Although no modern phylogenetic studies have addressed the affinities of the Chihuahuan raven, previous researchers have generally assumed that it is closely related to the common raven (e.g. Goodwin 1976; Jollie 1978). Chihuahuan raven sequences are on average 1.9% divergent from the California clade, which suggests that these two groups split from each other ca. 1 Myr ago (assuming divergence of 1.6% per Myr; Fleischer et al. 1998).

Our findings of paraphyly in the common raven are analogous to other studies that have revealed widespread highly divergent species with restricted-range species nested within them. For example, Slade & Moritz (1998) documented a cryptic break in cane toads (Bufo marinus, widespread in Central and South America) and found Bufo pananensis (restricted eastern South America) nested within it. Melnick et al. (1993) documented a mtDNA split in Rhesus macaques (Macaca mulatta), which is paraphyletic with respect to two island species, the Japanese (Macaca fuscata) and Taiwan (Macaca cyclopis) macaques. Neither of the widespread paraphyletic species show evidence of corresponding phenotypic differentiation.

Many similar studies in diverse taxa report similar paraphyly involving widespread and restricted range species, such as mallard ducks (Avise et al. 1990; also see Omland 1997a), pocket gophers (Patton & Smith 1994) and several insect species (Brown et al. 1994; Funk et al. 1995). Many researchers are quick to suggest hybridization as the primary explanation for finding two species’ DNA intermixed (e.g. Lehman et al. 1991). In our sequence data, the distinctive sequence of the Chihuahuan raven is inconsistent with recent hybridization. It is possible that our findings could represent hybridization between Chihuahuan and common ravens hundreds of thousands of years ago; additional microsatellite data will allow further testing of this alternative hypothesis. However, modes of speciation in which one species buds off from another will result in paraphyletic patterns (Harrison 1998); there is no reason to doubt that speciation could explain the paraphyly of the common raven.

(d) Are the two clades of common raven separate species?

The level of divergence between the California and Holarctic clades of common raven is a level typical for many pairwise differences between well-recognized bird species pairs (Klicka & Zink 1997). The mitochondrial haplotypes form two monophyletic groups that are strongly supported by bootstrap analysis. Both of these factors could suggest the recognition of two distinct species, especially following the phylogenetic species concept (e.g. Zink & McKitrick 1995).

However, more data are needed before the recognition of two species is warranted; for example, we need to know whether there are phenotypic differences between the two clades. Furthermore, several aspects of our findings suggest that the two clades of common raven are not
behaving as separate biological species (Mayr 1942). First, representatives of the two clades were mixed within the same populations in California, Idaho and two Washington populations. Second, these four populations are all separated from each other by as much as 1500 km over a huge area of the western United States. Both of the individuals sequenced in New Mexico as well as one California individual have Holarctic clade mtDNA, showing that there is not simply a narrow contact zone in the Pacific North-West of the United States, but possibly extensive introgression. Future research should be directed towards determining which of the myriad properties of species (Avise & Wollenberg 1997; de Queiroz 1998) are present in the two groups of ravens.

Our findings do correspond to some aspects of the subspecific taxonomy of common ravens in North America. Within the western United States, several authors (Oberholzer 1992; Willett 1941; Phillips 1986) suggested that ravens in coastal states from Washington, Oregon, California and Baja California would be a distinct subspecies, *Corvus corax clarionensis*. Our findings to date are consistent with the possibility of a distinct population in this region. (Fig. 5 in Phillips (1986) corresponds well with that aspect of our results.) However, none of these authors suggested that West Coast (US) ravens were dramatically distinct from all other common ravens in the western US and throughout the world, as suggested by our data.

The position of the Chihuahuan raven nested within the common raven makes it tempting to split or lump taxa to create monophyletic species. However, evolutionary processes at the species level may result in patterns that are not conducive to the designation of monophyletic species (e.g. Rieseborg & Brouillet 1994; Harrison 1998). As just discussed, more information is needed to evaluate the taxonomy of this group. If there is free gene flow between the California and Holarctic clades with intermixing of genotypes across a large area of the western US, then perhaps the common raven should not be split. On the other hand, the distinctive size, neck feathers, call characteristics and ecology of the Chihuahuan raven (Goodwin 1976) suggest that it should retain its designation as a separate species. Gene flow between the two clades of common raven, but not between Chihuahuan raven and common raven would not be surprising; phylogenetic relationships do not necessarily predict which taxa hybridize (e.g. Gill et al. 1993; Zink & McKitrick 1995; Burns 1998; Omland et al. 1999).

If the two common raven clades are randomly interbreeding and genes are intermixing completely across large areas of the western US, then this may be a striking example of geographical isolation and molecular differentiation, followed by subsequent remixing of the two gene pools (also see Fleischer & Rothstein 1988; Quinn 1992; Shapiro 1998). The geographical distribution and phylogenetic patterns of ravens make them an excellent group for studying the process of speciation, or possibly the processes that allow separate populations to merge and not differentiate as species.

The following individuals and institutions assisted us with tissue loans: Robert Zink and John Klicka, University of Minnesota Bell Museum of Natural History; Frederick Sheldon and Donna Dittmann, Louisiana State University Museum of Natural Science; John Bates and David Willard, Field Museum of Natural History; Patricia Parker, Ohio State University; Scott Edwards, Siwert Holwer and Chris Wood, Burke Museum of the University of Washington; and Donald Caccamise, New Mexico State University. Keya Nikki Jenkins assisted in the laboratory with the sequencing of cytochrome b, Carla Cicero, Kevin Johnson, Robert Lovich, Peter Marra, Townsend Peterson, Sonja Schieffer, Robert Zink and our colleagues in the Molecular Genetics Laboratory at the National Zoo provided assistance or helpful comments on the manuscript. J.M.M. thanks Bernd Heinrich for support of field research in Maine, and thanks the many individuals who helped with fieldwork in Maine, Idaho and Washington. This work was funded in part by the US Geological Survey, US Army ( Ft Irwin), US Air Force (Edwards AFB), US Fish and Wildlife Service, Friends of the National Zoo and a Smithsonian Molecular Evolution Fellowship to K.E.O.

REFERENCES


