

## PROCEEDINGS B

**On the origin of mongrels: Evolutionary history of free-breeding dogs in Eurasia**

Journal:	<i>Proceedings B</i>
Manuscript ID	RSPB-2015-2189.R1
Article Type:	Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Pilot, Małgorzata; Polish Academy of Sciences, Museum and Institute of Zoology; University of Lincoln, School of Life Sciences Malewski, Tadeusz; Polish Academy of Sciences, Museum and Institute of Zoology Moura, Andre; University of Lincoln, School of Life Sciences Grzybowski, Tomasz; Ludwik Rydygier Collegium Medicum, Nicolaus Copernicus University, Division of Molecular and Forensic Genetics, Department of Forensic Medicine Oleński, Kamil; University of Warmia and Mazury, Department of Animal Genetics Ruść, Anna; University of Warmia and Mazury, Department of Animal Genetics Kamiński, Stanisław; University of Warmia and Mazury, Department of Animal Genetics Fadel, Fernanda; University of Lincoln, School of Life Sciences Mills, Daniel; University of Lincoln, School of Life Sciences Alagaili, Abdulaziz; King Saud University Mohammed, Osama; King Saud University Kłys, Grzegorz; University of Opole, Department of Biosystematics Okhlopov, Innokentiy; Siberian Branch of Russian Academy of Sciences, Science Institute of Biological Problems Cryolithozone Suchecka, Ewa; Polish Academy of Sciences, Museum and Institute of Zoology Bogdanowicz, Wieslaw; Polish Academy of Sciences, Museum and Institute of Zoology
Subject:	Genomics < BIOLOGY, Evolution < BIOLOGY
Keywords:	Canis lupus familiaris, Biogeographic reconstruction, Expansion wave, Free-breeding dogs, Dog origin
Proceedings B category:	Evolutionary Biology



1 **On the origin of mongrels: Evolutionary history of free-breeding dogs in**  
2 **Eurasia**

3 Małgorzata Pilot<sup>1,2</sup>, Tadeusz Malewski<sup>1</sup>, Andre E. Moura<sup>2</sup>, Tomasz Grzybowski<sup>3</sup>, Kamil  
4 Oleński<sup>4</sup>, Anna Ruś<sup>4</sup>, Stanisław Kamiński<sup>4</sup>, Fernanda Fadel<sup>2</sup>, Daniel Mills<sup>2</sup>, Abdulaziz N.  
5 Alagaili<sup>5</sup>, Osama B. Mohammed<sup>5</sup>, Grzegorz Kłys<sup>6</sup>, Innokentiy M. Okhlopov<sup>7</sup>, Ewa  
6 Suchecka<sup>1</sup> and Wiesław Bogdanowicz<sup>1\*</sup>

7 **Affiliations**

8 <sup>1</sup>Museum and Institute of Zoology, Polish Academy of Sciences, Wilcza 64, 00-679  
9 Warszawa, Poland. E-mails: tmalewski@miiz.waw.pl (TM); esuchecka@miiz.waw.pl (ES);  
10 wieslawb@miiz.waw.pl (WB)

11 <sup>2</sup>School of Life Sciences, University of Lincoln, Green Lane, Lincoln LN6 7DL, UK. E-  
12 mails: MPilot@lincoln.ac.uk (MP); amoura@lincoln.ac.uk (AEM); ffadel@lincoln.ac.uk  
13 (FF); dmills@lincoln.ac.uk (DM)

14 <sup>3</sup>Division of Molecular and Forensic Genetics, Department of Forensic Medicine, Ludwik  
15 Rydygier Collegium Medicum, Nicolaus Copernicus University, Skłodowskiej-Curie 9, 85-  
16 094 Bydgoszcz, Poland. E-mail: tgrzyb@cm.umk.pl

17 <sup>4</sup>Department of Animal Genetics, University of Warmia and Mazury, Oczapowskiego 5, 10-  
18 711 Olsztyn, Poland. E-mails: kamel@uwm.edu.pl (KO); anna.rusc@uwm.edu.pl (AR);  
19 stachel@uwm.edu.pl (SK)

20 <sup>5</sup>KSU Mammals Research Chair, Department of Zoology, College of Science, King Saud  
21 University, P.O. Box 2455, Riyadh 11451, Saudi Arabia. E-mails: aziz99@gmail.com  
22 (ANA); obmkkwrc@yahoo.co.uk (OBM)

23 <sup>6</sup>Department of Biosystematics, University of Opole, Oleska 22, 45-052 Opole, Poland. E-  
24 mail: gklys@uni.opole.pl

25 <sup>7</sup>Science Institute of Biological Problems Cryolithozone, Siberian Branch of Russian  
26 Academy of Sciences, Yakutsk 677980, Russia. E-mail: imokhlopkov@yandex.ru

27 \*Author for correspondence: Wiesław Bogdanowicz, e-mail: wieslawb@miiz.waw.pl

28

29 **Abstract**

30 **Although a large part of the global domestic dog population is free-ranging and free-**  
31 **breeding, knowledge of genetic diversity in these free-breeding dogs (FBDs) and their**  
32 **ancestry relations to pure-breed dogs is limited, and indigenous status of FBDs in Asia is**  
33 **uncertain. We analyse genome-wide SNP variability of FBDs across Eurasia, and show**  
34 **that they display weak genetic structure, and are genetically distinct from pure-breed**  
35 **dogs rather than constituting an admixture of breeds. Our results suggest that modern**  
36 **European breeds originated locally from European FBDs. East Asian and Arctic breeds**  
37 **show closest affinity to East Asian FBDs, and they both represent earliest-branching**  
38 **lineages in the phylogeny of extant Eurasian dogs. Our biogeographic reconstruction of**  
39 **ancestral distributions indicates a gradual westward expansion of East Asian indigenous**  
40 **dogs to the Middle East and Europe through Central and West Asia, providing evidence**  
41 **for a major expansion that shaped the patterns of genetic differentiation in modern**  
42 **dogs. This expansion was probably secondary and could have led to the replacement of**  
43 **earlier resident populations in Western Eurasia. This could explain why earlier studies**  
44 **based on modern DNA suggest East Asia as the region of dog origin, while ancient DNA**  
45 **and archaeological data point to Western Eurasia.**

46

47 **Keywords:** *Canis lupus familiaris*, Biogeographic reconstruction, Dog origin, Expansion  
48 wave, Free-breeding dogs

49

50

## 51 **1. Introduction**

52           The global dog population has been estimated at 1 billion individuals [1], with about  
53 75% of this population classified as free-ranging [2]. Free-ranging dogs may be owned but  
54 not permanently restrained, semi-feral or feral [2]. Their common characteristic is that they  
55 are not artificially restricted in individual mate choice, i.e. are free-breeding (a term we use  
56 after [3]). While the population genetic structure of pure-breed dogs is defined by human  
57 breeding practices [4, 5], the genetic structure of free-breeding dogs (FBDs) is expected to be  
58 largely defined by ecological and evolutionary processes (like dispersal patterns, mate choice,  
59 natural selection), while still being affected by certain human activities (e.g. translocations,  
60 introduction of non-native dog breeds). Unrestricted mate choice has thus major evolutionary  
61 implications.

62           Close breeding practices resulting in the development of modern dog breeds have only  
63 been introduced in the last few centuries [6], and the breed formation process was associated  
64 with severe bottlenecks and a large increase in linkage disequilibrium [7, 8]. Therefore, FBDs  
65 that did not experience these breeding practices may be better suited to reconstruct events at  
66 earlier stages of the dog history preceding the origin of modern breeds. However, this depends  
67 on whether they represent indigenous populations (i.e. deriving from ancestors native for a  
68 region they occupy) instead of being a recent admixture of modern breeds or originating from  
69 recent translocations.

70           The indigenous status of FBDs has been explicitly assessed in Africa [9, 10], the  
71 Americas [10, 11], and recently also in Oceania and southern parts of Europe and Asia [10].  
72 African FBDs were shown to be a mosaic of indigenous dogs genetically distinct from non-  
73 African breed dogs, and non-native, mixed-breed individuals [9, 10]. FBDs from South and  
74 North America (except for the Arctic regions) and from South Pacific mostly descend from

75 European dogs, with indigenous American dogs contributing to only a small fraction of the  
76 modern gene pool [10, 11]. In contrast, in FBDs from central and south Asia native ancestry  
77 components predominate [10].

78         Although Eurasia is a particularly important region in dog's evolutionary history,  
79 being the continent where the domestication took place [5, 10, 12-17], earlier studies focused  
80 mostly on FBD populations from southern parts of Asia [10, 12, 15-17], while little is known  
81 about FBDs from central and northern Eurasia. Recently, it has been shown that Arctic dog  
82 breeds trace a part of their ancestry to ancient Siberian wolves [18], implying that north Asia  
83 is an important region for the dog's evolutionary history. Therefore, for accurate  
84 reconstruction of this history, the analysis of genetic variability in populations from both  
85 southern and northern parts of Eurasia is required.

86         Understanding the ancestral status of Eurasian FBDs may also shed light on the origin  
87 of pure-breed dogs. Some breeds, mostly of non-European origin (Supplementary Table 1),  
88 have been classified as “ancient” based on their early branching in the phylogeny of pure-  
89 breed dogs [4, 5], and it has been suggested that they “may be the best living representatives  
90 of the ancestral dog gene pool” [4]. Alternatively, this branching pattern can simply reflect  
91 geographic isolation of these breeds and their consequent genetic differentiation from modern  
92 European breeds [6]. Reconstructing the phylogenetic relationships between these breeds and  
93 regional FBD populations may improve our understanding of relationships between different  
94 breeds and provide correct interpretation for the observed branching pattern.

95         In this study, we compared genome-wide SNP profiles of 200 FBDs from across  
96 Eurasia (Fig. 1) with 51 “ancient” and modern breeds (combining newly-generated and  
97 published datasets; [19]) in order to understand the relationships between these groups, assess  
98 the indigenous status of Eurasian FBDs, reconstruct their population genetic structure, and  
99 infer past phylogeographic events in Eurasia.

100

101 **2. Material and Methods**102 **Datasets**

103 We collected blood samples from 234 free-breeding domestic dogs from 14 sites across  
104 Eurasia (Fig. 1A). Multiple samples were collected from each site (Supplementary Table 2).  
105 All these samples were genotyped with CanineHD BeadChip (Illumina) at 167,989 autosomal  
106 SNP loci (referred to as 168K) and 5,660 X chromosome SNP loci, together with four grey  
107 wolves from Armenia (the South Caucasus). We identified and removed closely related  
108 individuals from this dataset (see Supplementary Text), reducing it to 200 unrelated  
109 individuals. This dataset will be referred to as "FBD dataset".

110 This dataset was complemented with two datasets of SNP genotypes of pure-breed dogs  
111 (Supplementary Table 3). The first dataset consisted of 96 pure-breed or crossed-breed dogs  
112 collected from across the United Kingdom using Performagene saliva sample collection kits  
113 (DNA Genotek). These dogs represented 30 breeds (88 individuals, with 1–9 individuals per  
114 breed; Supplementary Table 3) and five types of crosses between two known breeds (8  
115 individuals, with 1–3 individuals per cross type). This dataset will be referred to as "UK  
116 dataset". The second dataset was a publicly available dataset from the LUPA project [19],  
117 which contained 446 pure-breed dogs representing 30 different breeds (with 10–26 individuals  
118 per breed). It will be referred to as "LUPA dataset".

119 These additional datasets were both generated using CanineHD BeadChip, the same as  
120 the FBD dataset, and therefore all three datasets could be merged without a reduction of the  
121 usable SNP set. Correct merging of the datasets was confirmed by the joint clustering of  
122 individuals representing the same breed, independent of whether they originated from the UK  
123 or LUPA datasets.

124 The initial set of 168K autosomal loci was pruned using PLINK [20] from loci with  
125 minor allele frequency (MAF) below 0.01 and those with missing data for more than 10% of  
126 individuals. The X chromosome loci were also removed from all datasets. This resulted in a  
127 set of 147,836 loci when the FBD dataset was analysed separately, and 147,485 loci when all  
128 the tree datasets were analysed together. For some analyses (highlighted throughout the text),  
129 a dataset pruned from loci in strong linkage disequilibrium (LD) was required. It was obtained  
130 by further pruning the dataset from SNPs with an  $r^2 < 0.5$  within a 50 SNP sliding window,  
131 with a 10 SNPs step size (where  $r^2$  is a squared correlation in genotype frequencies between  
132 loci). The LD-pruning resulted in a set of 108,610 loci when the FBD dataset was analysed  
133 separately, and 104,769 loci when all the three datasets were analysed together.

134 In LD and PCA analyses, we also included a newly generated dataset of 79 grey wolves  
135 from different parts of Asia: the Caucasus (26 individuals), Mongolia (14 individuals), Saudi  
136 Arabia (2 individuals) and Siberia (37 individuals). This dataset was also generated using  
137 CanineHD BeadChip, and therefore could be merged with the three dog datasets without  
138 reducing the usable SNP set. The combined dataset was pruned to remove loci with  
139 MAF < 0.01 and those with missing data for more than 10% of individuals as well as X-  
140 chromosome loci, which resulted in a set of 147,483 loci. The LD-pruning ( $r^2 < 0.5$ ) resulted  
141 in 110,112 loci.

142

### 143 **Analysis of genetic differentiation in Eurasian FBDs**

144 We analysed population genetic structure using the LD-pruned FBD dataset. We used  
145 the Bayesian clustering methods with no prior population information as implemented in  
146 ADMIXTURE [21] and STRUCTURE [22]. In addition, we carried out a spatially explicit analysis  
147 of genetic structure using the software GENELAND [23]. Because genetic clustering methods

148 do not perform well in populations that exhibit an isolation-by-distance pattern of the spatial  
149 distribution of genetic diversity [24], we assessed whether such pattern is present in the FBDs  
150 across Eurasia using a simple (univariate) Mantel test implemented in GENALEX 6.5 [25]. We  
151 also used GENALEX to carry out a spatial autocorrelation analysis based on pair-wise  $F_{ST}$   
152 values between 14 sampling sites. The details of all these analyses are described in  
153 Supplementary Materials.

154 We calculated pair-wise isolation-by-state (IBS) distances between all individuals  
155 from the FBD dataset in PLINK, and used a matrix of (1-IBS) values to construct a  
156 neighbour-joining tree representing genetic differentiation among individuals from different  
157 local populations, using the software MEGA6 [26]. To identify the dominant components of  
158 variability within the FBDs, we performed a Principal Components Analysis (PCA) using the  
159 SMARTPCA program from EIGENSOFT [27] package. EIGENSOFT was also used to estimate  
160 average divergence between and within 14 sampling sites, as well as pair-wise  $F_{ST}$  between  
161 the sites.

162

### 163 **Heterozygosity, autozygosity and linkage disequilibrium analysis**

164 Estimates of heterozygosity and linkage disequilibrium (LD) are dependent on sample  
165 size. Therefore, we randomly selected 9 unrelated individuals from all FBD populations that  
166 originally had larger number of samples. This way we obtained the equal sample size of 9  
167 individuals for all but two local populations (with sample sizes of 4 and 5 individuals).

168 Using PLINK, we calculated observed and expected heterozygosity in each population.  
169 We also assessed autozygosity levels by identifying runs of homozygosity (ROHs), longer  
170 than 100 kb and spanning at least 25 SNPs, in individuals from each population. The LD-

171 pruned dataset (with the threshold of  $r^2 < 0.5$ ) was used in this analysis to avoid identifying  
172 ROHs resulting from strong LD rather than from autozygosity.

173 For selected local populations representing each of the four main regions (East Asia,  
174 the Middle East, Central/West Asia and Europe), we calculated genome-wide pair-wise  
175 genotypic association coefficient  $r^2$  between all autosomal SNPs with  $MAF > 0.15$ , which  
176 provided us with an estimate of LD. In addition, we also analysed LD for grey wolf  
177 populations from Mongolia and Armenia, based on 9 individuals each. We estimated effective  
178 population sizes ( $N_E$ ) based on the extent of LD, to compare the demographic trends between  
179 FBDs from different regions (see details in Supplementary Material).

180

#### 181 **Analysis of genetic differentiation between FBDs and pure-breed dogs**

182 For the combined SNP dataset consisting of FBDs and pure-breed dogs (LUPA and  
183 UK dataset), we constructed a PCA plot as well as neighbour-joining trees of both inter-  
184 population  $F_{ST}$  coefficients and inter-individual IBS distances, as described above. We also  
185 re-ran this analysis with the addition of the grey wolf data.

186 Because uneven sample sizes of different populations may bias the PCA results [28],  
187 we also carried out the PCA after randomly removing individuals from some populations to  
188 obtain more even sample sizes: 93 FBDs (7 per sampling site except for two sites with lower  
189 sample number), 87 individuals representing European breeds (1 or 2 individuals per breed),  
190 31 individuals representing East Asian and Arctic breeds (all available), 12 Eurasiers (all  
191 available) and 83 grey wolves (all available).

192

#### 193 **Analysis of admixture among dog breeds and FBDs**

194 We analysed patterns of admixture among pure-breed and FBDs using the program  
195 TREEMIX [29]. This analysis was carried out for the combined dataset consisting of free-  
196 breeding and pure-breed dogs (FBD, UK and LUPA datasets), with Caucasian grey wolves as  
197 an outgroup. The UK dataset included some known cross-breed individuals, which were  
198 removed from this analysis, because the presence of cross-breeds affected the tree topology  
199 by clustering parental breeds together (e.g. Labradors and other retrievers clustered with  
200 Poodles in the presence of Labradoodles; see Supplementary Fig. 10). We also removed dog  
201 breeds represented by one individual only, and a few pure-breed individuals that did not  
202 cluster with their alleged breed.

203 We constructed the maximum likelihood trees containing both the pure-breeds and  
204 free-breeding populations assuming (a) no post-divergence gene flow among populations, and  
205 (b) 10 gene flow events. Although the LD-pruned dataset was used, to further account for LD  
206 we constructed the trees using blocks of 100 SNPs rather than individual SNPs. For the tree  
207 with no gene flow, we generated 1,000 bootstrap replicates by re-sampling blocks of 100  
208 SNPs. For the tree with 10 gene flow events, we generated only 100 replicates due to long  
209 computational time. For this tree, a jackknife analysis was used to assess whether the  
210 inclusion of each migration edge significantly improved the fit of this phylogenetic model to  
211 the data. We also constructed a tree assuming 15 gene flow events, to test whether the  
212 addition of more migration edges changes the tree topology.

213

#### 214 **Reconstruction of distributions of ancestral dog populations**

215 We used the software RASP [30], which estimates the occurrence of migration and  
216 vicariance events along a user-defined phylogenetic tree, to reconstruct the distribution of  
217 ancestral dog populations. For this purpose, we used the phylogenetic trees constructed in

218 TREEMIX (one assuming no post-divergence gene flow and the second assuming 10 gene flow  
219 events). The following distribution ranges were considered: Europe, Central/West Asia,  
220 Middle East, East Asia, and East Russia. Arctic breeds were assigned to East Asia, according  
221 to their primary origin [31-33]. Eurasier was assigned to both Europe and East Asia. Grey  
222 wolf, used as an outgroup, occurs in all the distribution ranges considered, and therefore was  
223 uninformative in this analysis.

224 We used the parsimony-based Statistical Dispersal-Vicariance Analysis (S-Diva) [34]  
225 and the Bayesian Binary (BB) Markov Chain Monte Carlo method [35] to estimate  
226 uncertainty in the reconstruction of ancestral distributions. A maximum of five geographic  
227 regions per node were considered, and an uninformative distribution was applied to the root.  
228 For the BB method, a total of 10 MCMC chains were run, 9 of which were heated, with  
229 10,000,000 iterations and 20% burn-in.

230

### 231 **3. Results**

#### 232 **Genetic differentiation in Eurasian FBDs**

233 Although genetic differentiation among Eurasian FBDs was relatively weak, different  
234 population clustering methods we used supported a division into three large-scale regions:  
235 East Asia, the Middle East, and Western Eurasia (Fig. 1). Western Eurasia was further sub-  
236 divided into Europe and Central/West Asia, based on both geographic proximity and genetic  
237 similarity of local populations (Fig. 1; Supplementary Figs 1-4). The population from East  
238 Russia, geographically belonging to East Asia, was genetically more similar to Central/West  
239 Asia. Therefore, it was considered as a separate region. We describe the results of population  
240 genetic and phylogenetic analyses that support this division in Supplementary Materials.

241           Despite the weak differentiation, some meaningful patterns could be identified. Most  
242 East Asian dogs (from China, Thailand and Mongolia) branched from basal nodes of the IBS  
243 tree, even though they did not group into a single clade (Supplementary Fig. 3). Multiple  
244 individuals from Thailand and one from Mongolia grouped together with dogs from western  
245 Eurasia, which may reflect recent gene flow of “western” dogs into East Asia. Individuals  
246 from Europe, West Asia and Central Asia did not group into clades consistent with  
247 geography, suggesting that they may constitute one genetic population (Supplementary Fig.  
248 3). Dogs from East Russia were also part of this large admixed group in the IBS tree, and  
249 clustered with West Eurasia in all other analyses (Supplementary Figs. 1-2).

250

#### 251 **Heterozygosity and linkage disequilibrium in Eurasian FBDs**

252           Observed heterozygosity in the 14 sampling sites varied between 0.30 and 0.35, and  
253 no consistent differences in heterozygosity were found between the four main regions of  
254 Eurasia (Supplementary Table 4). No consistent differences occurred between autozygosity  
255 levels, either (Supplementary Fig. 5, Supplementary Table 5).

256           The Chinese FBD population had lower linkage disequilibrium (LD) compared with  
257 other populations for all genetic distance classes (Supplementary Fig. 6A). FBD populations  
258 from Europe, Central/West Asia and the Middle East had similar LD levels for small distance  
259 classes (1.25–115 kb), supporting their common origin. The populations from Thailand and  
260 Mongolia had intermediate  $r^2$  values between the Chinese population and all other populations  
261 for distance classes between 1.25 and 60 kb, while their  $r^2$  values for larger distance classes  
262 were comparable with West Eurasian FBD populations (Supplementary Fig. 6B, C).

263           For short distance classes (1.25–40 kb) grey wolf populations had lower LD than  
264 FBDs (Supplementary Fig. 6B), as expected for an ancestral group. However, for long

265 distance classes (275–1,000 Kb) LD was higher in wolves than in FBDs (Supplementary Fig.  
266 6C), consistent with a long-term decline in wolf numbers in Eurasia [36, 37]. LD decayed  
267 below  $r^2 = 0.5$  at 3.75 kb in Chinese FBDs, 5–7.5 kb in other FBD populations, and 2.5 kb in  
268 wolves. Chinese FBDs had higher  $N_E$  estimates (inferred from LD) than any other FBD  
269 population throughout all the time periods assessed. The populations from Thailand and  
270 Mongolia had intermediate  $N_E$  estimates between the Chinese population and all the  
271 remaining populations until about 2,500 years ago (see Supplementary Fig. 7 and  
272 Supplementary Text).

273

#### 274 **Genetic differentiation between FBDs and pure-breed dogs**

275 PCA placed FBDs in intermediate positions between groups of pure-breed dogs  
276 (Supplementary Fig. 8). The majority of modern European breeds were clustered together,  
277 and only a few free-breeding individuals grouped within this cluster, suggesting relatively low  
278 gene flow from pure-breed dogs into FBDs. FBDs from Slovenia and Poland were placed  
279 closer to the cluster of European breeds than any other FBD populations, which is consistent  
280 with a local origin of modern European breeds. German shepherd occupied an outlier position  
281 in the PCA (Supplementary Fig. 8A), which was unexpected, but consistent with other  
282 analyses in this study.

283 Breeds of East Asian (Shar Pei, Shiba Inu) and Arctic origin (Greenland Sledge Dog,  
284 Alaskan Malamute, Siberian Husky) were placed at the opposite end of PC1 relative to the  
285 European breeds cluster. The Arctic breeds originated in East Asia [31-33], so close  
286 clustering of these two groups of breeds reflects their common origin. FBDs were placed  
287 between these two extremes, with East Asian FBDs grouping closer to East Asian breeds, and  
288 European FBDs closer to European breeds (Supplementary Fig. 8C).

289           The inclusion of grey wolves into the PCA shows that dogs and wolves form clearly  
290 separated clusters, suggesting that gene flow between FBDs and wolves (revealed in our  
291 TREEMIX analysis; see below) has not affected the genetic integrity of these populations  
292 (Supplementary Fig. 9A). At PC1, East Asian and Arctic breeds showed closest proximity to  
293 wolves of all the pure-breed dogs and FBDs. However, when this analysis was re-run with  
294 more balanced sample sizes for all groups (FBDs, breed dogs and wolves), East Asian and  
295 Middle Eastern FBDs showed similar level of proximity to wolves at PC1 as East Asian and  
296 Arctic breeds (Supplementary Fig. 9B). Despite these differences, all the PCA plots were  
297 consistent in showing genetic distinctiveness of FBDs from breed dogs, and distinctiveness of  
298 East Asian and Arctic breeds from modern European breeds.

299

### 300 **Phylogenetic reconstruction of relationships between FBDs and pure-breed dogs**

301           A maximum-likelihood tree of population divergence constructed in TREEMIX and a  
302 neighbour-joining tree of  $F_{ST}$ -distances among FBDs and pure-breed dogs, consistently  
303 inferred the earliest divergence for East Asian and Arctic breeds, followed by East Asian  
304 FBDs (Fig. 1B, Supplementary Figs. 10, 11). The early branching of East Asian breeds was  
305 inferred with 98–99% bootstrap support, and that of East Asian FBDs with 92–95% support  
306 (Fig. 1B, Supplementary Fig. 11). Modern European breeds were grouped in one clade, which  
307 also included Slovenian FBDs. All other FBDs were placed outside of this clade, with the  
308 Middle Eastern populations forming a distinct group from European and Central/West Asian  
309 populations.

310           Individual-based IBS distances produced a similar topology (Supplementary Fig. 12).  
311 Some individual FBDs clustered with particular dog breeds, suggesting their mixed-breed

312 origin. However, most FBDs formed separate groups from breed dogs, indicating their genetic  
313 distinctiveness.

314 In an IBS tree including pure-breed dogs only, East Asian and Arctic breeds branched  
315 from basal nodes, showing the consistency in this branching pattern for phylogenies with and  
316 without FBDs. Spitz-type breeds of European origin (Keeshond, Elkhound, Finnish spitz,  
317 German spitz, Schipperke) or mixed European and East Asian ancestry (Eurasier) were placed  
318 outside of the modern European breed clade (Supplementary Fig. 13), suggesting their genetic  
319 distinctiveness.

320

#### 321 **Admixture between FBDs, pure-breed dogs and grey wolves**

322 The TREEMIX analysis assuming 10 admixture events revealed post-divergence gene  
323 flow from grey wolves to Middle Eastern FBDs (Supplementary Fig. 14). We also identified  
324 gene flow from Keeshond to Eurasier, consistent with the origin of this last breed, which was  
325 developed by crossing Keeshond females with Chow Chow males [6]. Unexpectedly, we also  
326 detected gene flow from German shepherd to multiple FBD populations in Europe and  
327 Central/West Asia (Supplementary Fig. 14A), suggesting frequent mixing between this breed  
328 and FBDs.

329 Because the presence of German shepherd prevented the detection of other admixture  
330 cases, we re-ran the analysis after removing this breed from the dataset. This analysis revealed  
331 additional gene flow events, including: (a) from modern breeds to FBDs in Thailand, East  
332 Russia and Europe, (b) between Mongolian and Chinese FBDs, (c) between Arctic and East  
333 Asian breeds, and (d) between ancestral populations of modern European breeds  
334 (Supplementary Figs. 14B and 15). An analysis allowing 15 migration events instead of 10  
335 further revealed gene flow from wolves to Greenland Sledge Dogs, and additional cases of

336 gene flow between ancestral populations of modern European breeds (Supplementary Fig.  
337 14C). The addition of migration edges significantly improved the fit of the phylogenetic  
338 model to the data (Supplementary Text and Supplementary Fig. 16).

339       Importantly, accounting for gene flow affected the topology of dog phylogeny. In  
340 TREEMIX trees accounting for gene flow (assuming either 10 or 15 migration edges), Chinese  
341 and Thai FBDs formed the earliest-branching clade together with East Asian breeds  
342 (Supplementary Figs. 14 and 15), suggesting that East Asian breeds and FBDs have a  
343 common origin. KIMTREE analysis (see Supplementary Text) provided higher support for this  
344 last topology than the topology where only pure-breeds branched from the most basal node  
345 (Supplementary Fig. 17).

346

#### 347 **Reconstruction of the geographic distribution of ancestral dog populations**

348       We used the software RASP [30] to reconstruct the geographic distribution of  
349 ancestral dog populations, based on the TREEMIX trees both without and with gene flow. For  
350 both trees, this analysis indicated that the most recent common ancestor of extant dogs  
351 originated in East Asia (Fig. 1B, Supplementary Figs. 18 and 19). It also suggested a gradual  
352 westward expansion of dogs along two migration routes from East Asia (a) to the Middle East  
353 and (b) to Europe through Central and West Asia (Fig. 1A, B). It is important to stress that  
354 this finding concerns the most recent common ancestor of extant dogs, rather than the most  
355 recent common ancestor all dogs shared with their ancestral grey wolf population.

356

#### 357 **4. Discussion**

358 *Origin of FBDs in relation to pure-breed dogs*

359 Our results show that Eurasian FBD populations are genetically distinct from pure-  
360 breed dogs. Although we found mixed-breed individuals among FBDs, they constituted a  
361 small fraction of the entire population. Another study has recently reached a similar  
362 conclusion for FBDs from south and central Asia [10]. Taken together, these results suggest  
363 that most FBD populations in Asia represent lineages distinct from modern European breeds  
364 and probably native to their respective locations. Furthermore, we provided evidence for the  
365 long-term continuity of FBD lineages in East Asia by demonstrating their clustering with the  
366 dingo (Supplementary Text and Supplementary Fig. 20), which originated from East Asia and  
367 was isolated from other dog populations for at least 3,500 years before the arrival of  
368 Europeans and their dogs to Australia [6, 17, 38]. This shows that East Asian FBD  
369 populations are indigenous; however, similar as FBDs in other parts of Asia, they include a  
370 small fraction of non-native mixed-breed individuals.

371

### 372 *Population structure and genetic diversity in Eurasian FBDs*

373 Even though we sampled populations from discrete and distant locations (a sampling  
374 pattern that typically leads to overestimation of population structuring), we found no strong  
375 spatial genetic structure among Eurasian FBDs. Such a pattern may suggest a relatively recent  
376 common origin of all Eurasian FBD populations and/or intense admixture between regions.  
377 Declining spatial autocorrelation of genetic distances for geographic distance classes between  
378 1,000 and 4,000 km (Supplementary Fig. 4B) shows the importance of geographic distance in  
379 shaping population differentiation of FBDs. However, higher genetic similarity of East  
380 Russian FBDs to dogs from Central Russia and other countries formerly belonging to USSR  
381 (Kazakhstan, Tajikistan, Armenia) than to geographically closer dogs from China suggests  
382 that genetic differentiation of FBDs is also shaped by cultural/political divisions in human  
383 populations.

384           The weak genetic differentiation among FBDs from different parts of Eurasia can  
385 explain the lack of consistent differences in autosomal genetic variability between the four  
386 main regions of Eurasia. This result is similar to that based on Y chromosome data, which  
387 revealed comparably high diversity in South-West Asia, South-East Asia, Europe, Africa and  
388 Oceania [10, 15]. However, mitochondrial DNA data showed instead the highest haplotype  
389 diversity in South-East Asia [10, 12, 39] – a pattern which some studies interpreted as  
390 evidence for East Asian origin of the domestic dog [12, 39].

391           Population genetic models of spatial expansion are typically based on a serial founder  
392 effect model, which assumes a continuous decline in diversity along a colonisation route due  
393 to a series of bottlenecks, and no major phylogeographic changes after the initial colonisation  
394 [40]. However, genetic clines may also result from alternative scenarios involving extensive  
395 post-colonisation admixture [40]. Our TREEMIX analysis revealed a number of admixture  
396 events among different FBD populations, which could have contributed to the contrasting  
397 diversity patterns between mtDNA and nuclear DNA, especially if admixture was sex-biased.

#### 398 *Geographic patterns of FBDs expansion in Eurasia*

399           We found that the Chinese FBD population had lower LD and higher  $N_E$  estimates  
400 than other FBD populations, throughout all the time periods assessed. Such a pattern is  
401 expected from an ancestral population in comparison with derived populations, as illustrated  
402 by genetic studies on the origin of modern humans. A LD-based estimate of temporal  $N_E$   
403 changes in human populations showed a large reduction in  $N_E$  in non-Africans compared to  
404 Africans lasting between 125,000 and 10,000 years ago, providing support for the "Out of  
405 Africa" migration event [41]. Other East Asian FBD populations from Thailand and Mongolia  
406 had intermediate LD estimates between the Chinese populations and West Eurasian  
407 population for small distance classes. Increased LD and reduced  $N_E$  in West Eurasian FBD

408 populations as compared with East Asian populations are consistent with a migration event  
409 from East Asia westwards.

410 Identifying a precise geographic location of the source population for the inferred  
411 expansion would require denser sampling of Asian FBDs. Another study has recently found  
412 lowest LD at short inter-SNP distances in FBDs from Mongolia and Nepal [10], but did not  
413 include samples from China, and classified Mongolia as Central rather than East Asia.

414 In accordance with the inference from the LD pattern, we found that East Asian breeds  
415 and FBDs branch from basal nodes in the phylogeny of extant dogs (Fig. 1B and  
416 Supplementary Fig. 15). The biogeographic reconstruction of ancestral distributions using  
417 RASP showed a clear pattern of a gradual expansion of modern dogs from East Asia towards  
418 the Middle East and Europe, indicating that East Asia was a source population in a major  
419 migration event.

420 Patterns of Y chromosome variability also suggest a large and rapid expansion of dogs  
421 from East Asia westwards [17]. This expansion was dated at between 4,000 and 11,000 years  
422 ago (5,800 years ago, SE 1,750 or 8,400 years ago, SE 2,500, depending on the calibration;  
423 [17]), which is considerably later than current estimates of the time when the domestication  
424 process was initiated (~19,000–40,000 years ago [13, 18]). Such timing implies that this was a  
425 secondary rather than primary expansion wave, which could have led to the replacement of  
426 dog lineages that had earlier occupied Western Eurasia [17], potentially diluting evidence for  
427 the primary expansion.

428 Although the dating of this expansion event is not precise, it could be linked with the  
429 neolithisation process [17], and it could have occurred via trade and/or in association with  
430 spatial and demographic expansion of Neolithic humans [42, 43]. The dogs from the new  
431 expansion wave could have admixed with earlier resident populations – in parallel with the

432 admixture of expanding Neolithic humans with resident Mesolithic populations [42, 43].  
433 Alternatively, the new immigrants could have replaced earlier resident populations, similar to  
434 what was seen after the expansion of European dogs (and their human owners) in North  
435 America [11]. If Europe was the place of the primary dog origin [13], the replacement  
436 scenario is more likely, because in the case of admixture we should expect higher genetic  
437 diversity in Europe as compared with East Asia, as demonstrated in an example of the honey  
438 bee *Apis mellifera*, where diversity of admixed populations is higher compared with native  
439 populations [44].

#### 440 *Integration with archaeological and ancient DNA data*

441       The occurrence of the secondary expansion wave replacing earlier resident populations  
442 in Western Eurasia can account for discrepancies between earlier studies based on modern  
443 DNA analysis, suggesting East or Central Asia as the region of dog origin [10, 12, 14], and  
444 evidence from archaeological and ancient DNA data, pointing instead to Europe or West Asia  
445 [6, 13]. The occurrence of the secondary expansion event may also explain why "none of the  
446 ancient breeds derive from regions where the oldest archaeological remains have been found"  
447 [6]. The early-branching dog breeds from East Asia and the Arctic can be considered as  
448 "ancient" in the sense that they likely represent lineages older than modern European breeds.  
449 However, this does not imply a direct line of descent from the first domesticated population,  
450 which may be extinct [18, 36] or swamped by admixture.

451       We acknowledge that a major expansion from East Asia may be not the only scenario  
452 consistent with our data, however this conclusion is also supported by other independent  
453 datasets of modern Asian FBDs based on different types of genetic markers [10, 12, 15, 17,  
454 39]. These earlier studies differ in the precise location of the source population (South-East vs  
455 Central Asia) and in the interpretation of this expansion as a primary [10, 12, 15, 39] or

456 secondary [17] wave. In our opinion, this cannot be resolved without extensive analysis of  
457 archaeological dog samples from different parts of Asia.

458

#### 459 *Admixture patterns*

460 A recent study [18] provided evidence for introgression from a lineage of ancient  
461 Siberian wolves into Arctic and East Asian dog breeds (Siberian Husky, Greenland Sledge  
462 Dog, and Shar-Pei). This past admixture with wolves could result in earlier branching of these  
463 breeds relative to East Asian FBDs in the phylogeny of Eurasian dogs. Therefore, we used the  
464 TREEMIX approach to directly account for the post-divergence gene flow in the phylogenetic  
465 reconstruction. This resulted in a tree where Chinese and Thai FBDs formed the earliest-  
466 branching clade *together* with East Asian breeds (Supplementary Fig. 15), implying that these  
467 two dog groups have a common origin, and lineages they represent are older than lineages of  
468 modern European breeds. Importantly, the TREEMIX analysis revealed post-divergence gene  
469 flow from grey wolves to Greenland Sledge Dogs (Supplementary Fig. 14C), so the  
470 admixture event documented in [18] was accounted for.

471 The TREEMIX analysis also revealed a number of other admixture events that may  
472 have an important effect on the inference of the dog evolutionary history. For example, it  
473 revealed post-divergence gene flow from grey wolves to Middle Eastern FBDs, consistent  
474 with the inference from whole-genome data [36]. Although gene flow in the opposite  
475 direction was also inferred from whole-genome data, it was less intense (6–9% vs 12–14%;  
476 [36]) and remained undetected here because of the limited number of migration events  
477 assumed (10 or 15).

478 Many gene flow events we detected were known from earlier studies or from breed  
479 histories, confirming that our results are accurate. For example, the cross-breed origin of

480 Eurasier, resulting from an admixture between European and East Asian spitz-type dogs, was  
481 accurately inferred in our TREEMIX analysis. Because the geographic origin of Eurasier was  
482 both in Europe and East Asia, this resulted in ambiguous inference of the geographic  
483 distribution for the common ancestor of Eurasier and East Asian dogs in the RASP analysis  
484 based on the TREEMIX tree with 10 migration edges (see Supplementary Fig. 19A).

485 We also detected gene flow from modern breeds to FBDs in different parts of Eurasia.  
486 However, the tree of individual-based IBS distances showed that this is due to the presence of  
487 individual cross-breed dogs among FBDs. Most FBDs clustered separately from pure-breed  
488 dogs, further supporting our conclusion that FBDs are distinct genetic units rather than the  
489 result of ongoing admixture between breeds.

490

## 491 **5. Conclusions**

492 We presented here a large-scale assessment of genome-wide variability of Eurasian  
493 FBDs, showing that they are genetically distinct from pure-breed dogs, and their inclusion is  
494 necessary for a complete representation of genetic variability of extant dogs. We provided  
495 evidence that East Asian FBD populations are indigenous (although they include a fraction of  
496 mixed-breed individuals), while FBDs from West Asia and Europe derive from an ancient  
497 expansion of East Asian dogs. This expansion was probably secondary [17] and could have  
498 led to the replacement of earlier resident populations in Western Eurasia. The occurrence of  
499 such secondary expansion wave can account for discrepancies between studies aimed at  
500 identifying the region of primary dog domestication based on modern DNA analysis with  
501 those based on archaeological and ancient DNA data. We also presented evidence for  
502 admixture between different FBD populations and for hybridisation with wolves. The picture

503 emerging from our results shows a very complex post-domestication history of the dog, which  
504 was as eventful as the history of humans.

505

506 **Ethics:** Blood samples from free-ranging dogs were obtained by veterinarians or veterinary  
507 technicians. Samples from pure-breed dogs were obtained using saliva sample collection, with  
508 the owners' consent. No animal was harmed for the purpose of this study. The study was  
509 approved by the National Science Centre in Poland and the Museum and Institute of Zoology,  
510 Polish Academy of Sciences.

511

512 **Data accessibility:** SNP genotypes generated in this study are available from Dryad:  
513 doi:10.5061/dryad.078nc. Geographic locations of samples are provided in the Supplementary  
514 Material.

515

#### 516 **Author contributions**

517 MP participated in the design of the study, carried out the data analysis and wrote the  
518 manuscript; TM collected the samples and participated in the data analysis; AEM carried out  
519 the data analysis and helped draft the manuscript; TG participated in the design of the study;  
520 KO, AR and SK produced the SNP genotype data; FF, ANA and OBM collected the samples  
521 and extracted DNA; DM, GK and IMK collected the samples; ES extracted DNA; WB  
522 participated in the design of the study, coordinated the project, collected the samples and  
523 helped draft the manuscript. All authors read and approved the final manuscript.

524

525 **Competing interests:** The authors declare no competing interests.

526

527 **Funding:** This project was funded by the National Science Centre in Poland (grant No.  
528 2011/01/B/NZ8/02978). Additional funding was provided by the University of Lincoln, UK  
529 (Returners Research Fund) and the Deanship of Scientific Research at the King Saud  
530 University, Saudi Arabia (project number RGP\_020).

531

### 532 **Acknowledgments**

533 We thank colleagues who helped with collecting samples of free-breeding dogs: Malik H. Ali,  
534 Sergei V. Aramilev, Mika Bagramyan, Nikolay Baskakov, Sergey Belokobylskij, Mateusz  
535 Golan, Alexander V. Gromov, Maria Hołyńska, Thanapol Nongbua, Ksenija Praper, Akylbek  
536 Ryspaev, La-orsri Sanoamuang, Elena A. Sazhenova, Jacek Szwedo, Mikhail P. Tiunov,  
537 Elena Tsingarska, and Odbayar Ts. We thank owners of pure-breed dogs who provided the  
538 samples used in this project. We also thank E. Yaruyan, E. A. Nikolayev, N. V. Mamayev,  
539 and Odbayar Ts for samples of wolves. We thank Matthew Webster for his advice regarding  
540 the LUPA dataset. We are grateful to Robert D. Owen, Cassandra M. Miller-Butterworth,  
541 Oscar Gaggiotti and two anonymous referees for their helpful comments on the manuscript.

542

### 543 **References**

- 544 1. Gompper ME. 2014. The dog-human-wildlife interface: Assessing the scope of the  
545 problem. In *Free-ranging dogs and wildlife conservation* (ed ME Gompper), pp. 9–54.  
546 Oxford, UK: Oxford University Press.  
547 (doi:10.1093/acprof:osobl/9780199663217.003.0001)
- 548 2. Hughes J, Macdonald DW. 2013. A review of the interactions between free-roaming  
549 domestic dogs and wildlife. *Biol Conserv.* **157**, 341–51.
- 550 3. Boyko RH, Boyko AR. 2014. Dog conservation and the population genetic structure of  
551 dogs. In *Free-ranging dogs and wildlife conservation* (ed ME Gompper), pp. 185–210.

- 552 Oxford, UK: Oxford University Press.  
553 (doi:10.1093/acprof:osobl/9780199663217.003.0008)
- 554 4. Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, Johnson GS,  
555 DeFrance HB, Ostrander EA, *et al.* 2004. Genetic structure of the purebred domestic dog.  
556 *Science* **304**, 1160–4. (doi:10.1126/science.1097406)
- 557 5. vonHoldt BM, Pollinger JP, Lohmueller KE, Han E, Parker HG, Quignon P, Degenhardt  
558 JD, Boyko AR, Earl DA, Auton A, *et al.* 2010. Genome-wide SNP and haplotype analyses  
559 reveal a rich history underlying dog domestication. *Nature* **464**, 898–902.  
560 (doi:10.1038/nature08837)
- 561 6. Larson G, Karlsson EK, Perri A, Webster MT, Ho SYW, Peters J, Stahl PW, Piper PJ,  
562 Lingaas F, Fredholm M, *et al.* 2012. Rethinking dog domestication by integrating  
563 genetics, archeology, and biogeography. *Proc Natl Acad Sci USA*. **109**, 8878–8883.  
564 (doi:10.1073/pnas.1203005109)
- 565 7. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M,  
566 Chang JL, Kulbokas EJ III, Zody MC, *et al.* 2005. Genome sequence, comparative  
567 analysis and haplotype structure of the domestic dog. *Nature*. **438**, 803–819.  
568 (doi:10.1038/nature04338)
- 569 8. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, Zhao K,  
570 Brisbin A, Parker HG, vonHoldt BM, *et al.* 2010. A simple genetic architecture underlies  
571 quantitative traits in dogs. *PLoS Biol.* **8**, e1000451. (doi:10.1371/journal.pbio.1000451)
- 572 9. Boyko AR, Boyko RH, Boyko CM, Parker HG, Castelhana M, Corey L, Degenhardt JD,  
573 Auton A, Hedimbi M, Kityo R, *et al.* 2009. Complex population structure in African  
574 village dogs and its implications for inferring dog domestication history. *Proc Natl Acad*  
575 *Sci USA*. **106**, 13903–8. (doi: 10.1073/pnas.0902129106)

- 576 10. Shannon LM, Boyko RH, Castelhana M, Corey E, Hayward JJ, *et al.* 2015. Genetic  
577 structure in village dogs reveals a Central Asian domestication origin. *Proc Natl Acad Sci*  
578 *USA*. (early edition) (doi: 10.1073/pnas.1516215112)
- 579 11. Castroviejo-Fisher S, Skoglund P, Valadez R, Vila C, Leonard JA. 2011. Vanishing native  
580 American dog lineages. *BMC Evol Biol.* **11**, 73. (doi:10.1186/1471-2148-11-73)
- 581 12. Pang J-F, Klütsch C, Zou X-J, Zhang A-B, Luo L-Y, Angleby H, Ardalán A, Ekström C,  
582 Sköllermo A, Lundeberg J, *et al.* 2009. MtDNA data indicate a single origin for dogs  
583 south of Yangtze river, less than 16,300 years ago, from numerous wolves. *Mol Biol Evol.*  
584 **26**, 2849–64. (doi:10.1093/molbev/msp195).
- 585 13. Thalmann O, Shapiro B, Cui P, Schuenemann VJ, Sawyer SK, Greenfield DL, Germonpré  
586 MB, Sablin MV, López-Giráldez F, Domingo-Roura X, *et al.* 2013. Complete  
587 mitochondrial genomes of ancient canids suggest a European origin of domestic dogs.  
588 *Science.* **342**, 871–874. (doi:10.1126/science.1243650)
- 589 14. Wang G-D, Zhai W, Yang H-C, Fan R-X, Cao X, Zhong L, Wang L, Liu F, Wu H, Cheng  
590 L-G, *et al.* 2013. The genomics of selection in dogs and the parallel evolution between  
591 dogs and humans. *Nature Commun.* **4**, 1860. (doi:10.1038/ncomms2814)
- 592 15. Ding ZL, Oskarsson M, Ardalán A, Angleby H, Dahlgren L-G, Tepeli C, Kirkness E,  
593 Savolainen P, Zhang Y-P. 2012. Origins of domestic dog in Southern East Asia is  
594 supported by analysis of Y-chromosome DNA. *Heredity.* **108**, 507–14.  
595 (doi:10.1038/hdy.2011.114)
- 596 16. Brown SK, Pedersen NC, Jafarishorijeh S, Bannasch DL, Ahrens KD, Wu J-T, *et al.*  
597 2011. Phylogenetic distinctiveness of Middle Eastern and Southeast Asian village dog Y  
598 chromosomes illuminates dog origins. *PLoS ONE* 6(12): e28496.  
599 (doi:10.1371/journal.pone.0028496 )

- 600 17. Sacks BN, Brown SK, Stephens D, Pedersen NC, Wu J-T, Berry O. 2013. Y chromosome  
601 analysis of dingoes and Southeast Asian village dogs suggests a Neolithic continental  
602 expansion from Southeast Asia followed by multiple Austronesian dispersals. *Mol Biol*  
603 *Evol.* **30**, 1103–18. (doi:10.1093/molbev/mst027)
- 604 18. Skoglund P, Ersmark E, Palkopoulou E, Dalén L. 2015. Ancient wolf genome reveals an  
605 early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr*  
606 *Biol.* **25**, 1–5. (doi:10.1016/j.cub.2015.04.019)
- 607 19. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S,  
608 Fall T, Seppälä EH, Hansen MST, Lawley CT, *et al.* 2011. Identification of genomic  
609 regions associated with phenotypic variation between dog breeds using selection mapping.  
610 *PLoS Genet.* **7**(10), e1002316. (doi:10.1371/journal.pgen.1002316)
- 611 20. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar  
612 P, de Bakker PIW, Daly MJ, *et al.* 2007. PLINK: a tool set for whole-genome association  
613 and population-based linkage analyses. *Am J Hum Genet.* **81**, 559–575.  
614 (doi:10.1086/519795)
- 615 21. Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in  
616 unrelated individuals. *Genome Res.* **19**, 1655–64. (doi: 10.1101/gr.094052.109)
- 617 22. Pritchard J, Stephens M, Donnelly P. 2000. Inference of population structure using  
618 multilocus genotype data. *Genetics* **155**, 945–59.
- 619 23. Guillot G, Santos F, Estoup A. 2008. Analysing georeferenced population genetics data  
620 with Geneland: a new algorithm to deal with null alleles and a friendly graphical user  
621 interface. *Bioinformatics* **24**, 1406–1407. (doi:10.1093/bioinformatics/btn136).
- 622 24. Meirmans PG. 2012. The trouble with isolation-by-distance. *Mol Ecol.* **21**, 2839–46.  
623 (doi:10.1111/j.1365-294X.2012.05578.x)

- 624 25. Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic  
625 software for teaching and research – an update. *Bioinformatics* **28**, 2537–9.  
626 (doi:10.1093/bioinformatics/bts460)
- 627 26. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. 2013. MEGA6: Molecular  
628 Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol.* **30**, 2725–2729.  
629 (doi:10.1093/molbev/mst197)
- 630 27. Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS*  
631 *Genet.* **2**(12), e190. (doi:10.1371/journal.pgen.0020190)
- 632 28. McVean G. 2009. A genealogical interpretation of Principal Components Analysis. *PLoS*  
633 *Genet.* **5**(10), e1000686. (doi:10.1371/journal.pgen.1000686)
- 634 29. Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-  
635 wide allele frequency data. *PLoS Genet.* **8**(11), e1002967.  
636 (doi:10.1371/journal.pgen.1002967)
- 637 30. Yu Y, Harris AJ, Blair C, He X-J. 2015. RASP (Reconstruct Ancestral State in  
638 Phylogenies): a tool for historical biogeography. *Mol Phylogenet Evol.* **87**, 46–9.  
639 (doi:10.1016/j.ympev.2015.03.008)
- 640 31. Brown SK, Darwent CM, Sacks BN. 2013. Ancient DNA evidence for genetic continuity  
641 in Arctic dogs. *J Archaeol Sci.* **40**, 1279–88. (doi:10.1016/j.jas.2012.09.010)
- 642 32. Brown SK, Darwent CM, Wictum EJ, Sacks BN. 2015. Using multiple markers to  
643 elucidate the ancient, historical and modern relationships among North American Arctic  
644 dog breeds. *Heredity* 24 June 2015. (doi:10.1038/hdy.2015.49)
- 645 33. van Asch B, Zhang A-B, Oskarsson MCR, Klütsch CFC, Amorim A, Savolainen P. 2013.  
646 Pre-Columbian origins of Native American dog breeds, with only limited replacement by  
647 European dogs, confirmed by mtDNA analysis. *Proc R Soc B London.* **280**, 2013.1142.  
648 (doi:10.1098/rspb.2013.1142)

- 649 34. Yu Y, Harris AJ, He X.-J. 2010. S-DIVA (Statistical Dispersal-Vicariance Analysis): A  
650 tool for inferring biogeographic histories. *Mol Phylogenet Evol.* **56**, 848–50.  
651 (doi:10.1016/j.ympev.2010.04.011)
- 652 35. Ronquist F. 2004. Bayesian inference of character evolution. *Trends Ecol Evol.* **19**, 475–  
653 81. (doi:10.1016/j.tree.2004.07.002)
- 654 36. Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, Silva PM,  
655 Galaverni M, Fan Z, Marx P, Lorente-Galdos B, *et al.* 2014. Genome sequencing  
656 highlights the dynamic early history of dogs. *PLoS Genet.* **10**, e1004016.  
657 (doi:10.1371/journal.pgen.1004016)
- 658 37. Pilot M, Greco C, vonHoldt BM, Jędrzejewska B, Randi E, Jędrzejewski W, Sidorovich  
659 VE, Ostrander EA, Wayne RK. 2014. Genome-wide signatures of population bottlenecks  
660 and diversifying selection in European wolves. *Heredity.* **112**, 428–42.  
661 (doi:10.1038/hdy.2013.122)
- 662 38. Oskarsson MCR, Klütsch CFC, Boonyaparakob U, Wilton A, Tanabe Y, Savolainen P.  
663 2012. Mitochondrial DNA data indicate an introduction through Mainland Southeast Asia  
664 for Australian dingoes and Polynesian domestic dogs. *Proc R Soc B London.* **279**, 967–  
665 974. (doi:10.1098/rspb.2011.1395)
- 666 39. Duleba K, Skonieczna K, Bogdanowicz W, Malyarchuk B, Grzybowski T. 2015.  
667 Complete mitochondrial genome database and standardized classification system for  
668 *Canis lupus familiaris*. *Forensic Sci Int Genet.* **19**, 123–9.  
669 (doi:10.1016/j.fsigen.2015.06.014)
- 670 40. Pickrell JK, Reich D. 2014. Toward a new history and geography of human genes  
671 informed by ancient DNA. *Trends Genet.* **30**, 377–89. (doi:10.1016/j.tig.2014.07.007)

- 672 41. McEvoy BP, Powell JE, Goddard ME, Visscher PM. 2011. Human population dispersal  
673 "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs.  
674 *Genome Res.* **21**, 821–829. (doi:10.1101/gr.119636.110)
- 675 42. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH,  
676 Schraiber JG, Castellano S, Lipson M, *et al.* 2014. Ancient human genomes suggest three  
677 ancestral populations for present-day Europeans. *Nature* **513**, 409–13.  
678 (doi:10.1038/nature13673)
- 679 43. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt  
680 S, Harney E, Stewardson K, *et al.* 2015. Migration from the steppe was a source for Indo-  
681 European languages in Europe. *Nature* **522**, 207–11. (doi:10.1038/nature14317)
- 682 44. Harpur BA, Minaei S, Kent CF, Zayed A. 2012. Management increases genetic diversity  
683 of honey bees via admixture. *Mol Ecol.* **21**, 4414–21. (doi: 10.1111/j.1365-  
684 294X.2012.05614.x)

685

686 **Figure legend**687 **FIG. 1 Genetic differentiation in FBDs and dog expansion routes in Eurasia. (A)**

688 Distribution of sampling sites with their division into geographic regions, and dog expansion  
689 routes in Eurasia inferred using RASP. **(B)** Maximum likelihood tree of genetic  
690 differentiation among FBDs and pure-breed dogs, constructed in TREEMIX. Distribution of  
691 ancestral populations was inferred using RASP (with uncertainty assessed using Bayesian  
692 Binary MCMC), and is marked on nodes using colour-codes. Black colour denotes  
693 undetermined distribution, and the colour codes are simplified compared with the original  
694 output (Supplementary Fig. 18A). In the RASP analysis, Arctic breeds were assigned to East  
695 Asia, according to their primary origin [31-33]. Bootstrap support (based on 1,000 replicates)  
696 is marked with black stars if above 90%, and with white stars if between 65 and 90%. **(C)**

697 Population genetic structure in Eurasian FBDs, inferred using ADMIXTURE assuming 2 to 4  
698 genetic clusters.

699

700 **Supplementary Material**

701 Supplementary material is provided as two PDF files. Supplementary File 1 includes

702 Supplementary Text, Supplementary Figures 1-20 and Supplementary Tables 1-5;

703 Supplementary File 2 presents Supplementary Fig. 12B.

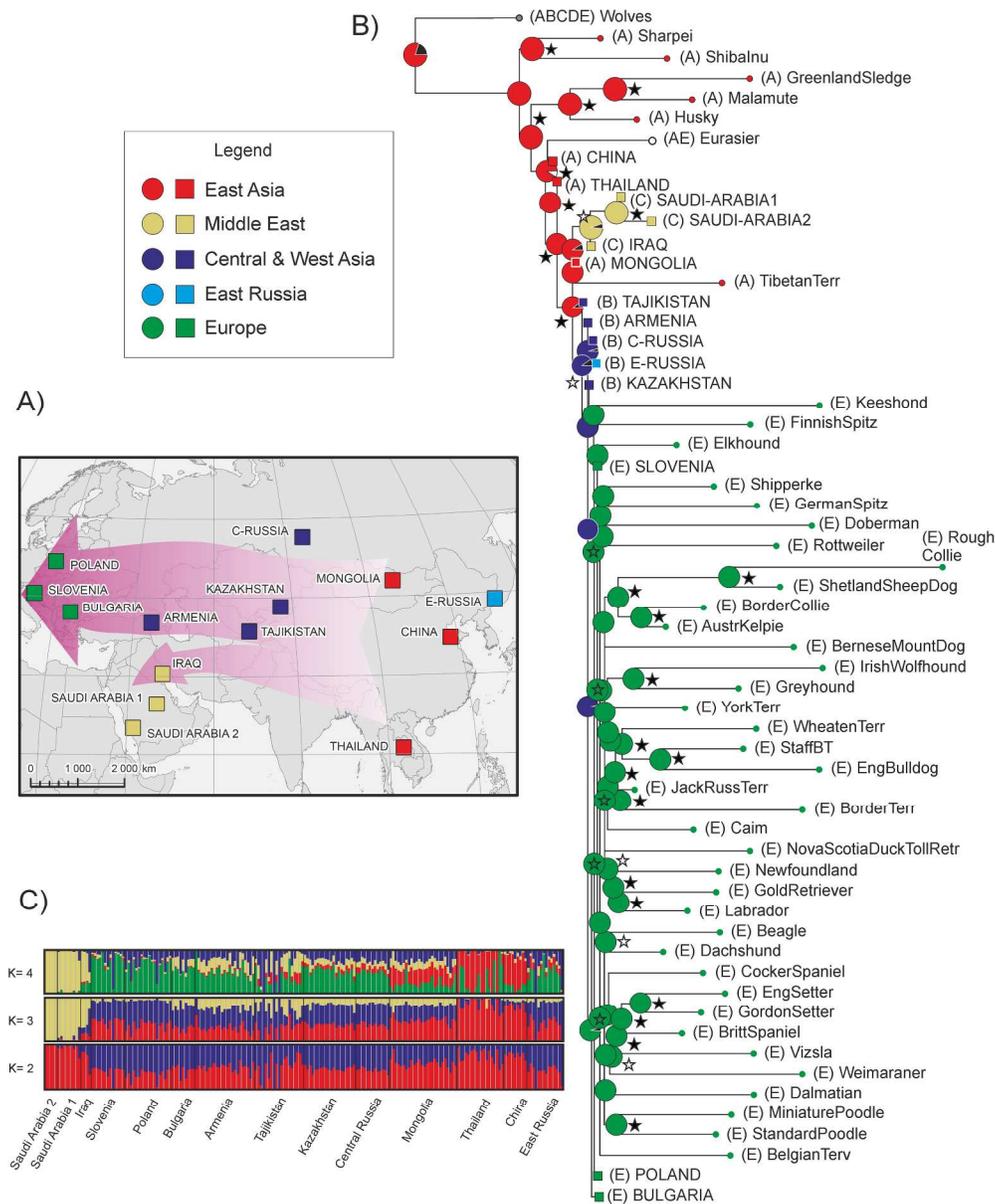


FIG. 1 Genetic differentiation in FBDs and dog expansion routes in Eurasia. (A) Distribution of sampling sites with their division into geographic regions, and dog expansion routes in Eurasia inferred using RASP. (B) Maximum likelihood tree of genetic differentiation among FBDs and pure-breed dogs, constructed in TREEMIX. Distribution of ancestral populations was inferred using RASP (with uncertainty assessed using Bayesian Binary MCMC), and is marked on nodes using colour-codes. Black colour denotes undetermined distribution, and the colour codes are simplified compared with the original output (Supplementary Fig. 18A). In the RASP analysis, Arctic breeds were assigned to East Asia, according to their primary origin [31-33]. Bootstrap support (based on 1,000 replicates) is marked with black stars if above 90%, and with white stars if between 65 and 90%. (C) Population genetic structure in Eurasian FBDs, inferred using ADMIXTURE assuming 2 to 4 genetic clusters.

239x289mm (300 x 300 DPI)