Supplementary information, Data S5 TreeMix results with our dataset and the F3/F4 test

We ran a TreeMix analysis that allowed a number of possible migratory tracks (0-3) using our 58 samples. When only one migration track was allowed, we saw a clear admixture for the northern Chinese indigenous dogs. The individual components are the southern East Asian components and dog breeds outside of Asia (Supplementary information, Figure S5). The second migration track is the migration from wolves to the African and Middle Eastern groups (i.e., African village dogs and Afghan Hound). Of note, the migration between wolves and Basenjis (also African origin) has also been found in a recent study[9]. Due to the limited number of samples that we have, the full migration picture between wolves and dogs remains to be further explored. The third migration track is the migration from southern Chinese indigenous dogs into the Arctic groups. This also matches with the observation in the structure/Frappe analysis where these groups show a mixture of components from the southern Chinese indigenous components and the out-of-Asia components (Figure 1D).

In order to further explore the landscape of admixture observed in the TreeMix analysis, we also employed the F3 test to test for admixture between candidate groups in our samples. In the F3 test (in the format of W: X,Y), one focal population (W) was tested as a mixture of population X and Y. A large negative value of f score (standardized to a Z score) would indicate a very strong signal of admixture of W as a mixture of X and Y. Using TreeMix's threepop module, we tested all combinations of W/X/Y and found these combinations with significant Z scores (≤ -2.0, calculated using the Jackknife procedure). In Supplementary information, Table S2, we can see that, the most significant combinations are combinations involving northern Chinese indigenous dog groups. This suggests that there are very strong signal of admixture in northern Chinese indigenous dogs.

It is worth pointing out that, the statistical power of the F3 test depends on whether the population, after admixture, had experienced a significant amount of genetic drift (e.g. bottleneck). The F3 test lacks the power to detect admixture if there have been a large amount of genetic drifts after the admixture event[6]. This might be the reason why many admixture events are not detectable in the tree using the F3 test. Northern

Chinese indigenous dogs did not experience a substantial bottleneck since the admixture and still maintain a large amount of genetic variations.

Other than the F3 test, a more sensitive method is the F4 test[6]. In the F4 test setting (e.g., (A,B), (C,D)), if the four populations are related by a tree topology, one of the three combinations: ((A,B), (C,D)), or ((A,C), (B,D)), or ((A,D),(B,C)), will give a close to zero f score f((A,B),(C,D)). In contrast, if there are admixture events in the history of these populations, all three combinations will show strong non-zero f values (or the standardized version, which is a Z score).

In our analysis, we ran all possible F4 test combinations with all our 23 populations (wolves, 12 individuals; southern Chinese indigenous dogs, 11 individuals; northern Chinese indigenous dogs,12 individuals; African village dogs, 4 individuals and 19 breeds, one individual each), thus there are 8,855*3 = 26,565 combinations. For each population, there are 1,540*3=4,620 combinations involving this population.

After extracting all possible four-population f test scores, we then combined the three f scores for the same group, (for each combination of four populations A,B,C,D, here three combinations: ((A,B), (C,D)) ,((A,C),(B,D)) and ((A,D), (B,C))), into a single value by taking the score with the smallest absolute value. This will give us 8,855 combinations and each single population is involved in 1,540 combinations.

Out of 1,540 possible f scores for the northern Chinese indigenous dogs, 579 are significant (Z score ≥ 3 or ≤ -3). This is much higher than for the other population groups (Supplementary information, Table S3). It matches our previous results with the F3 test and suggests that there is very strong statistical support for admixture in the northern Chinese indigenous dogs.

Since all combinations involving the admixed populations would show significant f scores, we thus removed all combinations involving the admixture populations and inspected the landscape “subtracting” the previously detected signals. We ran an iterative procedure where the top candidate of admixture was removed each time. We next removed all combinations involving the northern Chinese indigenous dogs, and the strongest signal left in the data is the African village dogs followed by the Afghan

Hound (AFG) (Supplementary information, Table S3). This matches with the findings in the TreeMix analysis (e.g., when the number of migration tracks is two, Supplementary information, Figure S5). Interestingly, we were able to detect admixture also in the other North African breed Sloughi (SLO) (discussed in a later section).

Following the admixture signals detected in the northern Chinese indigenous dogs, the African village dogs, and the Middle Eastern/North African breeds; the next top signals are the Arctic groups. We were able to detect admixture in the Arctic clade (ALM, GRD, SIH), which matches with our observation in the TreeMix analysis. More interestingly, we also found potential admixture signal in East Siberian Laika and Samoyed (Supplementary information, Table S3).

It is worth pointing out that in the F4 test, when an individual population is a source population (i.e. donor) for many admixed populations, it will also appear in many combinations. For example, the southern Chinese indigenous dogs and wolves also showed high number of combinations with high f scores. This is because they are the source for multiple populations. Removing them will remove many true signal of admixture.

Interestingly, even though indigenous dogs from Oceania, i.e., dingoes and New Guinea singing dogs, bear a signal of introgression from gray wolves, we found no evidence of wolf admixture in group 1 breeds and their closeness to wolves is not due to potential admixture with gray wolves.

Despite the fairly good concordance between the TreeMix and F4 test results, some subtle differences between them exist. The reason for this is quite complex and has not been adequately explored. We discuss two possibilities here. First, the TreeMix method is an iterative maximum likelihood procedure, where the likelihood function can greatly improve if migration tracks are allowed in “large” true admixed clades. For example, the migration track to the African/Middle Eastern clade and the migration track to Arctic dog groups (SIH, GRD and ALM) will tend to improve the likelihood function significantly. Thus, there might be a general tendency for the TreeMix method to detect admixture in large clades. Secondly, we have chosen a greedy algorithm in teasing apart the true signal in the F4 test. It remains a challenge to aggregate the piece-wise signal in the individual F4 test and obtain for a consensus that respects all the constraints. Nevertheless, we observe fairly good concordance between the F3/F4 test and TreeMix results.