INTERNATIONAL BIRD STRIKE COMMITTEE
IBSC 24/WP 9

Stara Lesna, Slovakia, 14 -18 September 1998

IDENTIFICATION OF BIRDSTRIKE REMAINS BY DNA ANALYSIS

J.R. Allan* C. Conyers" A. A1acNicholl" A. Baxter*

Central Science Laboratory
*Birdstrike Avoidance Team
#Biochemistry & Molecular Genetics Team
Sand Hutton
York, Y041 1LZ
UK

Summary

Engineers and bird controllers need to know which species of birds are being struck by aircraft. This allows the levels of damage to be related to bird weight and bird management to be targeted at the right species. Traditionally, identification has been achieved by visual comparison of whole feather remains or microscopic examination of feather structures. If the feather remains are limited, or if only blood smears remain identification to species may not be possible.

CSL has been developing the use of DNA sequences from the cytochrome-b gene for bird remains analysis. Comparison of material from birdstrike incidents with library sequences shows that a match of 97-99% is possible if a sequence from the same species or a congeneric species is available. Birds from the same family give matches of S7-95%. More distantly related species cannot be matched reliably. Thus, for this technique to be successful, a library of sequences of commonly struck species needs to be developed, so that there are at least examples of the family, and preferably the genus, of birds likely to be struck.

This system has the advantage that it works on any organic material that contains DNA and gives precise specific identification if the DNA sequence is in the comparison library (double blind tests of unknown material were 100% successful). The disadvantage is the cost required to set up the comparison library. We estimate that around 100 sequences would be required to cover the families of birds commonly struck in Europe at a cost of around US$ 15,000. Individual samples would then cost around US$ 150 each to process.

Keywords: Identification. DNA analysis

1. Introduction

The accurate identification of bird remains following birdstrike incidents is important for two reasons. Firstly engineers need to know what weight of bird was involved in a collision so that they can relate levels of damage to bird mass and impact speed and thus identify possible weaknesses in components. Secondly, those involved in birdstrike prevention need to know what species are most frequently involved in birdstrikes and which cause the greatest $damag_e$ so that steps can be taken to implement the most appropriate control techniques for the species involved. 'there is also an increasing trend towards using detailed species breakdown of the birdstrike records of airports to evaluate the performance of their bird control measures over time for which accurate identification to species level is needed.

All of the foregoing require bird remains to be identified as accurately as possible. but in the United Kingdom, where birdstrike reporting is generally good, and where there is a free expert identification service available to airports, the proportion of strikes where the bird species is reported as `unknown' is around 21% for civil birdstrikes and 54°% for military. Many of these unknowns result from strikes that occurred in the airport approaches or during low level military operations when no bird remains were found, and when there

were no feather remains left on the aircraft on subsequent inspection (blood smears or tissue remains are frequently reported but no identification is usually made). Unidentified strikes also occur when any feather remains that are found are too fragmentary to allow identification or when only blood smears or other non-feather tissue remains are present.

At present remains identification is based on comparison of whole feather remains with skin collections or on the microscopic examination of the downy barbules of feathers, which are known to have characteristic structures for different orders of birds. Microscopic examination requires considerable expertise to be carried out effectively, and even the most expert in the field may fail to separate birds to species, especially gulls (Lai-us), hawks *(Accipiteriformes)* and passerines *(Passeriformes)* if only fragmentary remains or non-downy feathers are present. For example, in 1995 only 56% of birdstrikes to civil aircraft in the UK were identified to species level. The frequency with which some of the difficult groups are struck i Lulls constitute 28.4% of the UK birdstrike sample) and the wide weight ranges involved (passerines range from the Goldcrest *(Regulus regulus)* weighing 5.6g- to the Raven *(Corv*[u]*s* corax) weighing II05g in the UK) mean that failure to identify the remains accurately imposes severe limitations on the usefulness of the birdstrike database.

Other techniques have been developed in an attempt to overcome these problems, of which leather protein electrophoresis has been the most successful. However this method still requires the availability of an adequate quantity of feather material which may allow a ;microscopic identification to be made anyway. and it cannot identify tissue or blood remains

An alternative to the techniques currently used is to employ DNA analysis to identify birdstrike remains. This has the advantage that only tiny quantities of material are required, and that any cells that have recoverable DNA (even dried blood smears on canopies etc.) can be used to accurately identify the bird species involved.

Most of the DNA present in a cell is contained in the nucleus, and is reassorted during sexual reproduction. Nuclear DNA is thus unique to each individual and is used in DNA fingerprinting etc.. DNA is also present in other parts of the cell, including the mitochondria. This DNA is passed unchanged from mother to offspring and is only modified by occasional chance mutation over long time periods. For example, Krajewski & King (1996) estimated a rate of change of 0.7 - 1/7% per million years since species divergence in the cytochrome-b gene of Cranes *(Gruidae)*. Thus, the mitochondria) DNA of individuals of the same species will be virtually identical, whilst that of more distantly related species will differ by a greater amount, depending upon the length of time, and hence number of mutations, since the two species became distinct.

The mitochondria) cytochrome b gene is commonly used as a means of identifying organisms, based on the sequence of nucleotides coding for the gene, which is unique to each species (Kocher *el* al., 1989; Bartlett and Davidson, 1992). A segment of the gene can be amplified from a DNA template using the polymerase chain reaction (PCR). The product of the PCR reaction can then be sequenced, and this sequence used to search a database of al l cytochrome b nucleotide sequences for the closest match. A match of 100% will identify the exact species from which the DNA came. Hence, in situations where no feathers are available to enable bird identification, or where identification from feathers is inconclusive, a sample of tissue or blood may be used to obtain a sample of DNA, which can then be used for cytochrome b gene amplification and sequencing. The use of PCR means that only a very small amount of DNA, and hence tissue sample, is required for analysis.

This paper describes a preliminary study undertaken at the Central Science Laboratory to evaluate the use of mitochondria) DNA for bird remains identification.

2. Methods

DNA extraction from each bird sample was carried out according to the proteinase-K digestion and phenol: chloroform extraction method of Sambrook *et al*., (1989). The DNA was resuspended in TE buffer (0.1 M Tris-HCI, 0.01 M EDTA, pH 8.0) at a final concentration of 100 ng ul^{-1} The PCR reaction was carried out in a total reaction volume of 50 ul containing 100 ng template, 1.2 uM forward and reverse primer, 2.5 U Taq DNA polymerase, 4 mM $MgCl_2$, 0.2 mM deoxynucleotidetriphosphates (dNTPs), 20 mM Tris-HCI pH 8.4, and 50 mM KCI. PCR was performed using a Techne Cyclogene thermal cycler with the following thermal cycling profile: 94 °C for 2 minutes (initial denaturation of double stranded DNA), followed by 35 cycles of: 94 °C for 1 minute (denaturation of double stranded DNA), 57 °C for 30 seconds (annealing of primer to DNA template), 72°C for 1 minute (extension of new DNA strands).

Complete extension of all new DNA strands was achieved by incubating at 72 °C for 10 minutes, after which the samples were held at 5 °C until PCR products were separated by electrophoresis through a 1% agarose gel containing 0.5 ug ml^{-1} ethidium bromide in order to visualise the DNA bands. PCR products of the correct length (358 bp) were excised from the gel, and the DNA eluted using a QIAEX gel extraction kit. The extracted PCR product was then directly sequenced by SequiServe (Germany). The nucleotide sequences were used to search GenBank to narrow down the possible bird species, and were then compared directly, over the 307 base pair core sequence to likely matching sequences using the GeneStream Align programme.

Further comparisons were run using both the data obtained by CSL and the sequences available on GenBank (a library of sequences available on the internet) to determine the range of percentage identity likely to be encountered for birds related at the genus, family, order and class levels This was carried out to determine if a certain level of percentage identity could be used to assign sequences which did not have a 100% match (i.e. the same species was not on the reference database) to a genus or family with reasonable certainty.

3. Results

3.1 Comparisons where a sequence of the same species is present on the database

A series of double blind tests were conducted using material from different individuals of the same species, both to determine the reliability of the technique and to discover if any differences existed in the cytochrome-b gene between individuals of the same species. In all cases, there was a 100% match between conspecific individuals of all species tested, showing that the technique is reliable and that, although a mutation in the cytochrome-b gene sequence could occur in an individual or its ancestors and hence give a match of less than 100% with another of the same species, the probability of encountering such an individual in a birdstrike is so small that it can probably be safely disregarded. The possibility of natural mutations, and hence different gene sequences, occurring in populations of the same species that have been genetically isolated for a considerable time should, however, be bore in mind if remains are to be identified using comparison sequences obtained from different parts of the world. Such differences have been found in populations Goose Barnacles (Van Syoc 1994) and some rodents (Smith & Patton 1991)

3.2 Comparisons where a sequence of the same species is not present on the database

If there is not a 100% match for an unknown sample when it is compared with the sequences on the database, assigning the sample to a particular taxon may be possible by finding the best match on the database and comparing the level of that percentage match with the level normally found in birds of the same genus, same family etc.

To illustrate this point, examples of the alignments 15 between a notional unknown sample (in fact it is Black-headed Gull I *Larus ridibundus)* and 3 other species of different levels of relatedness are shown in Figures I to 3.

Figure I shows the, percentage identity of the cytochrome b sequence between two birds within the same genus *(Laridae)* - Black Headed Gull *(Larus ridibundus)* and Lesser Black Backed Gull *(Larus* fuscus). The match is high at 96.1% because they are closely related species, having diverged from a common ancestor relatively recently.

Figure 1. Alignment of cytochrome b gene sequences from Lesser Black Backed Gull *Larus fuscus* and, Black Headed Gull, *Larus ridibundus*.

Numbers above the sequence indicate the gene base number.

```
                100         110         120         130         140
L.ridibundus    CTTCGGGTCCCTACTAGGCATTTGCCTACTAACACAAATCCTAACAGGAC
                     *
L.fuscus        CTTCGGATCCCTACTAGGCATTTGCCTACTAACACAAATCCTAACAGGAC

                150         160         170         180         190
L ridibundus    TCCTACTAGCCATACACTACACCGCAGACACAACCCTAGCCTTCTCATCC
                    *       *       *
L fuscus        TCCTGCTAGCTATACATTACACCGCAGACACAACCCTAGCCTTCTCATCC

                200         210         220         230         240
L.ridibundus    GTCGCCCACACATGTCGAAACGTACAATACGGCTGACTAATCCGAAACCT
                                                    *
L. fuscus       GTCGCCCACACATGTCGAAACGTACAATATGGCTGACTAATCCGAAACCT

                250         260         270         280         290
L. ridibundus   CCACGCAAACGGAGCATCATTCTTCTTTATTTGTATTTACCTACACATCG

Lfuscus         CCACGCAAACGGAGCATCATTCTTCTTTATTTGTATTTACCTACACATCG

                300         310         320         330         940
L ridibundus    GACGAGGATTCTACTACGGCTCATACCTCTACAAAGAAACCTGAAATACA
                                *                   *
L.fuscus        GACGAGGATTCTACTATGGCTCATACCTCTATAAAGAAACCTGAAATACA

                350         360         370         380         390
L.ridibundus    GGAGTCATTCTCCTCCTAACCCTAATAGCAACTGCCTTCGTAGGATATGT
                                *                                   *
L. fuscus       GGAGTCATTCTCCTCTTAACCCTAATAGCAAC7'GCCTTCGTAGGGTATGT

                400
L. ridibundus   ACTACCA

L.fuscus        ACTACCA
```

* 9 out of 307 nucleotides were non-identical giving a 96.1% identity

Figure 2 shows the percentage identity between birds from different orders - Black Headed Gull *(Larus ridibundus) (Charadriiformes)* and Carrion Crow *(Corvus corone) (Passeriformes)*. In this case, the match is only 84.7%, reflecting the greater time since the species diverged.

Figure 2. Alignment of cytochrome b gene sequences from Black Headed Gull. *Larus ridibundus* and Carrion Crow, *Corvus corone.*

Numbers above the sequence indicate the gene base number.

```
                100           110           120           130           140
L.ridibundus      CTTCGGGTCCCTACTAGGCATTTGCCTACTAACACAAATCCTAACAGGAC
                    *   *                 *  *        *  *  *          *  *       *
C.corone        CTTTGGATCCCTACTAGGCCTATGCCTAATCATACAAATCATTACAGGTC

                150           160           170        180           190
L ridibundus     TCCTACTAGCCATACACTACACCGCAGACACAACCCTAGCCTTCTCATCC
                   *               *           *        *    **              *  *   *
C.corone        TGCTACTAGCCATGCACTACACAGCAGATACCTCCCTAGCCTTCGCTTCT

                200           210           220           230           240
L.ridibundus      GTCGCCCACACATGTCGAAACGTACAATACGGCTGACTAATCCGAAACCT
                    *           *     *     *               *     *
C.corone        GTAGCCCACATATGCCGAGACCTACAATTCGGATGACTAATCCGAAACCT

                250           260           270           280          290
L. ridibundus    CCACGCAAACGGAGCATCATTCTTCTTTATTTGTATTTACCTACACATCG
                                  *   *        *   *        *   *
C.corone        CCACGCAAACGGAGCTTCCTTCTTTTTCATTTGCATCTACCTACACATCG

                300           310           320           330           940
L ridibundus      GACGAGGATTCTACTACGGCTCATACCTCTACAAAGAAACCTGAAATACA
                    *           *           *           *                    **
C.corone        GCCGAGGATTTTACTACGGTTCATACCTAAACAAAGAAACCTGAAATATC

                350           360           370           380           390
L.ridibundus     GGAGTCATTCTCCTCCTAACCCTAATAGCAACTGCCTTCGTAGGATATGT
                        *   *     *           *                              *     *
C.corone        GGAGTAATCCTTCTCCTAACACTAATAGCAACTGCCTTCGTAGGCTACGT

                400
L. ridibundus     ACTACCA
                 *     *     *
C.corone        TCTGCCT
```

* 47 out of 307 nucleotides were non-identical giving a 84.7% identity.

Figure 3 shows the percentage identity between two animals from different classes: Black Headed Gull *Larus ridibundus (Aves)* and Human Homo *Sapiens (Mammalia),* which by comparison is a low match of 71.7 %.

Figure 3. Alignment of cytochrome b gene sequences from Black Headed Gull, *Larus ridibundus* and Human, *Homo sapiens*.

Numbers above the sequence indicate gene base number.

```
                100           110           120           130           140
L.ridibundus     CTTCGGGTCCCTACTAGGCATTTGCCTACTAACACAAATCCTAACAGGAC
                       *   *   *   *      ***        ** * * * *         ***
H.sapiens       CTTCGGCTCACTCCTTGGCGCCTGCCTGATCCTCCAAATCACCACAGGAC

                150           160           170           180           190
L ridibundus    TCCTACTAGCCATACACTACACCGCAGACACAACCCTAGCCTTCTCATCC
                  ** *           *          *  * *       *  * *  *** *       *       *
H.sapiens       TATTCCTAGCCATGCACTACTCACCAGACGCCTCAACCGCCTTTTCATCA

                200           210           220           230           240
L.ridibundus     GTCGCCCACACATGTCGAAACGTACAATACGGCTGACTAATCCGAAACCT
                *             * ** *      *        *  *    *         *  *       **
H.sapiens       ATCGCCCACATCACTCGAGACGTAAATTATGGCTGAATCATCCGCTACCT

                250           260           270           280           290
L. ridibundus   CCACGCAAACGGAGCATCATTCTTCTTTATTTGTATTTACCTACACATCG
H.sapiens       TCACGCCAATGGCGCCTCAATATTCTTTATCTGCCTCTTCCTACACATCG

                300           310           320           330           940
L ridibundus    GACGAGGATTCTACTACGGCTCATACCTCTACAAAGAAACCTGAAATACA
                  *       * *  *   *       *       **         **                  *  **
H.sapiens        GGCGAGGCCTATATTACGGATCATTTCTCTACTCAGAAACCTGAAACATG

                350           360           370           380           390
L.ridibundus     GGAGTCATTCTCCTCCTAACCCTAATAGCAACTGCCTTCGTAGGATATGT
                   ** *    *       *   **  ****           *       *      *
H.sapiens       GGCATTATCCTCCTGCTTGCAACTATAGCAACAGCCTTCATAGGCTATGT

                400
L. ridibundus    ACTACCA
                *   *    *
H.sapiens       CCTCCCG
```

* 87 out of 307 nucleotides sere non-identical giving a 71.7% identity

If a large number of matches are undertaken, it is possible to estimate the maximum percentage match for birds of the same order but different family. same family but different genus or same genus but different species. If the best match obtained exceeds the maximum value for a particular degree of relatedness an unknown bird can be assigned to the next level taxon with a reasonable degree of certainty. Table 1 shows the maximum, minimum and mean percentage identity for a series of 41 sequence comparisons between birds of different taxa.

Table 1

	Same Species	Same genus but different species	Same Family but different genus	Same Order but different family	Different Order
Min Relatedness (%)	100	87.6	86.6	82.7	79.5
Max. Relatedness (%)	100	99.0	91.7	87.0	85.8
Mean Relatedness	100	93.8	89.8	85.1	82.1

The data shown in Table 1 are preliminary and limited in extent (no category contains more than 10 comparisons), but if these figures are representative of the full dataset, a simple rule can be formulated where, if the best percentage match was greater than the maximum percentage match within a class, the unknown bird could be assigned to a higher class viz:

Identity > 85.8%	Same order
Identity > 87%	Same family
Identity > 91.7%	Same genus
Identity 100%	Same species

For example, the notional unknown bird in table 1 has a best percentage match of 96.1 % with Lesser Black-backed Gull. This would be sufficient to assign it to the same genus *(Larus) as* it exceeds the maximum value yet recorded for birds of different genus but the same family (91.7%).

These data, and hence the cut-off percentages for assigning unknown birds to taxa, would obviously change with time as more comparisons became available. It would also be possible to re-compare birds previously allocated to family, genus or order level when new reference material was obtained and so refine the identifications retrospectively.

4 Discussion

4.1 Potential for the use of mitochondria) DNA in remains analysis

The routine use of mitochondria(DNA presents the first opportunity to advance the level of accuracy in birdstrike remains identification for many years. Providing that a DNA sequence from the same species or genus is present in the library of comparison traces, an accurate identification to the species or genus level is possible even if the tiniest amounts of flesh or blood smears are present on the aircraft. CSL is in the process of developing a simple kit to allow engineers to quickly and easily take swabs from blood smears for DNA analysis. If this ystem is implemented, there should be no reason not to submit a sample following a birdstrike providing that the impact point on the aircraft can be identified. No system will eliminate all 'unknowns' from the birdstrike sample, but the provision of DNA analysis should ensure that there is no excuse for failing to identify the species involved whenever remains are found.

4.2 Limitations of the system

The main limitations to this system are the need to extract DNA from feathers when no tissue or blood smears are present, and the costs involved in setting up the library of reference sequences.

4.2.1 Extraction of DNA from feathers

Feathers consist largely of keratin and grow out from living cells at the base of the feather close to the point of insertion in the wing. There is little information available concerning whether the tips of feathers contain any mitochondrial DNA and how this can be extracted and sequenced. Other research is in progress in the UK to address this issue, but until the results of this work are available it may prove impossible to conduct a DNA analysis if, for example, only a feather tip is available. It is likely, however, that blood smears will be found following an impact that leaves feather remains on the aircraft and an analysis could be carried out on these if needed.

4.2.2 Need for a full reference database

It is obviously not possible to identify a bird to the species level unless there is a sample from the same species already on the database to match with. To cover every species regularly found in the UK would therefore require a library of around 200 sequences to be established. However, it is frequently only necessary to establish identification to genus when the different species within that genus have similar lifestyles (and so similar bird control requirements) and/or similar weights. To enable accurate identification to genus of all regularly occurring birds in the UK, a reference sample from each genus would be needed. A library of 119 sequences would thus be required. If those species which are highly unlikely to be struck are excluded on a pragmatic basis, the number of traces needed falls to around 64. Additional sequences would be required to allow identification to species level within genera that are frequently struck and where either the ecology of the species varies (e.g. Woodpigeon and Feral Pigeon, *Columba)* or the weights vary widely (e.g. Black-headed Gull and Great Black-backed Gull, *Larus*). A library of around 80 sequences, pragmatically selected, would allow identification, to the necessary accuracy, of any bird likely to be struck in the UK. It should be possible to obtain the samples required from museum skins, birdstrikes, roadkills etc. thus avoiding the need for collection of birds from the wild in most cases. If access to skins in large museum collections can be arranged, the library of sequences can be expanded to include other countries or continents with relative ease.

5 Conclusion

Providing that a technique for extracting DNA from feathers can be developed successfully, sequencing of the cytochrome-b gene from the mitochondrial DNA of birds or other living material can be used as a more accurate and comprehensive identification technique than any currently available. This technique allows even the tiniest fragments of material to be identified to species or genus, the only limitation being the presence of a conspecific or congeneric sample on the reference database. This method also allows blood smears from en-route birdstrikes where no other remains can be recovered to be accurately identified. The cost of setting up the initial reference database is substantial, but if a pragmatic approach to which birds need to be identified to species and which to genus is taken, the cost can be significantly reduced. CSL will continue to seek funding partners for this work with a view to setting up a reference database for UK and, where possible, European, bird species in the near future.

6 References

Bartlett, S.E., amd Davidson, W.S. (1992). FINS (Forensically Informative Nucleotide Sequencing): A procedure for identifying the animal origin of biological specimens. BioTechniques 12: 408-411 (See erratum, BioTechniques 13: 518)

Kocher, T.D., Thomas, W.K., Meyer, A., Pääbo, S., Villablanca, F.X., and Wilson, A.C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86: 6196-6200.

Krajewski, C. & King, D.G. (1996) Molecular divergence and phylogeny: rates and patterns of cytochrome b divergence in cranes. Mol. Biol. Evol. 13:21-30.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning – A Laboratory Manual. 2nd Edition. Cold Spring Harbor Laboratory Press, USA.

Smith. M.F. & Patton, J.L. (1991). Variation in mitochondrial cytochrome b sequence in natural populations of South American akodontine rodents (Muridae:Sigmodontinae) Mol. Biol. Evol 8: 85-103.

Van Syoc, R.J. (1994). Genetic divergence between subpopulations of the eastern Pacific goose barnacle Polliceps elegans: mitichondrial cytochrome c subunit 1 nucleotide sequences. Mol. Mar. Biol. Biotechnol. 3:338-346.