Cluster analysis of the genetic heterogeneity and disease distributions in purebred dog populations

G. J. Ubbink, J. van de Broek, H. A. W. Hazewinkel, J. Rothuizen

Purebred dog populations have been subject to strong selection which has resulted in extreme differences between breeds and decreased heterogeneity within breeds. As a result, breed-specific inherited diseases have accumulated in many populations. The aim of this study was to analyse genetic heterogeneity in relation to the distribution of elbow dysplasia in labrador retrievers, portosystemic shunts in Irish wolfhounds, and hepatic copper toxicosis, in Bedlington terriers. Decreased heterogeneity was demonstrated in the multiple genetic interrelations in the three populations. In pedigrees containing seven generations of ancestors, the average number of common ancestors in all pair-wise combinations of dogs was five to six (range 0 to 18). These complex interrelationships were resolved by a cluster analysis on matrices of relatedness. This analysis gave clusters of highly related animals, the average relatedness of these clusters, and the average relatedness of the entire population, as expressions of its genetic variability. The mean relatedness was 0.032 for Irish wolfhounds and Bedlington terriers, and 0.002 for labrador retrievers. The labrador retriever cohort was resolved into 31 clusters, and all cases of elbow dysplasia were concentrated in five highly related clusters with an overall incidence of 17 per cent. The Bedlington terrier cohort consisted of 12 clusters which all contained cases of copper toxicosis, with an overall incidence of 46 per cent. The Irish wolfhounds were divided into 14 clusters with a disease incidence of 4 per cent. Dogs with portosystemic shunts were found in four averagely related clusters. A genetic distribution became obvious only when relatedness due to common ancestors of the cases was used as a criterion, and the cases were then concentrated in five highly related clusters.

PUREBRED dog populations represent genetically homogeneous populations with varying population structures. Since only a small fraction of the population is used as breeding stock, strong founder effects may have been introduced (Bouw 1982). Thus, selection has led to a complex web of relatedness among breed members due to their varying exposures to different founders. A preliminary analysis of the Dutch Kennel Club breeding records of purebred dog populations indicated that only 3 to 5 per cent of all dogs registered in the Netherlands in the last two to three decades have contributed to the present populations. As a result, there has been a permanent reduction of the gene pool which has resulted in a high incidence of certain genetic diseases in different breeds (Ubbink and others 1992). Each breed now has a number of common diseases which are probably associated with particular founders.

Despite the loss of genetic variability suffered by all dog breeds, the web of genetic relationships within breeds may be quite different for different breeds. Knowledge of the genetic variation and the structure of a population in terms of highly related subgroups, in conjunction with the distribution of an inherited disease, may improve the mode of selection against the disease while maintaining the highest possible genetic heterogeneity. If the diseased animals occur in highly related subgroups it may also be possible to

G. J. Ubbink, DVM, H. A. W. Hazewinkel, DVM, PhD, J. Rothuizen, DVM, PhD, Department of Clinical Sciences of Companion Animals, J. van de Broek, PhD, Centre of Biostatistics, Faculty of Veterinary Medicine, University of Utrecht, PO Box 80 154, 3508 TD Utrecht, The Netherlands

Acknowledgements. – Without the financial help of the Grass Sickness Fund who covered many of the expenses involved and of the International League for the Protection of Horses who support the nursing staff, this success story would have been impossible. The authors wish to thank all the nurses and owners who took part in the study for their help, cooperation and hard work. Thanks are due to Mr J. C. Wainwright who submitted intestinal tissues from a case in his care.

References

GREIG, R. J. (1942) Transactions of the Highland and Agricultural Society of Scotland, p. 1
Mclaren, L. (1918) Veterinary Record 30, 473
MILNE, E. M. (1997) In Practice 19, 128
OBEL, A. L. (1955) Journal of Comparative Pathology 65, 334
PINSENT, P. J. N. (1989) Veterinary Record 9, 169

The Veterinary Record, February 28, 1998

209

are normal. Similar remarks apply to sweating which in a few cases also persisted for years, although not continuously. When the horse was calm and at rest, no sweating was evident, but after mild excitement clearly visible, often large patches of sweat appeared, particularly around the mane.

At the outset of this project the authors had no clear idea of the proportion of working horses likely to resume effective work, and to find that over 80 per cent of the respondent’s animals were able to return to their original or originally designated work, was extremely pleasing. Such a high proportion indicates that the recuperative powers of intestinal neurons must be considerably greater than what was considered possible a decade ago. Of the 42 treated cases of dysautonomia, six have died and five could not be traced. The remaining 31 are well and have clinically normal intestinal function despite the fact that they have suffered a severe depletion of enteric neurons. It seems likely either that the remaining neurons have taken on extra duties, or that some degree of neuron regeneration has occurred. It is not possible from present evidence to prove that neurons have regenerated but enteric neurons isolated from adult horses into tissue culture can be stimulated to divide and grow (H. John, personal communication); the possibility of some neuronal regeneration is therefore not as unlikely as has previously been suggested.

The jejunum of the horse from which tissue was received was distended and had a thickened wall. It is possible that jejunal hypertrophy may have occurred to compensate for the poor function of the ileum.

Much time and effort was put in by the nursing staff and the owners to achieve these recoveries. Owners may have to treat a case as an individual to be specially pampered for many weeks or even months, and to do this in an establishment run as a business is especially difficult. It is therefore gratifying that all the owners consulted considered their efforts to have been worthwhile. Grass sickness is still a major cause of mortality, especially in Scottish horses, but the prognosis is good for a proportion of chronic cases whose owners are prepared to work long and hard to help them survive.
recognise the otherwise unnoticed inherited nature of a disease (Peters and Ubbink 1994).

In this study, a cluster analysis has been applied to the evaluation of the genetic structure and heterogeneity of certain breeds, on the basis of the pedigree-registrations held by the kennel clubs. It is a hierarchical analysis of a matrix of the estimated overall genetic similarity of all existing relations, which unravels the population in terms of the number of related clusters and estimates of the degree of relatedness of these clusters. Because of the theoretical disadvantages of hierarchical methods, the procedure has been validated by comparing it with a non-hierarchical method. The method has been applied to three Dutch dog populations with different population structures, each having a specific inherited disease. These populations were Irish wolfhounds with liver failure due to inherited portosystemic shunts with an incidence of 4 per cent (Meyer and others 1994), Bedlington terriers with copper toxicosis due to impaired hepatic excretion of copper into bile (Johnson and others 1980) with an incidence of 46 per cent and labrador retrievers (used as guide dogs for the blind) with elbow dysplasia due to fragmented coronoid process with an incidence of 17 per cent (Hazewinkel and others 1988).

Materials and methods

Pedigrees were obtained from Kennel Club Stud Books of international breeding associations listed in the 'Federação Cynológica Internacional' and the Kennel Club. For all the dogs in the cohorts studied pedigree information was available for seven generations.

Populations screened

Cluster analysis was applied to three Dutch registered populations, and all the animals in the cohorts under study were tested by standardised protocols to evaluate the presence of disease. The populations were:

1) a labrador retriever subpopulation, owned by the Royal Dutch Guide Dog for the Blind Association. All the 252 dogs owned (bred and imported) by the association between January 1, 1988 and December 31, 1992 were screened for clinical signs of fragmented coronoid process causing chronic arthrosis of the elbow joint (Hazewinkel and others 1988) by two experienced dog trainers. The dogs with clinical signs were further examined, their elbows were radiographed (Voorhout and Hazewinkel 1987), and the disease was treated by arthroscopy. The clinical incidence in this cohort was 17 per cent. The mode of inheritance of the disease is unknown (Guthrie and Piddock 1990, Gröndalen and Lingaa 1991);

2) the 613 Irish wolfhounds born and registered between January 1, 1985 and December 31, 1992. They were screened for the presence of a congenital portosystemic shunt by measuring the concentration of ammonia in plasma at six to eight weeks of age; a concentration above 125 nmM was taken to be a positive diagnosis and the incidence was 4 per cent. The mode of inheritance is unknown (Meyer and others 1994);

3) the 155 Bedlington terriers born and registered between January 1, 1977 and December 31, 1985. The disease in this breed was hepatic copper toxicosis due to defective biliary excretion of copper. The mode of inheritance is autosomal recessive (Johnson and others 1980). They were screened by the measurement of copper in liver biopsies, either by histochemistry or cytochemistry, or by atomic activation analysis (Teske and others 1992). The biopsies were taken when the dogs were one year of age and the incidence of the disease was 46 per cent.

Determination of common ancestry and overall relatedness

In order to find a quantitative measure for the complex sources (founders) of relatedness between individual dogs, the set of common ancestors of the cohort (K_{all}) was determined in each cohort. A common ancestor was defined as an ancestor shared by at least two cohort members to which the common genes were passed on by different descendants (Wright 1922). Secondly, the number of common ancestors (k \in K_{all}) that contributed to each of the relations of cohort members was determined. Thirdly, the extent of relatedness of each relation was determined using the relation-coefficient originally defined by Wright (1922, 1925). All the estimates resulted in matrices of relatedness in which the expected overall similarity of each cohort member to every other member was determined.

Formula (1):

\[ R_{ij} = 2 \sum_{k=1}^{K_{all}} \frac{[\text{K}_{ij}'_k]_k}{\sum_{z=1}^{Z} (1 + F_{z})} \]  

In which, \( R_{ij} \) = the genetic similarity between cohort members \( i \) and \( j \), \( k = \) the common ancestor of \( i \) and \( j \), \( z = \) the theoretical pathway in which an allele of \( k \) could be inherited by both \( i \) and \( j \), \( n_{1z} \), \( n_{2z} \) = the number of generations separating \( i \) and \( j \), respectively, from ancestor \( k \) in pathway \( z \), \( F_{z} = \) the inbreeding coefficient of ancestor \( k \), \( K_{all} = \) the set of common ancestry of all cohort members.

Formula (2):  

\[ F_k = \frac{1}{2} R_{ab} \]  

in which, \( F_k = \) the inbreeding coefficient of ancestor \( k \), \( R_{ab} = \) relatedness estimate of \( a \) and \( b \), the parents of \( k \), estimated as in formula (1).

Determination of relatedness due to specific founders

In order to focus more specifically on the disease, a subset within the set of founders (K_{all}) was selected. This subset consisted of all ancestors that contributed to the relatedness between cases (K_{cases}).

Formula (3) \( K_{cases} \subseteq K_{all} \)

Thus the extent of relatedness between the dogs in the cohort caused by common ancestors of cases could be estimated for ancestors \( k \) with \( k \in K_{cases} \) using formula (1).

Cluster analysis of overall (\( k \in K_{all} \)) and specific relations (\( k \in K_{cases} \))

A hierarchical cluster analysis with average linkage between clusters was applied to the matrices of relatedness. Animals with the highest genetic similarity were grouped in clusters, defined as animals having an average relationship \( > 0.125 \). The average relatedness between clusters was displayed in dendrograms.

Validation of the hierarchical cluster analysis

In a hierarchical procedure the order in which data are incorporated into the cluster analysis, and/or the lack of variation in the data, may influence the reproducibility of the final results. The reproducibility was analysed by reading the relatedness matrices in eight different sequences, using data sets derived from two, three, four, five, six and seven-generation pedigrees. These figures did not include the individual dog's generation. The resulting clusters were compared with those obtained using a non-hierarchical K-means method (Johnson and Wichern 1992), which is not sensitive to the order in which data are entered. The number of clusters used in the K-means method was equal to the number of hierarchical clusters. Inaccuracies were defined as the fraction of the animals in a cluster that was added to a different cluster in at least one or more of the eight analyses by the hierarchical method.

To test the descriptive resolution of the hierarchical procedure the fraction of animals that was added to a different cluster was determined by comparing the results based on seven generations with the results based on six, five, four, three and two generations.
FIG 1: Relative frequencies of the numbers of ancestors that contributed to the genetic similarity of pairs of dogs in the three populations

Results

Common ancestry in the cohorts

The total number of common ancestors in the three populations varied (Table 1). In each cohort there was a complex network of genetic interrelationships. As shown in Fig 1, the median numbers of shared ancestors per combination of labrador retrievers, Irish wolfhounds, and Bedlington terriers were five (range 0 to 15), five (range 0 to 18) and six (range 0 to 17), respectively. Thus, in all three cohorts there were dogs that seemed to be unrelated on the basis of seven-generation pedigrees, whereas other dogs had 15 to 18 ancestors in common.

Validation of hierarchical cluster analysis

The influence of the order of the entry of the data was strong when estimates of relatedness based on only a few generations were incorporated, owing to the limited variation in the estimates, but the differences between the hierarchical and the non-hierarchical analysis decreased when more generations were incorporated. When pedigrees of five or more generations were used the clusters of relatedness were nearly identical in both analyses, with less than 5 per cent of the dogs being assigned to a different cluster (Fig 2a). When in the hierarchical procedure relatedness was determined by using two, three, four, five, six or seven-generation pedigrees, the results were virtually identical for five, six and seven generations, but the results based on pedigrees containing four or fewer generations differed considerably from those based on seven generations (Fig 2b).

Cluster analysis of overall relatedness of the three populations

The results of cluster analyses using seven-generation pedigrees and overall relatedness estimates (Kcase) were quite different in the three cohorts (Figs 3a, 4a and 5a). The labrador retriever population showed considerable heterogeneity; there were 31 small clusters with a low degree of relatedness between the clusters. In the Irish wolfhounds and the Bedlington terriers 14 and 12 clusters, respectively, were found with a high degree of relatedness between the clusters, indicating less genetic variability. The distribution of diseased animals among the clusters was different in the three populations (Figs 3a, 4a and 5a). The labrador cases were concentrated in five highly related clusters; in the Irish wolfhounds portosystemic shunts occurred in four clusters which were no more than averagely related; in the Bedlington terriers copper toxicosis occurred in all the clusters.

Cluster analysis based on relatedness with common ancestors of cases

The numbers of common ancestors of cases (Kcase) in the three cohorts are shown in Table 1. In the labrador retriever population cluster analysis based on relatedness with common ancestors of cases showed an increased difference between the relatedness of clusters at risk and other clusters (Fig 3b). In the Irish wolfhounds a difference was found in the relatedness of clusters with disease and clusters without disease which had remained unnoticed when the analysis was based on overall relatedness (Fig 4a and b). In the Bedlington terriers the results of the two types of cluster analysis were similar (Figs 5a and b).

TABLE 1: Populations of dogs subjected to a hierarchical cluster analysis of genetic heterogeneity and disease prevalence

<table>
<thead>
<tr>
<th>Birth cohorts</th>
<th>Number studied</th>
<th>Population</th>
<th>Kophilia</th>
<th>Kcase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labrador retriever</td>
<td>252</td>
<td>386</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(1988-1992)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irish wolfhounds</td>
<td>613</td>
<td>3283</td>
<td>109</td>
<td>97</td>
</tr>
<tr>
<td>(1985-1992)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bedlington terriers</td>
<td>155</td>
<td>2650</td>
<td>182</td>
<td>85</td>
</tr>
<tr>
<td>(1977-1985)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The population is the number of dogs registered by the Dutch Kennel Club, representing the available breeding stock. The population of the labrador retrievers could not be established because this cohort was only a fraction of the entire Dutch population.
† Numbers of common ancestors of the cohorts (Kophilia).
‡ Subsets of common ancestors of the affected animals (Kcase).

FIG 2: a) Influence of the order of input of data and of the extent of pedigree information when applied in the hierarchical average-between-cluster analysis. The histogram shows the fraction of dogs that was added to the same cluster as in the non-hierarchical method in all of the eight sequence orders tested. When five or more generations were used there was a negligible difference between the non-hierarchical and the hierarchical methods. b) The fraction of the dogs that was added to another cluster when data containing pedigrees of two, three, four, five and six generations of ancestors were compared with the results based on seven generations. The clusters obtained with five, six or seven generations were virtually identical.
FIG 3: a) The labrador retriever population depicted in a dendrogram of the clusters found when seven generations of ancestors of the entire population were used. Clusters were formed when dogs had a relatedness of \( \geq 0.125 \). The fraction of the dogs with fragmented coronoid process is indicated in black, and the percentage of the affected dogs is given below these clusters. b) The labrador retriever population when the cluster analysis was based only on ancestors of affected dogs.

Discussion

This study has evaluated cluster analysis as a method for studying the relatedness of purebred dog populations, for resolving the complex structure of these populations to estimate their level of relatedness and genetic heterogeneity, and for assessing the genetic background of diseases.

A hierarchical clustering procedure was applied because this approach makes it possible to find any suitable number of clusters with a given average degree of relatedness (here \( \geq 0.125 \)). Hierarchical clustering thus makes it possible to fit the model to the variable and unpredictable structures of different populations. Its theoretical disadvantage is that the result may be influenced by the order in which the data are incorporated into the analysis and by the extent of variation in the data set. During a hierarchical cluster analysis an animal (or already formed cluster) is added in successive rounds to a cluster with which it has the highest (average) relationship. At any stage during the analysis equal relationships might occur in the matrix, and a random choice between them may influence the hierarchical order. Thus, when the cluster analysis is repeated a different random choice may produce clusters with a different composition. The influence of limited variation and inaccuracies in sequence may be expected to be more pronounced when fewer generations are being used, because with more generations, and hence more variation in the data, there will be less chance that exactly equal relationships may occur. Non-hierarchical systems are not sensitive to the order in which data are introduced, but they require a fixed number of clusters, which is not suitable for analysing populations of unpredictable complexity (Johnson and Wichern 1992). In order to investigate whether there is a critical quantity of data beyond which this potential source of inaccuracy becomes negligible, the results of hierarchical and non-hierarchical cluster analysis have been compared in each of the three cohorts, using relationships derived from two, three, four, five, six and seven generations. The results showed that in the dog populations considered, pedigrees contain-

FIG 4: a) Dendrogram showing the results of clustering in the Irish wolfhound population found when seven generations of ancestors of the entire population were used; the relatedness of dogs within the clusters was \( \geq 0.125 \). The fraction of the dogs with a congenital portosystemic shunt is indicated in black, and the percentage of the affected dogs is given below these clusters. b) The dendrogram of the Irish wolfhound population when the cluster analysis was based only on ancestors of affected dogs.
may have a high incidence, as in the Bedlington terriers. Secondly, as in the Irish wolfhounds, only a few clusters may be affected, but the relationship between these clusters was no greater than the overall relatedness of all the clusters. As shown before, genetic similarity between dogs is based mainly on multiple common ancestry but the relatedness due to the ancestors that actually distributed the risk factors may be overshadowed by other ancestors in the overall relatedness. To overcome this problem a cluster analysis was focused on the common ancestry of the affected dogs. In this population the analysis clearly separated the clusters at risk from the other clusters, and elucidated the genetic background of portosystemic shunts:

The distribution of dogs with a fragmented coronoid process in the labrador retrievers revealed five highly related clusters at risk. The mean risk of the population was 17 per cent but in these five clusters it exceeded 27 per cent, and was 0 per cent for the remaining clusters. In this case, cluster analysis provides a tool to select against the disease directly, while the heterogeneity of the population may remain preserved. In the Bedlington terrier population all the clusters were affected by this known genetic disease. There was little difference between the ancestry of the diseased dogs and that of the overall population. This explains why the two methods of cluster analysis (Figs 5a and b) gave nearly the same result. However, the absence of unaffected clusters would hamper the eradication of the disease by the selection of breeding stock. All the dogs were closely related to affected dogs, and as a result all unaffected dogs are at high risk of being heterozygous carriers.

The above strategies of cluster analysis demonstrate the feasibility of this approach in three purebred dog populations with different population structures and a variety of inherited diseases. It is concluded that a hierarchical cluster analysis provides a better insight into the population structure and genetic heterogeneity of a population of dogs and thus of the available breeding stock. It may make it possible to detect the genetic nature of a disease, when the highly related clusters at risk can be identified. However, the absence of a clear risk distribution does not exclude inheritance, as was shown in the cohort of Bedlington terriers.

A knowledge of the distribution of an inherited disease among clusters of related animals may help to provide the means to select against the disease, while maintaining the heterogeneity of the population. The clusters of dogs at risk may also provide a basis for the selection of animals for molecular genetic studies to search for markers of the genes underlying the disease (Lisitsyn and others 1993, 1994, Nelson and others 1993, Aldous 1994).

Acknowledgements. — The authors would like to thank Dr B. A. van Oost for a critical reading of the manuscript, and Dr W. Nuy for computing the algorithms. The Royal Guide Dog for the Blind Association is acknowledged for its cooperation in the study, and the Dutch Kennel Club for the financial support which made it possible.

References


BOUW, J. (1982) Tijdschrift voor Diergeneeskunde 107, 1


MEYER, H., ROTHUIZEN, J. & UBBINK, G. J. (1944) Veterinary Record 136, 13


PETERS, M. E. & UBBINK, G. J. (1994) Veterinary Record 134, 444


UBBINK, G. J., KNOL, B. W. & BOUW, J. (1992) Veterinary Quarterly 14, 137


WRIGHT, S. (1925) Journal of Agricultural Research 31, 376
Cluster analysis of the genetic heterogeneity and disease distributions in purebred dog populations


Veterinary Record 1998 142: 209-213
doi: 10.1136/vr.142.9.209

Updated information and services can be found at:
http://veterinaryrecord.bmj.com/content/142/9/209

References
Article cited in:
http://veterinaryrecord.bmj.com/content/142/9/209#related-urls

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/