Session G3.23 – (Abstract no. 504)
email: info@geneticdiversity.net

Cluster analysis of kinship in the Iceland Dog as a small pedigreed population

P.A. Oliehoek[1], *G.J. Ubbink*[2] *and R.F. Hoekstra*[3]

[1]*Wageningen-UR, Animal Sciences Group, Marijkeweg 40, 6709 PG Wageningen, The Netherlands,*
[2]*University of Utrecht, Department of Clinical Sciences of Companion Animals, Yalelaan 8, 3584 CM Utrecht, The Netherlands,*
[3]*Wageningen-UR, Laboratory of Genetics, Arboretumlaan 4, 6703 BD Wageningen*

Cluster analysis of kinship can elucidate the population structure, since this method divides the population in clusters of individuals in a dendrogram. Previous research shows that the incidences of breed-specific diseases are bound to specific clusters.
Research has been carried out on the entire Iceland Dog population, a sheep-herding breed. When kinships were calculated up to seven generations, as has been done in previous research, the population split up in 5 clusters, which is much lower than other dog populations. Next cluster analysis was based on kinships calculated up to the founder-population. The results demonstrate that the cluster-analysis dendrogram is rather different when based on all generations instead of seven. This contradicts predictions of previous research. Furthermore, the results suggest that kinship-based clustering reveals animals with low mean kinship (animals that are genetically important for genetic diversity). The research shows that despite increasing population size, the Iceland Dog population lost considerable genetic variation. Average mean kinship of the population was 0.22. The number of puppies per litter decreased significantly with 0.22 per 10% increase of inbreeding coefficient. Cluster analysis of kinship coefficient provides insight in the distribution of the available genetic variation.

1. INTRODUCTION

Genetic variation is the mean expected heterozygosity per locus in a population. Managing genetic variation within populations is necessary for the adaptive potential as well as the fitness of the population. Many dog-breeds suffer from breed-specific (recessive) genetic diseases. Incidence of diseases is increased by the reduction of genetic variation, resulting in unavoidable higher levels of kinship followed by higher levels of inbreeding.

Though it is possible to estimate the relative contribution towards genetic variation of specific individuals within a population (Meuwissen, 1997), it is difficult to get an overview of the population. Especially populations that are not bred by one authority, it is not clear how genetic variation can be maintained.

Ubbink et al. (1998ab, 1999) used cluster analysis of kinship coefficients to elucidate the familial structure of purebred dog populations and to demonstrate the correlation with a genetic disease present in those populations. Instead of 'looking at a large pile of pedigrees' or a table with mean kinship, hierarchical cluster analysis permits the fitting of the model to variable and unpredictable structures of different pedigreed populations (Ubbink et al., 1998a). Using this clustering method, a population is separated in highly related family groups that have a certain level of kinship (relationship) among each other. The family groups represented by specific individuals are demonstrated as bars in the dendrogram. Branches demonstrate the relationship among these family groups. (For further explanation, see appendix).

2. MATERIAL AND METHOD

The following countries have sent data containing information of their population of Iceland Dogs: the Netherlands, Sweden, Iceland, Germany, Norway, Denmark and Finland. These countries cover almost the entire global Iceland Dog population. These data-files were transferred into one database-table.

Then cluster analysis was performed twice on the individuals born in the years 1991 till 1998, once as in the method described in Ubbink et al., (1998): kinship coefficient based on seven generations and once as described in Ubbink et al., (1998) but with kinship coefficients based on pedigrees up to the founders.

A cohort was assembled including all reproductive (fertile-living) individuals. If the reproductive state of an individual was unknown, the individuals were included unless they were older than 11 years for females and 16 years for males. Cluster analysis protocol, with kinships calculated using all generations, was performed on this cohort.

Finally, the population was checked on the presence of inbreeding depression on litter size. Correlation coefficient and vector coefficient was calculated for the inbreeding coefficient (F) and the number of puppies registered per litter. The litters included were pedigreed for at least five generations. Using this method one does not calculate the correlation between the F and the actual number of puppies, but the correlation between F and the number of puppies that survive the first eight weeks.

3. RESULTS

Of the 4680 dogs, 35 did not have any parents registered. All other dogs in the pedigree file descended from these 35 founder animals. The base population or founder stock in this research was defined as all founder animals

The correlation coefficient of number of puppies registered and the inbreeding coefficient of 1218 litters was -0.10374, p is 0.0003 (SAS Institute, 1989). A 10% increase in the inbreeding coefficient resulted in a decrease of 0.22 puppies (p is 0.0003; SAS Institute, 1989).

3.1 Cluster Analysis Methods Compared

The first cohort was composed of all individuals born between 1-1-91 and 1-1-98 and contained 2557 dogs, represented by 512 unique parent combinations. The cohort was analysed twice in order to compare different kinship coefficient calculation methods.

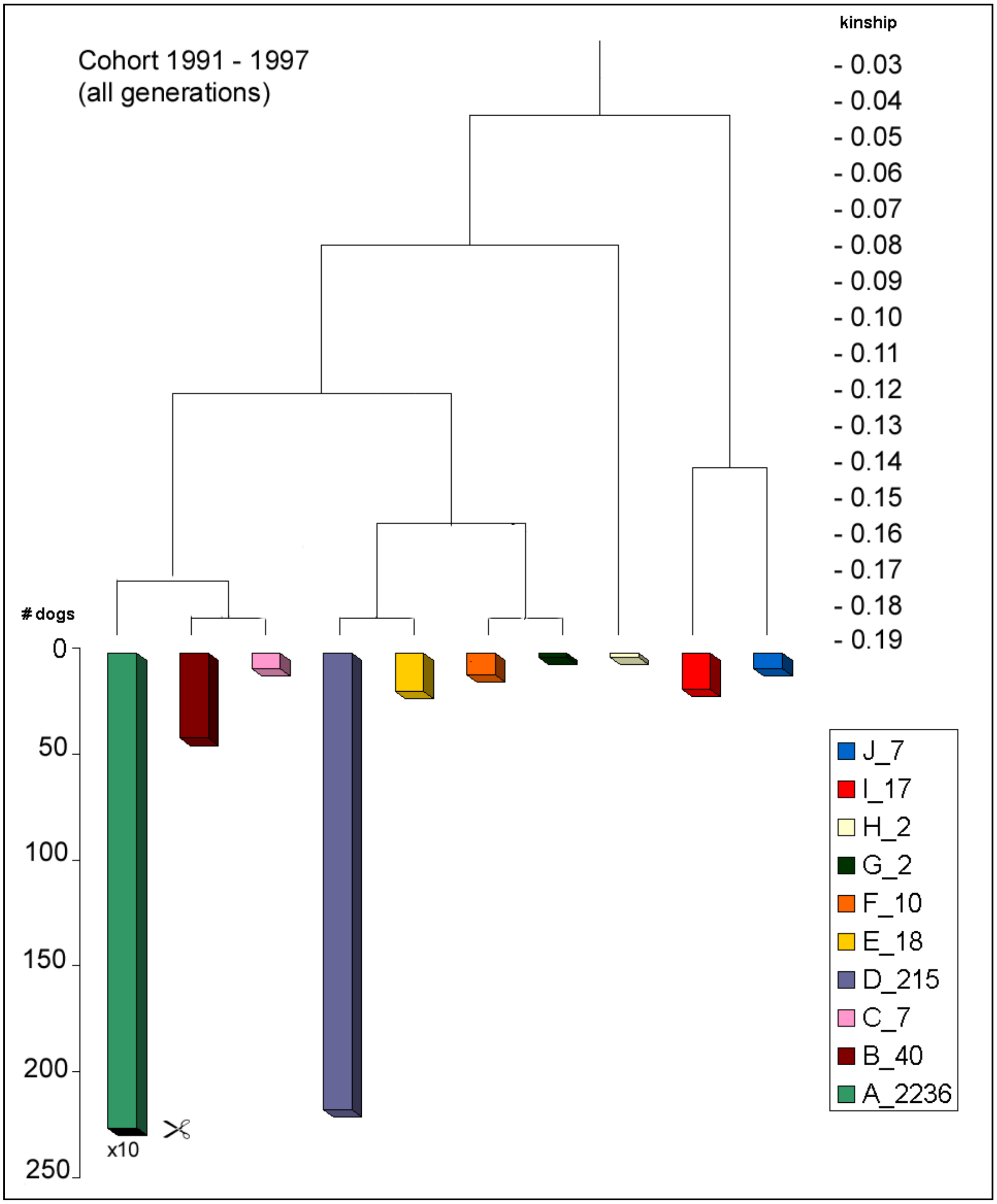

Figure 1. Dendrogram showing the result of clustering based on kinship coefficient calculated up to the founders (**all generations**). A legend is added in order to compare this dendrogram in figure 2. The legend has a code corresponding with each cluster. Except for cluster A_2236, the number of individuals can be estimated by the length of the cluster. Cluster A_2236 represented 2236 animals.

The first analysis used all generations. Family groups were defined as having an average kinship (relation) greater than 0.19, which resulted in ten family groups and is shown in a dendrogram (Figure 1). The bars in the lower part of the dendrogram correspond with the family groups. The vertical branch lengths in the upper part correspond with the relationship (kinship) between the family groups. The kinship-scale in the upper part of the dendrogram is linear and range from 0 till 0.19, but is only shown till 0.03. The value 0.19 was chosen because this value would result in enough family groups to be able to compare this cluster analysis with the second cluster analysis on the same cohort.

The second cluster analysis on the same cohort was based on kinship coefficient calculated for the first seven generations. The analysis produced five clusters (family-groups) when (family) groups of animals having an average relation greater than 0.0625 were defined to be a cluster (figure 2). The kinship-scale in the upper part of figure 2 range from 0 till 0.0625. The value of 0.0625 is chosen to allow comparison with the research of Ubbink et al. (1998ab).

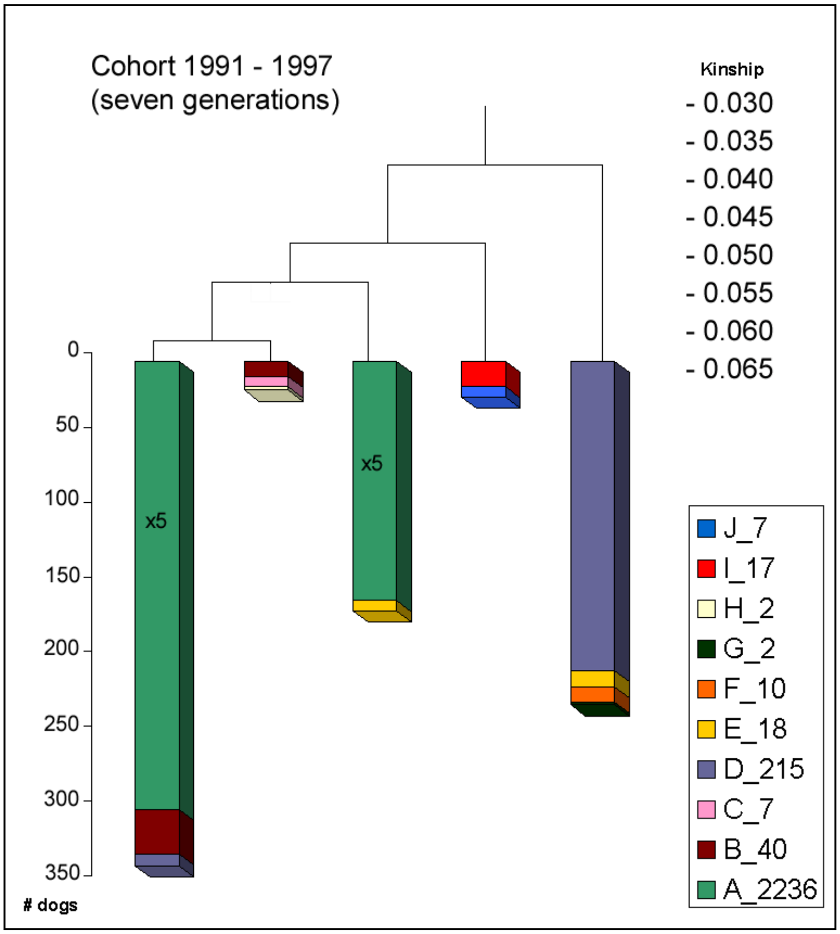

Figure 2. Dendrogram of the same cohort as depicted in figure 3, except, calculation of kinship coefficients was based on **seven generations**. The legend represents the family groups as demonstrated in figure 3. Except for the first and the third cluster, the length of the cluster corresponds with the cluster size. The length of the 'green' A_2236 fraction corresponds with five times the actual size.

The dendrogram of the cluster analysis based on seven generations (figure 2) differs from the dendrogram of the cluster analysis based on calculation up to the founder animals (18 generations: figure 1). Apparently the kinship coefficients within and among the family groups of the second analysis is much lower than in the first analysis. Hence it appears that cluster analysis based on kinship coefficients based on calculations up to the founder animals, increase the average kinship tremendously only for some (parts of the) family groups, compared with cluster analysis based on seven generations. This is the case for group D_215, E_18, F_10, G_2 (increase from 0.035 to 0.12). Other clusters or family groups increase just moderately (group H_2) or even not at all (group I_17, J_7).

3.2 Cluster analysis combined with country of birth

A cohort was assembled including all reproducible (fertile and living) individuals. If the reproductive state of an individual was unknown, the individuals were included unless they were older than 11 years for females and 16 years for males. Cluster analysis protocol was performed on this cohort (all generations). The level of kinship between clusters in the dendrogram was 0.18

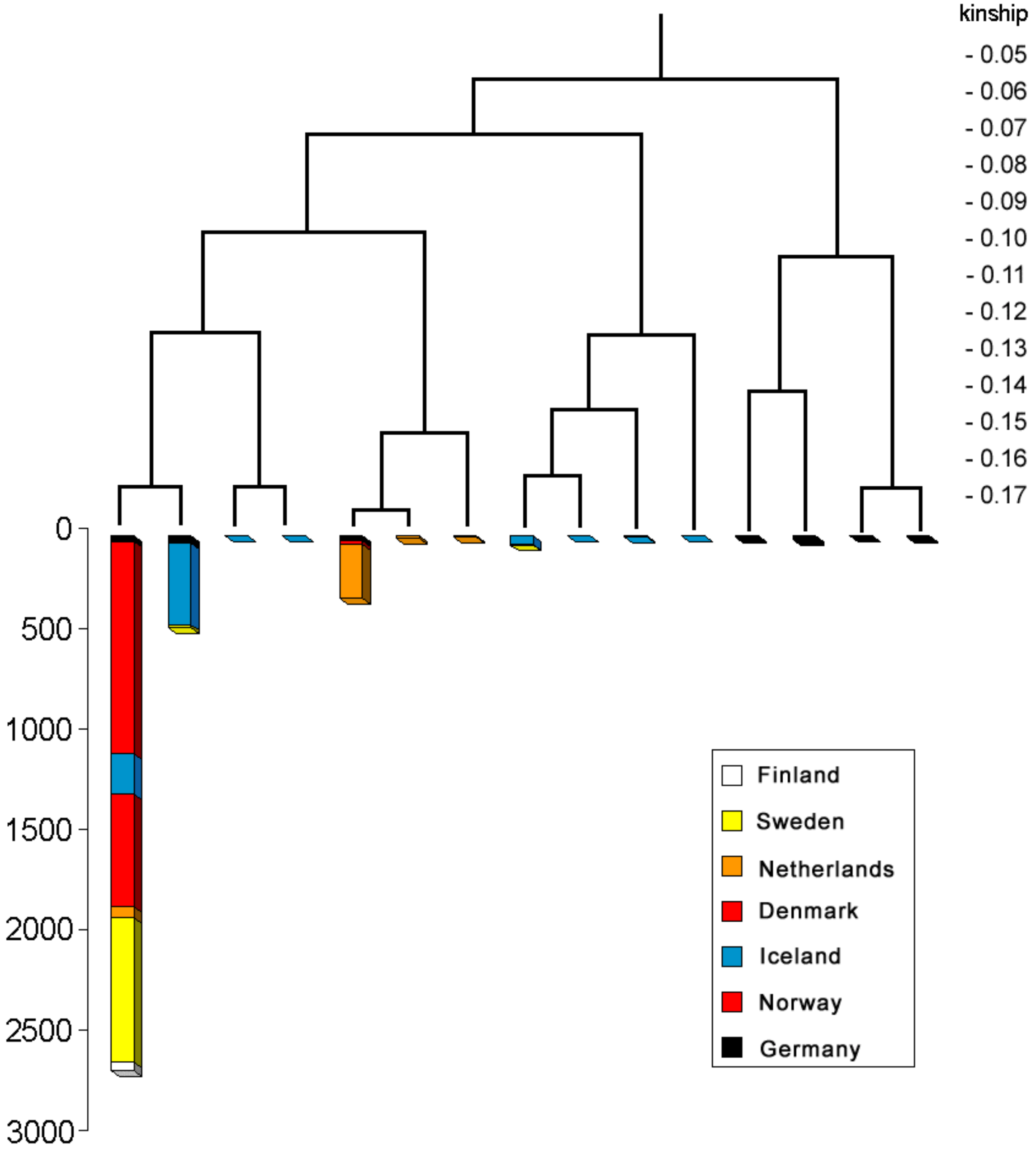

Figure 3. Dendrogram of all reproductive Iceland Dogs showing the result of clustering based on kinship coefficient calculated up to the founders (all generations included). The index shows the country of birth.

The cohort consists of 3572 animals, which were preclustered in 770 unique parent combinations. In this dendrogram the country of birth is added to each individual of each cluster. The first and largest cluster contains the entire Norwegian Iceland Dog population and almost every Iceland Dog that was born in Sweden, the Danish population as well as the Finnish population. The Icelandic Iceland Dog population is fragmented. A part is found in the largest cluster. A second part has average kinship of 0.16 with the large 'N-S-DK-SF family group'. A third part is branching off at average kinship of 0.12 and represents 3 dogs in two clusters. The last part of the Icelandic population has a kinship of 0.07 with the other parts. Almost the entire Dutch Iceland Dog population is clustered together if clusters were defined as groups having an average kinship of 0.15. Apparently the Dutch population is rather unrelated to the other countries. The German Iceland Dog population is spread over the entire dendrogram.

4. DISCUSSION

The research shows that the Icelandic sheepdog is relatively insensitive for inbreeding depression on the number of puppies that are born and survive the first eight weeks. The decrease is probably even lower in the Icelandic Sheepdog than in the Icelandic goat. The latter had a decrease of 0.8 in total number of kids born per 10% increase in the inbreeding coefficient (Aðalsteinsson et al., 1994). In the Icelandic Sheepdog, a 10% increase in the inbreeding coefficient resulted in a decrease of 0.22 number of puppies born (and survived the first eight weeks).

Figure 3 elucidates the subpopulations of the Iceland Dog. One large cluster (family group) contains almost every dog of Scandinavia. The reason behind this is not solely the founder-effect. Many imports from Iceland have been carried out to obtain more diversity within each country. However, each time, the import was from “the best family within Iceland”, since the trouble going trough the importation process is rather costly. Without knowing Scandinavian mainland-countries imported highly related dogs over and over again. On pedigree the close relationship did not show. This resulted in one large highly related family group.

First cluster analysis experiment was conducted in order to compare cluster analysis based on seven generations and cluster analysis based on pedigrees up to the founders (18 generations). Ubbink et al. (1998a) have shown that the inclusion of five, six or seven generations yielded virtually identical and reproducible results. Ideally, the most accurate description of the actual kinship coefficient should be based on pedigrees up to the original unrelated ancestors (founders, Ubbink et al., 1998a). Cluster analysis of the Iceland Dog (cohort 1991-1997) including seven generations was not identical cluster analysis of the same cohort based upon kinship coefficients up to the founders. An ‘ancestor accumulating effect’ probably causes this difference. If a specific individual is inbred on its grandfather (half sib-mating) the inbreeding coefficient is the same as if its four great-grandfathers were the same.

A double number of the same ancestors also double the inbreeding coefficient, while going back one generation bisects inbreeding coefficient. The pedigree of every individual in the cohort will end up with specific overrepresented ancestors, which might appear on different generation levels. Consequently, inbreeding coefficient and kinship coefficients that are calculated up to these ancestors could be much higher than calculations that are based on a limited number of generations (the ‘ancestor accumulation effect’).

Meuwissen (1997) priveds an optimum method for selecting animals while maintaining diversity, but that method is difficult to use in populations with no central management. Cluster analysis of kinship coefficient is a powerfully tool in small unmanaged populations. Like mean kinship, cluster analysis can be used to select breeding animals. Mean kinship already proved to be an accurate method to accomplish increase and maintenance of genetic variation by putting animals in breeding having low mean kinship (Ballou and Lacy, 1995). However, mean kinship does not elucidate the relationship between individuals with low mean kinship. Individuals with low mean kinship are often full sibs (Ballou and Lacy, 1995) or maybe closely related in another way. This corresponds with the dendrograms of cluster analysis of cohort of 1991-1997 (figure not shown) implemented with the mean kinship of the individuals in the cohort. The dendrogram pointed out that animals having low mean kinship clustered together. Other individuals with low mean kinship, however, can be found in different family groups (clusters).

In conclusion, cluster analysis of kinship coefficient reveals the population structure and provides better insight in the breeding potential in terms of genetic variation. The dendrogram of figure 3 shows that the genetically important animals can be found in Iceland, Holland and Germany. Cluster analysis is suitable especially for analysing small non-supervised closed pedigreed populations. In larger highly related populations (up to 1500 individuals), genetic variation can be maximised by cluster analysis of the preclustered parent combinations followed by the calculation of mean kinship only for the individuals of underrepresented family groups (small family groups, that are most distant from large ones).

5. REFERENCES

Aðalsteinsson, S., Ó.R. Dyrmundsson, S. Bjarnadóttir, E. EyÞórsdóttir (1994). Inbreeding in Icelandic Goats. Icelandic Agricultural Science 8, pp. 99-105.

Ballou, J.D. and R.C. Lacy (1995) Identifying genetically important individuals for management of genetic variation in pedigreed populations. In: Ballou, J.D., Gilpin, M. and Foose T.J. (eds.), Population management for survival and recovery, ISBN 0 231 10177 5, Columbia University Press, New York, pp. 76-111.

Meuwissen, T.H.E. (1997) Maximising the response of selection with a predefined rate of inbreeding. Journal of Animal Science 75: 934-940.

Ubbink, G.J., B.W. Knol, J. Bouw (1992) The relationship between homozygosity and the occurrence of specific diseases in Bouvier Belge des Flandres dogs in The Netherlands. Veterinary Quarterly 14 (4; Dec), pp. 137-140.

Ubbink, G.J., J. van de Broek, H.A.W. Hazewinkel, J. Rothuizen (1998a). Cluster analysis of the genetic heterogeneity and disease distributions in purebred dog populations. Veterinary Record 142 (28; Feb), pp. 209-213.

Ubbink, G.J., J. van de Broek, H.P. Meijer, J. Rothuizen (1998b). Prediction of inherited portosystemic shunts in Irish Wolfhounds on the basis of pedigree analysis. American Journal of Veterinary Research 59 (12; Dec), pp. 1553-1556.

Ubbink, G.J., H.A. Hazewinkel, J. van de Broek, J. Rothuizen (1999). Familial clustering and risk analysis for fragmented coronoid process and elbow joint incongruity in Bernese Mountain Dogs in the Netherlands. American Journal of Veterinary Research 60 (9; Sep), pp. 1082-108.