

## **Identification of forces shaping the commensal Escherichia coli genetic structure by comparing animal and human isolates**

Patricia Escobar-Páramo, Arnaud Le Menac'H, Tony Le Gall, Christine Amarin, Stéphanie Gouriou, Bertrand Picard, David Skurnik, Erick Denamur

► **To cite this version:**

Patricia Escobar-Páramo, Arnaud Le Menac'H, Tony Le Gall, Christine Amarin, Stéphanie Gouriou, et al.. Identification of forces shaping the commensal Escherichia coli genetic structure by comparing animal and human isolates. Environmental Microbiology, Wiley-Blackwell, 2006, 8 (11), pp.1975-1984. <10.1111/j.1462-2920.2006.01077.x>. <hal-00554799>

**HAL Id: hal-00554799**

**<http://hal.univ-brest.fr/hal-00554799>**

Submitted on 11 Jan 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Identification of forces shaping the commensal *Escherichia coli* genetic structure by comparing animal and human isolates**

Patricia Escobar-Páramo,<sup>1</sup> Arnaud Le Menac'h,<sup>2</sup> Tony Le Gall,<sup>1</sup> Christine Amorin,<sup>1</sup> Stéphanie Gouriou,<sup>3</sup> Bertrand Picard,<sup>1</sup> David Skurnik,<sup>4</sup> and Erick Denamur<sup>1\*</sup>

INSERM U722 and Université Paris 7 Denis Diderot, Paris,<sup>1</sup> INSERM U707, Faculté de Médecine Saint Antoine, Paris,<sup>2</sup> Laboratoire de Microbiologie, CHU Brest,<sup>3</sup> Laboratoire de Bactériologie, Groupe Hospitalier Bichat-Claude Bernard, AP-HP, and EA 3964, Université Paris 7 Denis Diderot, Paris,<sup>4</sup> France

Running title: Commensal *Escherichia coli* ecological structure

Key words: *Escherichia coli*, commensal, animal, genetic structure, niche adaptation

\*Corresponding author. Mailing address: INSERM U722, Faculté de Médecine Paris 7 Denis Diderot, Site Xavier Bichat, 16 rue Henri Huchard, 75018 Paris, France. Phone: 33 1 44 85 61 56. Fax: 33 1 44 85 61 49. E-mail: [denamur@bichat.inserm.fr](mailto:denamur@bichat.inserm.fr)

## Summary

To identify forces shaping the *Escherichia coli* intraspecies ecological structure, we have characterized in terms of phylogenetic group (A, B1, D and B2) belonging, presence/absence of extraintestinal virulence genes (*pap*, *sfa*, *hly* and *aer*) and intra-host phylotype diversity a collection of 1,898 commensal isolates originating from 387 animals (birds and mammals) sampled in the 1980s and the 2000s. These data have been compared to 760 human commensal isolates, sampled from 152 healthy subjects in the 2000s, and analysed with the same approach. The prevalence of the *E. coli* phylogenetic groups in birds, non-human mammals and humans is clearly different with a predominance of D/B1, A/B1 and A/B2 strains, respectively. A major force shaping the ecological structure is the environment with a strong effect of domestication and the year of sampling followed by the climate. Host characteristics, as the diet and body mass, also influence the ecological structure. Human microbiota are characterised by a higher prevalence of virulence genes and a lower intra-host diversity than the non-human mammal ones. This work identifies for the first time a group of strains specific to the animals, the B1 phylogenetic group strains exhibiting the *hly* gene. In conclusion, a complex network of factors seems to shape the ecological structure of commensal *E. coli*, with anthropogenic factors playing a major role and perturbing natural niche equilibrium.

## Introduction

Intraspecies polymorphism in bacteria results both from mutation and lateral gene transfer with or without subsequent recombination (Maynard Smith et al., 1993). An important issue is to identify the forces that shape and maintain this polymorphism. The existence of ecological structure, i.e. different strains of a species are adapted to particular ecological niches (Maynard Smith 1991), could explain such a polymorphism. Assuming that a given genetic innovation (mutation or horizontal gene transfer) will not be advantageous in all environments, periodic selection will purge genetic variation within an ecological population, but not in the species as a whole (Cohan 1994).

An interesting candidate to validate these hypotheses is the *Escherichia coli* species. Members of this species are mainly commensal of numerous vertebrates including mammals, birds and reptiles, but can also be pathogenic, leading to intestinal (diarrhea) and extraintestinal (urinary tract infection, meningitis, bacteremia) diseases (Donnenberg 2002). *E. coli* can be considered having a clonal genetic structure with a low level of recombination (Selander et al., 1987; Desjardins et al., 1995). Four main phylogenetic groups, A, B1, D and B2, constitute the bulk of the species (Herzer et al., 1990). It is now well accepted that extraintestinal pathogenic strains belong mainly to phylogenetic groups B2 and, at a lesser extend, D and possess numerous specific virulence genes that are absent in commensal strains (Picard et al., 1999). **These genes code for proteins involved in polysaccharide coatings, adhesins, toxins, invasins, iron capture systems and can disseminate via horizontal gene transfer (Mühldorfer and Hacker, 1994).** Non-human mammal commensal strains can be distinguished from their human counterparts by various phenotypic and genotypic markers (Gouillet and Picard 1986; Johnson et al., 2004; Ram et al., 2004; Ahmed et al., 2005). Within humans, ecological

structure seems to be present, with a major role of the climate (temperate versus tropical) (Duriez et al., 2001; Escobar-Páramo et al., 2004b). Within animals, such structure also seems to exist. Souza et al. (1999) showed on a collection of 202 strains from Australian and American wild mammals and birds that the genetic diversity observed by multilocus enzyme electrophoresis is not randomly distributed but reflects geographic effects and host taxonomic groups. Similarly, several studies based on a smaller number of animal species were able to evidence species specific genetic structures (Hartel et al., 2002; Ram et al., 2004; Ahmed et al., 2005). Recently, Gordon and Cowling (2003) have reported in a large collection of 2,300 non-domesticated vertebrate hosts living in Australia that the relative abundance of *E. coli* phylogenetic groups A, B1, D and B2 in mammals depended on animal group, climate, host diet and body mass.

To get further insights in the factors shaping the ecological structure in *E. coli*, we have characterised in terms of phylogenetic group belonging, presence/absence of extraintestinal virulence genes and intra-host diversity a collection of 1,898 commensal isolates originating from 387 wild and domestic mammals and birds living in temperate (mainland France) and tropical regions. We have also compared these data to 760 human commensal isolates analysed with the same approach originating from 152 healthy subjects living in mainland France, Colombia and Benin.

## **RESULTS**

### *Global analysis of animal isolates*

When one randomly selected isolate per sample from all the 387 animals (Tables 1 and 2) were analysed together, strains from phylogenetic groups A (34.1 %) and B1 (33.1 %) were predominant as compared to phylogenetic group D (20.4 %) and B2 (12.4 %) strains. This repartition does not differ when all the 1,898 isolates are considered (data not shown). To assess global relationships between the characteristics of the host and of its environment (mammal or bird, diet, domestic or wild, tropical or temperate climate, year of sampling) and the characteristics of the strain (phylogenetic group and virulence gene content), a multiple component analysis (MCA) was conducted with the 387 *E. coli* isolates as individuals and the characters as qualitative variables. Projections of the variables on the plane F1/F2 (Figure 1), which accounted for 27.7 % of the total variance, clearly distinguish the bird, carnivorous, wild, temperate, 2000s, D and B2 phylogenetic group and virulence gene (except *aer*) variables on the positive values of the F1 axis from the mammal, omnivorous and herbivorous, domestic, tropical, 1980s, A and B1 phylogenetic group and *aer* variables on the negative values of the F1 axis. The association of the *sfa*, *pap* and, at a lesser extend, *hly* virulence genes and the B2 phylogenetic group that are projected on the positive values of both axes is a well-known phenomenon (Picard et al., 1999). No association was evidenced with the host family variable (data not shown).

The clear delineation between bird and mammal isolates, associated to the major physiologic differences between these two groups of animals, prompt us to analyse them separately.

#### *Isolates from birds*

The distribution between the phylogenetic groups is different between bird and mammal isolates (Khi2,  $p < 0.001$ ) with an increase in D (35.9 %) and B2 (18.8 %) and a decrease in A (15.6 %) and B1 (29.7 %) group isolates in birds as compared to mammals (Figure 2B).

According to the small number of isolates studied and their heterogeneous distribution within the variables, the only significant data within bird isolates was the fact that wild birds exhibit more B2 strains and fewer A strains than domestic ones (Fisher,  $p < 0.01$ ).

### *Isolates from mammals*

*Phylogenetic groups.* The distribution of the mammal isolates within the four *E. coli* phylogenetic groups is given in **Figure 2A**. Multinomial logistic regression shows that year of sampling ( $p < 0.001$ ), wild or domestic status ( $p < 0.01$ ), diet ( $p < 0.05$ ) and body mass ( $p < 0.01$ ) affect this distribution. An increase in B2 (2.6 to 16.0 %) and D (10.3 to 21.4 %) and a decrease in A (53.8 to 28.6 %) group strains were observed between the 1980s and the 2000s (**Figure 3A**). Wild animals possess more B2 (19.7 versus 6.7 %) and D (24.1 versus 13.7 %) strains and fewer A (32.1 versus 40.7 %) and B1 (24.1 versus 38.9 %) strains than domestic animals (Khi-2,  $p < 0.0001$ ) (**Figure 3B**), as observed from the birds. It still holds for B1 and B2 group strains when wild and domestic variants of a same species (*Oryctolagus cuniculus*, *Sus scrofa*) or of very closely related species (*Ovis ammon* and *Ovis aries*) are studied (Khi2,  $p < 0.0001$ ) (Table 3). A fact not evidenced on the MCA is that herbivorous are distinguished from non-herbivorous. Indeed, they have fewer A (32.8 vs 47.4 and 52.2 % for carnivorous and omnivorous, respectively) but more D (19.1 vs 15.8 and 11.6 % for carnivorous and omnivorous, respectively) strains than non-herbivorous (Khi2,  $p < 0.05$ ) (**Figure 3C**). The A/D phylogenetic group ratio decreases with the increase of the body mass. Although evidenced in the MCA analysis (Fig 1), the role of the climate has not been retrieved in the subsequent analyses probably due to the small sizes of populations (**Figure 3D**).

*Virulence gene content.* Thirty six isolates (11.2 %) exhibit at least one of the four virulence genes tested (*pap*, *sfa*, *hly* and *aer*) with 26 (8.1 %) and 10 (3.1 %) strains having one and two of these genes, respectively. Logistic regression shows that the presence of virulence genes is dependent on the phylogeny of the strain ( $p < 0.05$ ) and the body mass ( $p < 0.01$ ). B2 strains possess the more virulence genes and A strains the fewer. As previously reported, *sfa* and *pap* are strictly and mainly found in B2 group strains, respectively, whereas *aer* is widespread among the four phylogenetic groups (Picard et al., 1999). Surprisingly, 15 B1 strains exhibit at least one virulence gene, and among them, 14 exhibit *hly*. Such association is unusual as B1 strains are usually devoid of virulence genes and none of the B1 strains in the ECOR collection possess *hly* (Johnson et al., 2001). These *hly* positive B1 strains do not belong to a specific group of mammal hosts as they originate from cows, deers, roe-deers, Pyrenean isard, rabbit, fox and bat. Strains from the 2000s possess more virulence genes than the 1980s ones (14.6 versus 5.1 %) (Khi2,  $p < 0.01$ ), but this could result from the increase in B2 strains in the 2000s collection. The body mass also correlates with the virulence gene content as the presence of virulence genes increases with the increase in body mass ( $r = 0.15$ ,  $p < 0.01$ ).

*Diversity.* The intra-animal isolate diversity was studied in 143 mammals from which 5 isolates were available per animal. The isolates were assigned to one of the 7 group/sub groups A<sub>0</sub>, A<sub>1</sub>, B1, D<sub>1</sub>, D<sub>2</sub>, B<sub>2</sub><sub>2</sub>, B<sub>2</sub><sub>3</sub> by using all the available combinations of the Clermont *et al.* technique (Clermont et al., 2000). The majority of animals (37.7 %) exhibit two types of isolates, 28.0 and 26.6 % exhibit three and one types of isolates, respectively, and 7.7 % exhibit four types of isolates (mean of 2.2 isolate types per animal). Multinomial logistic regression shows a role of the year of sampling ( $p < 0.001$ ) and the wild or domestic status ( $p < 0.01$ ) on the intra-animal isolate diversity: (i) animals sampled in the 1980s have a higher *E. coli* microbiota diversity than those sampled in the 2000s (mean of 2.3 isolate



types/individual vs 1.8) and (ii) wild animals have a higher diversity than domestic animals (mean of 2.3 isolate types/individual vs 1.9). This effect of domestication is also retrieved when phylogenetically very close hosts corresponding to wild and domesticated variants are studied (Wilcoxon test,  $p < 0.05$  for *Oryctolagus* and *Ovis*, non significant for *Sus*) (Table 3), strengthening again the impact of the domestication process over the phylogenetic grouping of the hosts in shaping the *E. coli* genetic structure. Furthermore, univariate analysis displays that the intra-animal *E. coli* microbiota diversity is also affected by the diet as carnivorous (mean 2.3 isolate types/individual) and herbivorous (mean 2.2 isolate types/individual) have higher diversity than omnivorous (mean 1.7 isolate types/individual) (Wilcoxon test,  $p < 0.03$ ). No difference was observed in microbiota diversity between temperate and tropical animals (2.16 versus 2.20,  $p = 0.44$ ).

#### *Comparison between human and non-human mammal isolates*

*Phylogenetic groups.* When one randomly selected isolate from the 152 human subjects were analysed and compared to the unique isolates from the 323 non human mammals, a significant difference in the repartition of the four major phylogenetic groups between human and non-human mammal isolates was observed (Figure 2A) (Khi2  $p < 0.0003$ ) with a higher proportion of B2 (23.7 %) and A (43.4 %) strains and a lower proportion of B1 (21.7 %) and D (11.2 %) strains in human isolates. As for the animal isolates, no difference is observed in the repartition of the phylogenetic groups whether one or five isolates per human individual are taken into account (data not shown). The repartition within the four phylogenetic groups is different between human isolates from tropical and temperate regions with a higher proportion of B2 strains in temperate area as previously reported (Escobar-Páramo et al., 2004b). When human isolates are split in temperate and tropical origins, the significant

difference between human and animal isolates is retrieved only with isolates from human living in temperate area (Khi2,  $p < 0.0001$ ).

*Virulence gene content.* Fifty four human isolates (42.2 %) exhibit at least one virulence gene with 34 (26.6 %), 16 (12.5 %), three (2.3 %) and one (0.8 %) strains having one, two, three and four genes, respectively. Thus, human isolates exhibit a clearly higher proportion of virulence genes than animal isolates (Khi2  $p < 0.0001$ ). Furthermore, the association of *pap* and *hly* genes in a unique strain, which is indicative of the presence of a pathogenicity island (Bingen-Bidois et al., 2002), is retrieved only in human isolates. The repartition of the virulence genes within the phylogenetic groups is identical to the animal one for *pap*, *sfa* and *aer* genes but in sharp contrast for *hly* gene as no *hly* are observed in the B1 phylogenetic group strains. This higher proportion of virulence genes in human isolates is not only the consequence of the higher proportion of B2 strains in humans as the same trend is observed when the analysis is restricted on the B2 isolates (data not shown).

*Diversity.* A significant difference in term of intra-host diversity between human and animal isolates was observed in temperate area (1.68 vs 2.16 isolate types; Wilcoxon test,  $p < 0.001$ ), but not in tropical area (2.27 vs 2.21,  $p = 0.8$ ). This absence of difference could be explained by the fact that, as previously observed (Escobar-Páramo et al., 2004b), the intra-host diversity in human isolates is higher in tropical area (2.27) than in temperate regions (1.68) (Wilcoxon test,  $p < 0.0001$ ).

## Discussion

*E. coli* is a microorganism worldwide distributed, found in most vertebrates present in very different habitats ranging from deserts to polar regions, constantly passed from species to species through water and food, and resistant to severe environmental conditions. All these characteristics may suggest that *E. coli* has unrestricted dispersal capabilities and has been able to colonize all different environments. Consequently, one could expect to find the different phylogenetic groups that constitute the species equally distributed around the world, unless selecting forces are acting differentially upon these groups. This work clearly shows that a strong ecological structure of population exists in commensal *E. coli* strains in terms of repartition in the four main phylogenetic groups, presence of extraintestinal virulence genes and intra-host diversity. The different categories of animals studied and the comparison with human isolates have allowed us to identify some of the forces shaping this structure.

One main force is the environment in which the host is living. This is demonstrated by the effects of human-animal interactions, with a predominance of anthropogenic modifications on animal microbiota. Domestication deeply affects the animal flora. Farming is a human activity involving numerous processes, which include breed selection, diet modification and exposition to a high antibiotic pressure. Farm animal *E. coli* microbiota is characterised, as compared to wild animals, by a higher proportion of A and B1 strains and a lower proportion of B2 and D strains (Figures 1 and 3B). It has been reported that B2 group strains (Johnson et al., 1991; Skurnik et al., 2005) are less prone to be resistant to antibiotics than non-B2 strains. This link between strain phylogeny and antibiotic resistance could explain that in farm animals subjected to a higher antibiotic pressure than wild animals, A and B1 strains are selected and B2 strains counter selected. We have indeed observed that the farm animal

isolates are clearly more resistant than wild animal isolates (Skurnik et al., 2006). An intriguing factor modifying the *E. coli* genetic structure in our collection is the year of sampling, with an increase in the proportion of B2 and D strains over time. This phenomenon has also been reported in human commensal isolates (Escobar-Páramo et al., 2004b) and attributed to the modification in diet and hygiene over the last 20-year-period. It can be hypothesised that the constant anthropogenic modifications in the environment could have been responsible for this shift. One of the regions where we sampled wild animals is a forest near Paris in which human density, represented by both people living close to the forest and hiking during weekends, has increased of 30 % over the last 20 years (<http://www.recensement.insee.fr>). Our work shows that a polymorphism also exists in the intra-host *E. coli* microbiota diversity, influenced by the same factors that shape the repartition of the strains in the different phylogenetic groups: a decrease in diversity is observed over time and with domestication. Such a decrease in *E. coli* microbiota diversity has already been reported between wild and penned deers (Hartel et al., 2003). Furthermore, in temperate area, human microbiota is less diverse than the animal one. According to the niche-width variation hypothesis (Van Valen 1965) and the environmental theory of genetic diversity (Nevo 2001), which describe an increase of intra-species diversity with the increase of habitat variability, this trend could be explained by a decrease in intraluminal gut heterogeneity over time and from wild animals to humans. Due to the lack of hygiene, wild animals are constantly in contact with numerous biotic and abiotic substances in nature as well as with numerous *E. coli* competitors (other bacteria). In addition, they are also in contact with a high diversity of *E. coli* sources, e.g., faeces of other animals. Over time increase of anthropogenic modifications of the environment where wild animals are living, responsible of a decrease in biodiversity, as well as domestication may reduce this gut heterogeneity. Also, hygiene rules in humans lead to a more homogeneous gut environment.

Of note, the observation of higher intra-host diversity in human *E. coli* isolates originating from tropical area than from temperate regions also fits in the environmental theory (Nevo 1988).

Conversely, an influence of animal microbiota on the human one can be suspected. Transmission of bacteria between domestic animals (pigs) and farmers has been reported (Armand-Lefevre et al., 2005). When comparing animal and human *E. coli* microbiota phylogenetic group repartition (Figure 2A), a general trend is emerging with a predominance of A and B1 strains in animals and of A and B2 strains in humans. However, the comparison of our data with a previous work on human microbiota (Escobar-Páramo et al., 2004b) shows that the prevalence of the phylogenetic groups in commensal *E. coli* from the specific population of French pig farmers, from which animal-human transmission of bacteria has been reported (Armand-Lefevre et al., 2005), is very close to the repartition that we observed in the animals. Beside human-animal interactions, another role of the environment is illustrated by the fact that B2 group strains are expected to be less frequently isolated from hosts living in the tropics than from hosts living in temperate climates (Figure 1). Interestingly, this trend has also been observed in Australian vertebrate (Gordon and Cowling, 2003) and human (Escobar-Páramo et al., 2004b) commensal isolates.

Another main force shaping the genetic structure of commensal *E. coli* strains comes from the host characteristics. Bird, non-human mammal and human isolates are clearly distinguished (Figures 1 and 2). This work shows the role of diet and intestine physiology, represented by the host phylogeny and body mass. For the repartition in the four main phylogenetic groups, this has been previously reported in a study based on Australian vertebrates (Gordon and Cowling, 2005). However, the effects do not appear to be similar between both studies. For example, Australian mammals omnivorous have the higher proportion (more than 50 %) of

B2 group strains compared to carnivorous and herbivorous mammals. At the opposite, we found that omnivorous mammals have the lower B2 strain proportion (less than 6 %) (Figure 3C). This indicates that the forces shaping the *E. coli* genetic structure are complex and multifactorial and that each combination of those forces gives a specific structure. An interesting new data emerging from our work is the effect of host characteristics on the virulence gene content of the strains. Genes coding for adhesins, toxins or proteins involved in iron capture have been called “virulence genes” as they are more frequently found in pathogenic strains than in commensal strains (Picard et al., 1999; Johnson et al., 2001) and as mutants lacking these genes are less virulent in animal models (Schubert et al., 2002). However, several studies have pointed out the fact that these genes, and particularly genes coding for adhesins, are associated to the persistence of the strains carrying them in commensal intestinal microbiota (Wold et al., 1992). This is an argument for the idea that these genes were at the origin selected and then maintained for giving a better fitness to strains within the intestine, the normal habitat of *E. coli*, and that their role in virulence is “coincidental” (Levin 1996). We observed an increase in the number of virulence genes with the increase in body mass in non-human mammals as well as enrichment in these genes in humans as compared to animals. As an increase of the body mass is correlated to an increase of the large intestine and as the retention times increase significantly with the body mass (Gordon and Cowling, 2003), it could be hypothesized that the increase of the virulence genes observed allows a better adaptation of the strains to a more complex ecological niche. Likewise, human commensal strains are subjected to numerous selective pressures due to exposure to various chemicals and antibiotics, which could favour the emergence of strains carrying these genes. We also evidenced a role of diet on the mammal microbiota diversity as omnivorous have a lower diversity than herbivorous and carnivorous. However, these data are in disagreement with the diversity theory discussed above (Van Valen 1965; Nevo 2001) as

carnivorous and herbivorous have the basically more unvaried diet and omnivorous the more heterogeneous. This discrepancy from diversity theory might indicate that the intra-host diversity is not a sole effect from the diet as such.

Up to now, the genetic structure of *E. coli* commensal strains was characterised by differences in allele frequencies between ecotypes (Gordon 2001), without identification of specific clones. The opposition between A/B1 and D/B2 phylogenetic groups on the MCA (Figure 1), previously reported for human commensal isolates (Escobar-Páramo et al., 2004b), suggests that A/B1 strains occupy similar commensal niches, which are distinct from the commensal niches occupied by D/B2 strains. Interestingly, this delineation corresponds to the phylogenetic history of the species as B2 and D strains were the first to emerge during the evolution followed by the A and B1 strains which are sister groups (Escobar-Páramo et al., 2004a). For the first time, we were able to identify a group of strains specific to the animals, the B1 phylogenetic group strains exhibiting the *hly* gene. These strains were isolated from wild and domestic mammals, essentially herbivorous, in several locations (Brittany, Pyrenees, Venezuela, Ethiopia). The preliminary analyses of these strains indicate a relative heterogeneity between them as sequencing of the *hly* gene identified two groups of sequences as in (Escobar-Páramo et al., 2004a) and serotyping evidenced at least three distinct serotypes (Clermont and Denamur, personal data).

In conclusion, by comparing commensal *E. coli* isolates from animals and humans of various origins with the same approach, we were able to distinguish very specific bacterial population characteristics, related to the host and the environment. A complex network of factors seems to shape these characteristics. Of note, anthropogenic factors play a major role, perturbing natural niche equilibrium.

## Experimental procedures

### *Animal bacterial isolates and host characteristics*

For the majority of animals, freshly deposited faeces were collected *ad libitum* when found on the ground and preserved at + 4° C until bacterial isolation. Cloacal sampling has also been performed in some cases of bird faeces. The stools were then cultured on Drigalski agar plates and one to 10 lactose-positive colonies were randomly picked from each sample. *E. coli* were identified by an API 20E gallery (bioMérieux). A total of 1,898 *E. coli* were isolated from 387 distinct animals. These animals encompass 323 mammals (35 species) and 64 birds (10 species). Among them, 135 were wild animals whereas 252 were domestic (extensively reared farm animals), with 122 and 265 sampled in the 1980s and 2000s, respectively. The animals were living principally in mainland France (temperate area) in three principal regions: Brittany, the Pyrenees and the Ile de France (93 from the 1980s and 248 from the 2000s) but some (29 from the 1980s and 13 from the 2000) were living in tropical areas (French Caribbean islands, Venezuela, Cameroon, Ethiopia, Gabon). The complete list of animals and isolates studied as well as their main characteristics are given in Tables 1 and 2. The body mass as well as the diet of the host were also considered. For the body mass, the log<sub>10</sub> of the mean value from the literature was used. For the diet, mammals were assigned to the carnivorous, herbivorous or omnivorous categories (Table 1).

### *Human bacterial isolates*

A collection of 760 isolates harvested from the faeces of 152 healthy human subjects (5 randomly sampled isolates per individual) was studied. This collection is mainly part of a



previously published work (Escobar-Páramo et al., 2004b) and is composed of isolates from 55 subjects living in mainland France with a temperate climate (28 from Brest in Brittany and 27 from Paris) and 97 subjects living in tropical area (69 from Cotonou, Benin and 28 from Bogotá, Colombia).

### *Bacterial genotyping*

The Clermont et al. method (Clermont et al., 2000) was first used for the assignment of *E. coli* isolates to one of the 4 major phylogenetic groups A, B1, D, or B2. A slightly modified version of this method was developed to be able to handle numerous strains at lower cost than PCR. Dot blots of bacterial lysates were hybridised with DIG-labelled probes generated by PCR amplification of the three DNA markers (the genes *chuA* and *yjaA*, and the anonymous DNA fragment Tspe4.C2) described by Clermont et al., (2000). The perfect concordance of the two approaches (PCR and dot blot hybridisation) was verified on 300 strains (data not shown). To increase the discriminative power of our analyses when several isolates per sample were considered, all the combinations of genetic markers described in this method were also used as follows: subgroup A<sub>0</sub> (group A), *chuA*<sup>-</sup>, *yjaA*<sup>-</sup>, Tspe4.C2<sup>-</sup>; subgroup A<sub>1</sub> (group A), *chuA*<sup>-</sup>, *yjaA*<sup>+</sup> Tspe4.C2<sup>-</sup>; group B1, *chuA*<sup>-</sup>, *yjaA*<sup>-</sup>, Tspe4.C2<sup>+</sup>; subgroup B2<sub>2</sub> (group B2), *chuA*<sup>+</sup>, *yjaA*<sup>+</sup>, Tspe4.C2<sup>-</sup>; subgroup B2<sub>3</sub> (groupB2), *chuA*<sup>+</sup>, *yjaA*<sup>+</sup>; Tspe4.C2<sup>+</sup>; subgroup D<sub>1</sub> (group D), *chuA*<sup>+</sup>, *yjaA*<sup>-</sup>, Tspe4.C2<sup>-</sup>; and subgroup D<sub>2</sub> (group D), *chuA*<sup>+</sup>, *yjaA*<sup>-</sup>, Tspe4.C2<sup>+</sup>. Thus, seven group/subgroups (A<sub>0</sub>, A<sub>1</sub>, B1, B2<sub>2</sub>, B2<sub>3</sub>, D<sub>1</sub> and D<sub>2</sub>) were defined. Extraintestinal virulence genes (*pap*, *sfa/foc*, *hly*, *aer*) were detected from DNA by PCR or hybridisation as previously described (Picard et al., 1999). These genes code for two adhesins (pyelonephritis associated pili system, S fimbrial adhesin), one toxin ( $\alpha$ -hemolysin) and one

iron capture system, respectively, and are good representatives of the intrinsic extraintestinal virulence of the strains (Picard et al., 1999).

### *Statistical analyses*

Analyses were performed on three data sets: (i) one randomly selected isolate per sample, (ii) all the isolates available per sample, (iii) 5 randomly selected isolates per sample. In a first step, a MCA (Tenenhaus and Young, 1985) was carried out from the animal isolate data to evaluate associations between parameters (phylogenetic groups, virulence genes, year of sampling, climate, diet, animal classes). **MCA uses a covariance matrix based on  $\chi^2$  distances. The computation determines a plane defined by two principal axes of the analysis; the first axis F1, accounts for most of the variance, and the second axis F2, orthogonal to F1, accounts for the largest part of the variance that is not accounted by F1. The plane F1/F2 allowed the positioning of the variables according to their coordinates on each of the factors (Figure 1). MCA was conducted from a 2 two-way table which had 387 rows, one for each *E. coli* strain and 19 columns corresponding to the 19 variables (host, environment and bacterial strain characteristics). For each column, each strain was coded as a binary code: present=1, absent=0. To confirm the significance of associations observed with the multiple component analysis and analyse the role of body mass, logistic regression analysis, Khi2 and Fisher tests were performed. To compare the diversity of *E. coli* isolates in mammals and humans, non-parametric Wilcoxon test was used. The analyses were carried out using SAS/STAT® Software (v. 9.1).**

## **Acknowledgments**

We are grateful to all the people (M. Denamur, J. L. Crampe from the “Office National de la Chasse et de la Faune Sauvage”, G. Lecointre from the “Museum National d’Histoire Naturelle”, P. Rouquet from the “Centre International de Recherches Médicales de Franceville” and B. Cadiou) who spent a lot of their time collecting the faeces, to Lorraine Domelovo for technical help, to Olivier Clermont for numerous discussions on the data and to an anonymous reviewer for her/his numerous constructive remarks. We thank E. Gadreau for editing the manuscript. This work was partially funded by the “Fondation pour la Recherche Médicale”.

## References

- Ahmed, W., R. Neller, and M. Katouli. 2005. Host species-specific metabolic fingerprint database for Enterococci and *Escherichia coli* and its application to identify sources of fecal contamination in surface waters. *Appl. Environ. Microb.* 71:4461-4468.
- Armand-Lefevre, L., R. Ruimy, and A. Andremont. 2005. Clonal comparison of *Staphylococcus aureus* isolates from healthy pig farmers, human controls, and pigs. *Emerg. Infect. Dis.* 11:711-714.
- Bingen-Bidois, M., O. Clermont, S. Bonacorsi, M. Terki, N. Brahimi, C. Loukil, D. Barraud, and E. Bingen. 2002. Phylogenetic analysis and prevalence of urosepsis strains of *Escherichia coli* bearing pathogenicity island-like domains. *Infect. Immun.* 70:3216-3226.
- Clermont, O., S. Bonacorsi, and E. Bingen. 2000. Rapid and simple determination of the *Escherichia coli* phylogenic group. *Appl. Environ. Microb.* 66:4555-4558.
- Cohan, F. M. 1994. Genetic exchange and evolutionary divergence in prokaryotes. *Trends Microbiol.* 9:175-180.
- Desjardins, P., B. Picard, B. Kaltenbock, J. Elion, and E. Denamur. 1995. Sex in *Escherichia coli* does not disrupt the clonal structure of the population: evidence from random amplified polymorphic DNA and restriction-fragment-length polymorphism. *J. Mol. Evol.* 41:440-448.
- Donnenberg, M. 2002. *Escherichia coli* virulence mechanisms of versatile pathogen. Elsevier Science Ed ed, San Diego California.
- Duriez, P., O. Clermont, S. Bonacorsi, E. Bingen, A. Chaventre, J. Elion, B. Picard, and E. Denamur. 2001. Commensal *Escherichia coli* isolates are phylogenetically distributed among geographically distinct human populations. *Microbiology* 147:1671-1676.
- Escobar-Páramo, P., O. Clermont, A. Blanc-Potard, H. Bui, C. Le Bouguenec, and E. Denamur. 2004a. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. *Mol. Biol. Evol.* 21:1085-1094.
- Escobar-Páramo P, K. Grenet, A. Le Menach, L. Rode, E. Salgado, C. Amorin, S. Gouriou, B. Picard, M. Cherif Rahimy, A. Andremont, E. Denamur, and R. Ruimy. 2004b. Large-scale population structure of human commensal *Escherichia coli* isolates. *Appl. Environ. Microbiol.* 70: 5698-5700.
- Gordon, D. M. 2001. Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. *Microbiology* 147:1079-1085.
- Gordon, D. M., and A. Cowling. 2003. The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology* 149:3575-3586.
- Goulet, P., and B. Picard. 1986. Comparative esterase electrophoretic polymorphism of *Escherichia coli* isolates obtained from animal and human sources. *J. Gen. Microbiol.* 132:1843-1851.

- Hartel, P. G., J. D. Summer, J. L. Hill, J. V. Collins, J. A. Entry, and W. I. Segars. 2002. Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. *J. Environ. Qual.* 31:1273-1278.
- Hartel, P. G., J. D. Summer, and W. I. Segars. 2003. Deer diet affects ribotype diversity of *Escherichia coli* for bacterial source tracking. *Water Res.* 37:3263-3268.
- Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam. 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J. Bacteriol.* 172:6175-6181.
- Johnson, J. R., P. Goulet, B. Picard, S. L. Moseley, P. L. Roberts, and W. E. Stamm. 1991. Association of carboxylesterase B electrophoretic pattern with presence and expression of urovirulence factor determinants and antimicrobial resistance among strains of *Escherichia coli* that cause urosepsis. *Infect. Immun.* 59: 2311-2315.
- Johnson, J., P. Delavari, M. Kuskowski, and A. L. Stell. 2001. Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*. *J. Infect. Dis.* 183:78-88.
- Johnson, L. K., M.B. Brown, E. A. Carruthers, J. A. Ferguson, P. E. Dombek, and M. J. Sadowsky. 2004. Sample size, library composition, and genotypic diversity among natural populations of *Escherichia coli* from different animals influence accuracy of determining sources of fecal pollution. *Appl. Environ. Microb.* 70:4478-4485.
- Levin, B. 1996. The evolution and maintenance of virulence in microparasites. *Emerg. Infect. Dis.* 2:93-102.
- Maynard Smith, J. 1991. The population genetics of bacteria. *Proc. R. Soc. B.* 245:37-41.
- Maynard Smith, J., N. H. Smith, M. O'Rourke, and B. G. Spratt. 1993. How clonal are bacteria ? *Proc. Natl. Acad. Sci. USA.* 90:4384-4388.
- Mühdorfer, I., and J. Hacker. 1994. Genetics aspects of *Escherichia coli* virulence. *Microb. Pathog.* 16:171-181.
- Nevo, E. 1988. Genetic diversity in nature: patterns and theory. *Evolutionary Biology.* 23:217-247.
- Nevo, E. 2001. Evolution of genome-phenome diversity under environmental stress. *Proc. Natl. Acad. Sci. USA.* 98:6233-6240.
- Picard, B., J. Garcia, S. Gouriou, P. Duriez, N. Brahimi, E. Bingen, J. Elion, and E. Denamur 1999. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect. Immun.* 67:546-553.
- Ram, J. L., R. P. Ritchie, J. Fang, F. S. Gonzales, and J. P. Selegan. 2004. Sequence-based source tracking of *Escherichia coli* based on genetic diversity of  $\beta$ -glucuronidase. *J. Environ. Qual.* 33:1024-1032.

Schubert, S., B. Picard, S. Gouriou, J. Heesemann, and E. Denamur. 2002. Yersinia High-Pathogenicity Island (HPI) contributes to virulence in *Escherichia coli* causing extraintestinal infections. *Infect. Immun.* 70:5335-5337.

Selander, R., D. Caugant, and T. Whittnam. 1987. Genetic structure and variation in natural populations of *Escherichia coli*, p. 1625-1648. In F. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society of Microbiology, Washington DC.

Skurnik, D., A. Le Menac'h, D. Zurakowski, D. Mazel, P. Courvalin, E. Denamur, A. Andremont, and R. Ruimy. 2005. Integron-associated antibiotic resistance and phylogenetic grouping of *Escherichia coli* isolates from healthy subjects free of recent antibiotic exposure. *Antimicrob Agents Chemother.* 49:3062-3065.

Skurnik, D., R. Ruimy, A. Andremont, C. Amorin, P. Rouquet, B. Picard, and E. Denamur. 2006. Effect of human vicinity on antibiotic resistance and integrons in animal faecal *Escherichia coli*. *J. Antimicrob. Chemother.* In press

Souza V, M. Rocha, A. Valera, and L. E. Eguiarte. 1999. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. *Appl. Environ. Microbiol.* 65:3373-3385.

Tenehaus, M. and F. W. Young. 1985. An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis, and others methods for quantifying categorical multivariate data. *Psychometrika.* 50:91-119.

Van Valen, L. 1965. Morphological variation and width of ecological niche. *American Naturalist.* 100:377-389.

Wold, A. E., D. A. Caugant, G. Lidin-Janson, P. de Man, and C. Svanborg. 1992. Resident colonic *Escherichia coli* strains frequently display uropathogenic characteristics. *J. Infect. Dis.* 165:46-52.

## Figure legends:

Figure 1: Graphical representation of the results of the multiple component analysis carried out with the whole data from 387 *E. coli* isolates sampled from 387 distinct animals. Projection of the variables on the plane F1/F2: class of animals (mammal: Mam, bird: Bird), diet (carnivorous: Cn, herbivorous: Hb, omnivorous: On, for mammals only), status (wild: Wild, domestic: Dom), climate (temperate: Temp, tropical: Trop), year of sampling (1980, 2000), phylogenetic groups (A, B1, D, B2) and virulence genes (*aer*, *hly*, *pap*, *sfa*). In grey: host and its environment characteristics, in black: bacterial **strain** characteristics.

Figure 2: Repartition within the four major phylogenetic groups (A, B1, D, B2) of the *E. coli* (one isolate per individual) isolated from (A) animals and humans and (B) birds and mammals p values are indicated only when the differences are significant.

Figure 3: Repartition within the four major phylogenetic groups (A, B1, D, B2) of the *E. coli* isolated from mammals (one isolate per individual) according to (A) the year of sampling, their (B) habitat and (C) diet and (D) the climate. p values are indicated only when the differences are significant.

**Table 1:** Complete list of the animals and *E. coli* isolates studied.

	Linean denomination	Common name	Family	Body mass <sup>a</sup> (Kg)	Diet <sup>b</sup>	Wild/ Domestic	Climate	Animal (n)	<i>E. coli</i> isolate (n)
Mammal	<i>Aepyceros melampus</i>	Impala	<i>Bovidae</i>	60	Hb	Wild	Tropical	3	29
	<i>Antidorcas marsupialis</i>	Springbok	<i>Bovidae</i>	35	Hb	Wild	Tropical	1	8
	<i>Bos taurus</i>	Domestic Cow	<i>Bovidae</i>	600	Hb	Domestic	Temperate/Tropical	29/4	126
	<i>Bubalus caffer</i>	Buffalo	<i>Bovidae</i>	625	Hb	Wild	Tropical	1	10
	<i>Capreolus capreolus</i>	Roe-Deer	<i>Cervidae</i>	25	Hb	Wild	Temperate	9	90
	<i>Cephalophus sylvicultor</i>	Yellow-backed Duiker	<i>Bovidae</i>	17.5	Hb	Wild	Tropical	2	20
	<i>Cervus elaphus</i>	Common Deer	<i>Bovidae</i>	175	Hb	Wild	Temperate	18	175
	<i>Capra hircus</i>	Goat	<i>Bovidae</i>	45	Hb	Domestic	Temperate	22	109
	<i>Equus caballus</i>	Horse	<i>Equidae</i>	425	Hb	Domestic	Temperate	37	37
	<i>Erinaceus europaeus</i>	Hedgehog	<i>Erinaceidae</i>	0.75	Cn	Wild	Temperate	2	20
	<i>Felis concolor</i>	Puma	<i>Felidae</i>	100	Cn	Wild	Tropical	1	8
	<i>Gorilla gorilla</i>	Gorilla	<i>Pongidae</i>	250	Hb	Wild	Tropical	2	20
	<i>Hydrochaeris hydrochaeris</i>	Capybara	<i>Hydrocherides</i>	50	Hb	Wild	Tropical	1	5
	<i>Lepus capensis</i>	Hare	<i>Leporidae</i>	4.15	Hb	Wild	Temperate	7	70
	<i>Loxodonta africana</i>	Elephant	<i>Elephantidae</i>	6000	Hb	Wild	Tropical	1	10
	<i>Marmota marmota</i>	Marmot	<i>Sciurides</i>	6	Hb	Wild	Temperate	4	40
	<i>Martes foina</i>	Beech marten	<i>Mustelidae</i>	1.8	Cn	Wild	Temperate	3	30
	<i>Martes martes</i>	Pine marten	<i>Mustelidae</i>	1.3	Cn	Wild	Temperate	4	40
	<i>Microchiropteres sp.</i>	Bat	<i>Vespertilionides</i>	0.01	On	Wild	Temperate	4	40
	<i>Microtus arvalis</i>	Common Voles	<i>Cricetidae</i>	0.0125	Hb	Wild	Temperate	1	10
	<i>Myocastor coypus</i>	Nutria	<i>Myocastoridae</i>	8	On	Wild	Tropical	1	10
	<i>Oryctolagus cuniculus</i>	Rabbit	<i>Leporidae</i>	1.775	Hb	Wild	Temperate	7	50
	<i>Oryctolagus cuniculus domestica</i>	Domestic Rabbit	<i>Leporidae</i>	1.775	Hb	Domestic	Temperate	16	35
	<i>Oryx Gazella</i>	Gemsbok	<i>Bovidae</i>	210	Hb	Wild	Tropical	1	6
	<i>Ovis ammon</i>	European Mouflon	<i>Bovidae</i>	37.5	Hb	Wild	Temperate	10	100
	<i>Ovis aries</i>	Sheep	<i>Bovidae</i>	22.5	Hb	Domestic	Temperate/Tropical	41/8	119
	<i>Rattus norvegicus</i>	Rat	<i>Muridae</i>	0.19	On	Wild	Temperate	3	21
	<i>Rupricapra pyrenaica</i>	Pyrenean Isard	<i>Bovidae</i>	27.5	Hb	Wild	Temperate	3	30
	<i>Sciurus vulgaris</i>	Red Squirrel	<i>Sciuridae</i>	0.375	Hb	Wild	Temperate	2	20
	<i>Sus scrofa</i>	Wild Boar	<i>Suidae</i>	125	On	Wild	Temperate	8	80
	<i>Sus scrofa domestica</i>	Pig	<i>Suidae</i>	90	On	Domestic	Temperate/Tropical	51/3	79
	<i>Tapirus terrestris</i>	Tapir	<i>Tapiridae</i>	247.5	Hb	Wild	Tropical	2	7
<i>Taurotragus oryx</i>	Eland	<i>Bovidae</i>	825	Hb	Wild	Tropical	2	15	
<i>Ursus arctos</i>	Brown Bear	<i>Ursidae</i>	825	Hb	Wild	Temperate	3	30	
<i>Vulpes vulpes</i>	Fox	<i>Canidae</i>	4.5	Cn	Wild	Temperate	6	59	
Bird	<i>Anser anser</i>	Geese	<i>Anatidae</i>	4		Domestic	Temperate	2	2
	<i>Gallus gallus</i>	Chickens	<i>Phasianidae</i>	1.5		Domestic	Temperate	33	102
	<i>Larus marinus</i>	Great Black-backed Gull	<i>Laridae</i>	1.1		Wild	Temperate	8	80
	<i>Meleagris gallopavo</i>	Turkey	<i>Meleagrididae</i>	10		Domestic	Temperate	6	6
	<i>Oceanodroma leucorhoa</i>	Leach's Storm-petrel	<i>Hydrobatidae</i>	0.45		Wild	Temperate	1	10
	<i>Phalacrocorax carbo</i>	Cormorant	<i>Phalacrocoracidae</i>	2.25		Wild	Temperate	3	30



<i>Phasianus colchicus</i>	Pheasant	<i>Phasianidae</i>	1.05	Wild	Temperate	7	70
<i>Pyrrhocorax pyrrhocorax</i>	Chough	<i>Corvidae</i>	0.15	Wild	Temperate	1	10
<i>Tetrao urogallus</i>	Capercaillie	<i>Tetraonidae</i>	1.5	Wild	Temperate	2	20
<i>Turdus merula</i>	Blackbird	<i>Turdidae</i>	0.095	Domestic	Temperate	1	10

<sup>a</sup> Mean value from the literature.

<sup>b</sup> For mammals only. Hb: Herbivorous, Cn: Carnivorous, On: Omnivorous.

**Table 2:** Main characteristics of the animals and *E. coli* isolates studied from the two periods of sampling (1980 and 2000).

	1980 <sup>a</sup>			2000 <sup>a</sup>			Total
	Wild	Domestic	Total	Wild	Domestic	Total	
Mammals	22 <sup>b</sup> (166) <sup>c</sup>	95 (227)	117 (393)	90 (887)	116 (278)	206 (1165)	323 (1558)
Birds	0 (0)	5 (20)	5 (20)	23 (230)	36 (90)	59 (320)	64 (340)
Total	22 (166)	100 (247)	122 (413)	113 (1117)	152 (368)	265 (1485)	387 (1898)

<sup>a</sup> Year of sampling.

<sup>b</sup> Number of animals studied.

<sup>c</sup> Number of strains studied.

**Table 3:** Repartition within the four major phylogenetic groups (A, B1, D, B2) and intra-host diversity of the *E. coli* isolated from wild and domesticated variants in a set of mammals.

	Phylogenetic groups (%)				Intra-host diversity <sup>a</sup>		
	Total <sup>b</sup> (n)	A	B1	D	B2	Total <sup>b</sup> (n)	Isolate types per animal <sup>c</sup>
Domestic Rabbits	35	17	66	11	6	20	1.75
Wild Rabbits	50	64	10	16	24	40	3.4
Domestic Pigs	79	62	28	6	4	20	2.0
Wild Boars	80	3	10	49	38	80	2.5
Domestic Sheeps	119	34	54	9	3	40	1.75
Wild Mouflons	100	35	21	20	24	50	2.7

<sup>a</sup>The intra-animal isolate diversity was studied in mammals from which at least 5 isolates were available per animal, using all the combinations of genetic markers described in Clermont et al. 2000.

<sup>b</sup>Number of studied isolates.

<sup>c</sup>Mean.

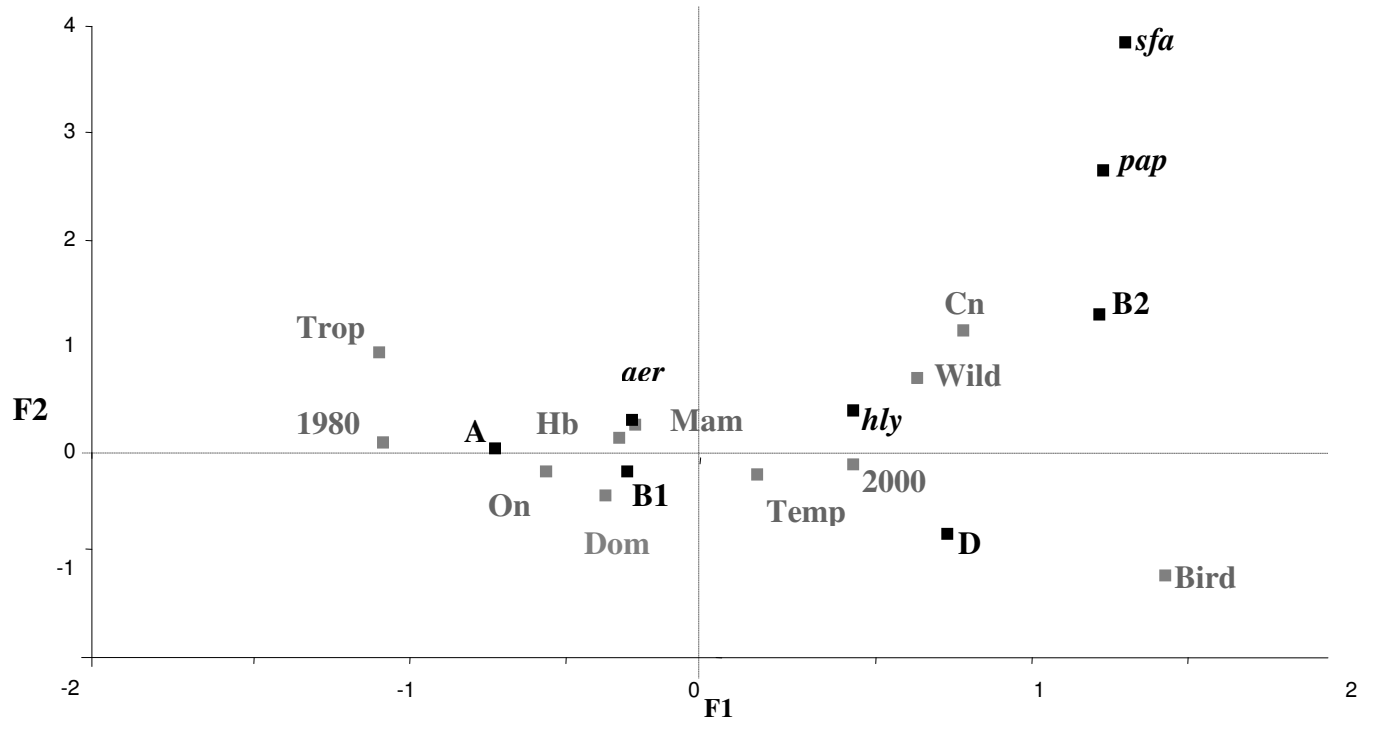


Figure 1

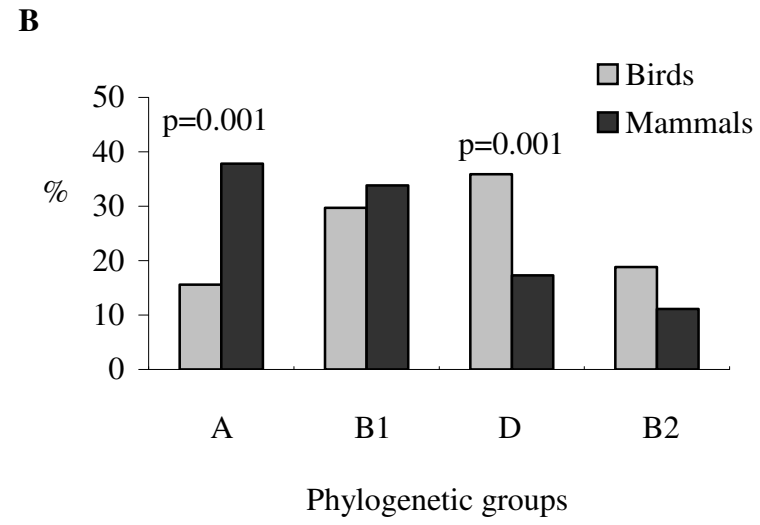
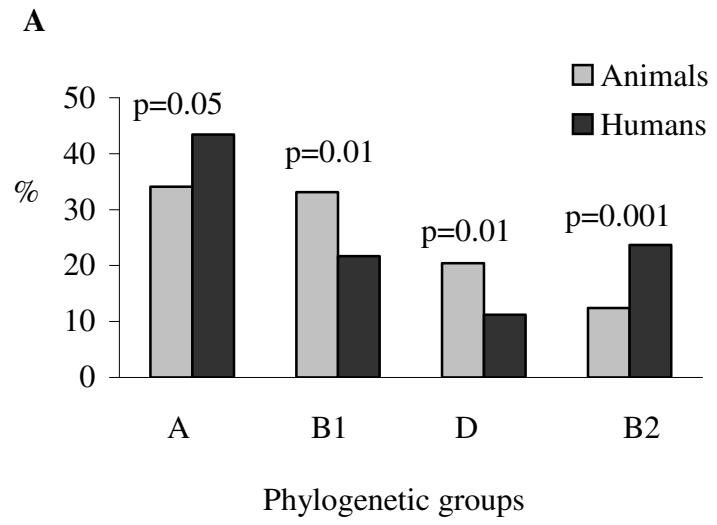


Figure 2

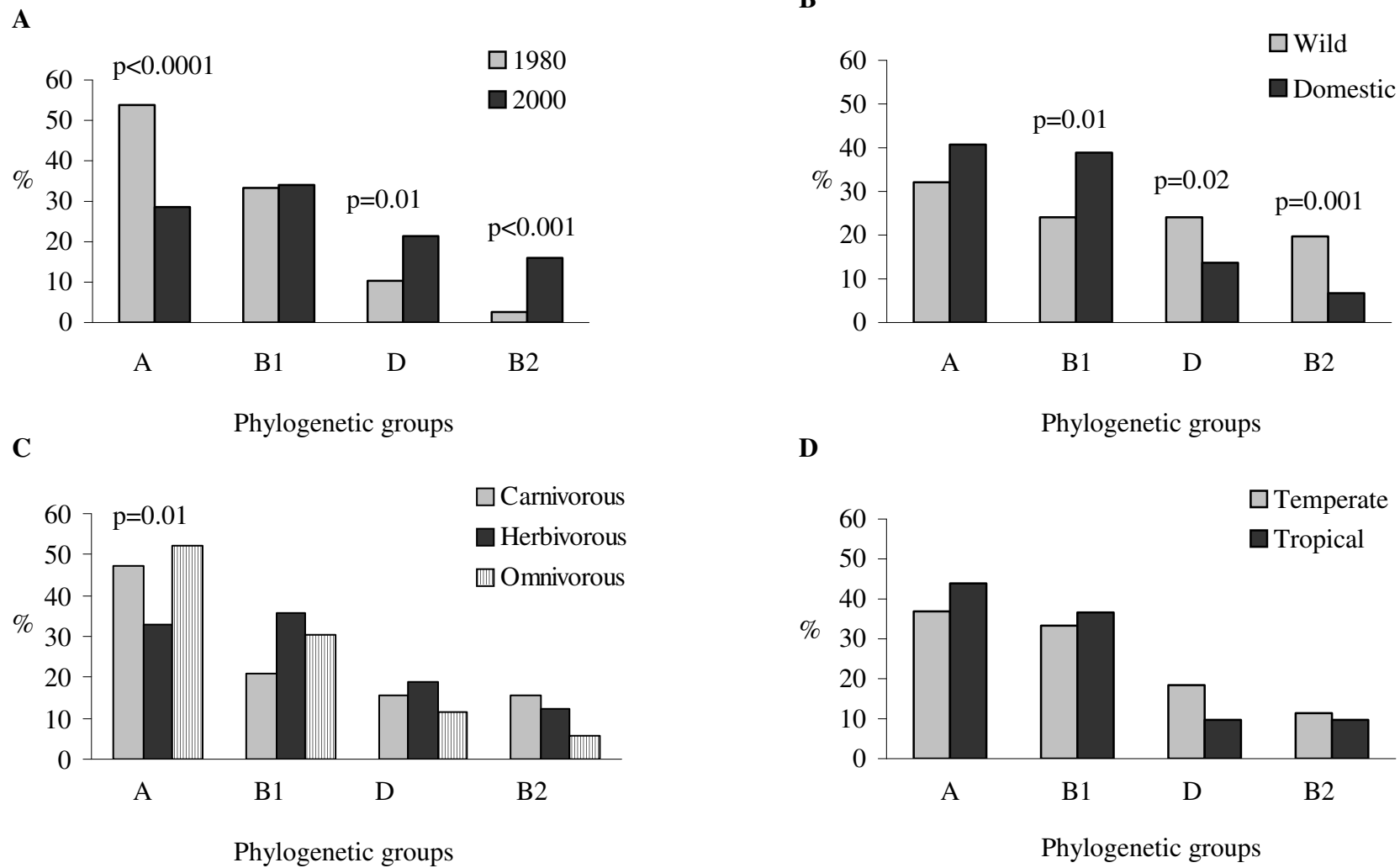


Figure 3