




ORIGINAL RESEARCH

Major inconsistencies of inferred population genetic structure estimated in a large set of domestic horse breeds using microsatellites

Stephan Michael Funk^{1,2}  | Sonya Guedaoura^{3,4} | Rytis Juras⁵ | Absul Raziq⁶ | Faouzi Landolsi⁷ | Cristina Luís⁸  | Amparo Martínez Martínez⁹ | Abubakar Musa Mayaki¹⁰ | Fernando Mujica¹¹ | Maria do Mar Oom¹² | Lahoussine Ouragh¹³ | Yves-Marie Stranger¹⁴ | Jose Luis Vega-Pla¹⁵ | Ernest Gus Cothran⁵ 

¹Centro de Excelencia de Modelación y Computación Científica, Universidad de La Frontera, Temuco, Chile

²Nature Heritage, St. Lawrence, UK

³Faculté des Sciences de la Nature et de la Vie, Université d'El-Tarf, El-Tarf, Algeria

⁴Faculté de Pharmacie, Université Laval, Québec City, QC, Canada

⁵College of Veterinary Medicine and Biomedical Science, Texas A&M University, College Station, TX, USA

⁶Society of Veterinary, Environment and Agriculture Scientists (SAVES), Quetta, Pakistan

⁷Ecole Nationale de Médecine Vétérinaire, Sidi Thabet, Tunisie

⁸Centro Interuniversitário de História das Ciências e da Tecnologia (CIUHCT), Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

⁹Departamento de Genética, Universidad de Córdoba, Córdoba, Spain

¹⁰Department of Veterinary Medicine, Usmanu Danfodiyo University, Sokoto, Nigeria

¹¹Instituto de Producción Animal, Universidad Austral de Chile, Valdivia, Chile

¹²CE3C – Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

¹³Institut Agronomique et Vétérinaire Hassan II, Rabat, Morocco

¹⁴, Marmande, France

¹⁵Laboratorio de Investigación Aplicada, Cria Caballar de las Fuerzas Armadas, Cordoba, Spain

Correspondence

Stephan Michael Funk, Nature Heritage, St. Lawrence, Jersey, UK.
Email: smf@natureheritage.org

Ernest Gus Cothran, College of Veterinary Medicine and Biomedical Science, Texas A&M University, College Station, TX 77843 USA.
Email: gcothran@cvm.tamu.edu

Abstract

STRUCTURE remains the most applied software aimed at recovering the true, but unknown, population structure from microsatellite or other genetic markers. About 30% of STRUCTURE-based studies could not be reproduced (*Molecular Ecology*, 21, 2012, 4925). Here we use a large set of data from 2,323 horses from 93 domestic breeds plus the Przewalski horse, typed at 15 microsatellites, to evaluate how program settings impact the estimation of the optimal number of population clusters K_{opt} that best describe the observed data. Domestic horses are suited as a test case as there is extensive background knowledge on the history of many breeds and extensive phylogenetic analyses. Different methods based on different genetic assumptions and statistical procedures (DAPC, FLOCK, PCoA, and STRUCTURE with different run scenarios) all revealed general, broad-scale breed relationships that

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

largely reflect known breed histories but diverged how they characterized small-scale patterns. STRUCTURE failed to consistently identify K_{opt} using the most widespread approach, the ΔK method, despite very large numbers of MCMC iterations (3,000,000) and replicates (100). The interpretation of breed structure over increasing numbers of K , without assuming a K_{opt} , was consistent with known breed histories. The over-reliance on K_{opt} should be replaced by a qualitative description of clustering over increasing K , which is scientifically more honest and has the advantage of being much faster and less computer intensive as lower numbers of MCMC iterations and repetitions suffice for stable results. Very large data sets are highly challenging for cluster analyses, especially when populations with complex genetic histories are investigated.

KEYWORDS

domestic horse, population genetic structure, Przewalski horse, STRUCTURE analysis

1 | INTRODUCTION

Molecular ecology and conservation biology heavily rely on the identification of population structure and genetic admixture between individuals and populations. Model-based inference of these parameters has become the central methodological approach especially for microsatellite markers (Pritchard, Stephens, & Donnelly, 2000). Bayesian statistics utilizing Markov Chain Monte Carlo, MCMC, simulations has been implemented in various computer programs such as BAPS (Corander, Marttinen, Sirén, & Tang, 2008) and STRUCTURE (Pritchard et al., 2000). The latter is the by far most applied software for microsatellites (188 and 17,997 citations for BAPS and STRUCTURE, respectively, in WEB OF SCIENCE by March 2019), acts as a benchmark, and remains popular even after the advent of alternative markers such as SNPs (1,188 STRUCTURE citations for 2018 compared with peak citations of 1,433 for 2014). STRUCTURE performs relatively well under various population models (Gilbert et al., 2012; Puechmaile, 2016; Putman & Carbone, 2014), but the reanalysis of STRUCTURE-based studies failed in ~30% of cases to reproduce results (Gilbert et al., 2012). Underpinning factors include the inherent stochastic nature of the model-fitting procedure, non-convergence of parameter estimates due to inappropriate number of MCMC iterations, inappropriate number of replicate runs R , weak population structure, too few informative microsatellite loci, and the estimation procedure per se (Gilbert et al., 2012; Putman & Carbone, 2014). The reanalysis used 34 studies with relatively small data sets (median = 254 individuals) including only one case with > 1,000 individuals. Three-quarter of studies used less than the recommended $R = 20$ (Evanno, Regnaut, & Goudet, 2005; Gilbert et al., 2012). The review concluded that, in general, additional replicates are highly important to achieve higher precision of the parameter estimates (Gilbert et al., 2012). Replicates are essential to achieve higher precision of the parameter estimates including their variance, which are the basis for the ad hoc statistic ΔK , the dominant statistic to identify

the optimal number of clusters that best explain the data (Evanno et al., 2005). In conjunction with the observation that STRUCTURE performs best with a small number of discrete populations (Pritchard et al., 2000), the nonrepeatability might point to as yet not evaluated performance issues of STRUCTURE and ΔK specific for large data sets.

Microsatellite data sets with large numbers of samples and animal populations have become increasingly available, especially in relatively easy to collect organisms such as near-sedentary wild animals, for example, more than 2,500 colonies of the Broadcast spawning coral *Acropora tenuis* (Lukoschek, Riginos, & van Oppen, 2016), and domestic animals, for example, 1,514 dogs from 61 breeds (Leroy, Verrier, Meriaux, & Rognon, 2009), 1,826 sheep from 49 breeds (Leroy et al., 2015), 1,924 cattle from 40 breeds (Martin-Burriel et al., 2011), 3,333 cattle from 81 breeds (Martínez et al., 2012), and 1,547 and 2,385 horses from 25 and 50 breeds, respectively (Cortés et al., 2017; van de Goor, van Haeringen, & Lenstra, 2011). While one study did not specify the number of repeats (Cortés et al., 2017), only low numbers of replicates ranging from $R = 4$ to $R = 10$ were employed except of $R = 50$ by Leroy et al. (2009, 2015). Moreover, the choices of R and MCMC iterations are rarely justified in the literature. STRUCTURE works best with relatively small numbers of demes or populations (Pritchard et al., 2000) but the performance with large data sets remains unevaluated. It is noteworthy that the largest study included in Gilbert et al. (2012) with 1,361 Anacapa deer mice (Ozer, Gellerman, & Ashley, 2011b) was nonrepeatably despite employing the recommended number of replicates. An evaluation of the performance of STRUCTURE and ΔK for large data sets is thus important and timely.

Domestic horses (*Equus caballus*) lend themselves as an empiric test case to evaluate these questions not only because of our large sample size but also detailed information on natural history and breeding management of modern breeds is available. Horses were relatively late domesticated at ≈ 5.5 KYA, but then rapidly gained

wide distribution due to their versatility (Levine, 2005; Librado et al., 2016; Warmuth et al., 2012). Widespread gene flow during and after domestication recurred from ancestral lineages including the Przewalski horse (*Equus przewalskii*), the wild horse *Equus ferus*, and an extinct wild, taxonomically undescribed horse population into domestics and vice versa (Der Sarkissian et al., 2015; Schubert et al., 2014; Warmuth et al., 2012). A strong sex bias to the overall gene pool not only characterizes domestication but also modern breed management where it is exaggerated by the use of a small, selected number of stallions (Wallner et al., 2017). The formation of systematic modern horse breeding is marked by the establishment of the earliest studbook, that of the Lipizzaner in 1580 and continues till today with newly emerging formal breeds (Galov et al., 2013). The large variety of breed management is exemplified by the co-existence of closed breeds without admixture from outside breeds, open breeds, landraces, and feral populations with little artificial selection but sometimes strong natural selection such as the adaptation to high altitudes (Hendrickson, 2013), and breeds with strong artificial selection for desired traits. Two large microsatellite studies investigated 67 and 41 breeds, respectively, and constructed distance-based phylogenies using a variety of phylogenetic methods (Conant, Juras, & Cothran, 2012; Pires et al., 2016). In general, these approaches revealed patterns largely reflecting known breed histories, but statistical support was consistently low (Cothran & Luís, 2005). Whether these low bootstrap values are a consequence of complex breed history and admixture or caused by the phylogenetic methods per se (Pardi & Scornavacca, 2015; Pritchard et al., 2000) remains unknown.

Using large-scale screening of domestic horse breeds alongside the Przewalski horse, we aimed to evaluate the robustness of STRUCTURE and ΔK with emphasis on the effects of the numbers of MCMC iterations and replicates. Results are compared with alternative clustering methods based on different model assumptions. STRUCTURE assumes linkage equilibrium, LE, and Hardy-Weinberg equilibrium, HWE (Pritchard et al., 2000). In contrast, the multilocus maximum likelihood approach (Paetkau, Calvert, Stirling, & Strobeck, 1995) as implemented in FLOCK (Pierre Duchesne & Turgeon, 2012) iteratively allocates genotypes into genetic clusters without assuming HWE and LE. Discriminant analysis of principal components, DAPC, as implemented in the software ADEGENET (Jombart, 2008; Jombart, Devillard, & Balloux, 2010) is a model-free multivariate method.

2 | MATERIALS AND METHODS

2.1 | Samples and genotyping

A total of 4,392 individuals from 94 horse populations (Table S1) were collected during long-term studies on horse genetics (Cothran, Canelon, Luís, Conant, & Juras, 2011; Cothran & Luís, 2005; Khanshour, Conant, Juras, & Cothran, 2013; Khanshour, Juras, Blackburn, & Cothran, 2015; Khanshour, Juras, & Cothran, 2013;

Luís, Cothran, & Oom, 2007; Pires et al., 2014). The sample includes the Przewalski horse and 89 internationally recognized breeds of which Arabian horse, Barb horse, and Spanish Pure Breed stem from five, three, and two different management schemes and studbooks, respectively. North American bred Spanish Pure Bred is commonly called Andalusian. Jointly, they represent 93 domestic breeds sensu lato plus the Przewalski, henceforth all denoted as “breed” for simplicity. The geographic origin of breeds was focussed on South and Central America ($n = 13$), the Iberian Peninsula ($n = 14$), the British Isles ($n = 16$), and Central and Eastern Europe ($n = 18$) with additional, representative breeds from North America ($n = 5$), Southeastern Europe ($n = 7$), Africa ($n = 10$), and Asia ($n = 10$). Horse breeds include highly divergent breeding strategies and histories including cold-blooded, Celtic and Iberian horses in Europe, and Arab and non-Arab horses in Africa (Table S1). Breeds were labeled in the figures according to their geographic origin rather than their breeding history as a neutral approach to visual interpretation of the graphs. Except for four breeds with less than 25 individuals each, we reduced sample sizes to 25 individuals per breed in order to account for the impact of uneven sampling that tends to bias the evaluation of population genetic structure (Puechmaile, 2016). In horses, sample sizes of 25 per breed have been demonstrated suitable for evaluating population genetic structure (Cothran & Luís, 2005). Individuals were excluded in order to remove horses from the same group or farm, which may be distantly related, although great care was taken not to sample relatives. In total, 2,323 horses were included in the analysis (Table S1).

We used 15 autosomal microsatellite markers, distributed on 14 chromosomes, from marker panels that are recommended for diversity studies by ISAG-FAO and the International Society for Animal Genetics (Cothran & Luís, 2005; FAO, 2011). Details on genotyping can be found in Juras, Cothran, and Klimas (2003) and Khanshour, Conant, et al. (2013).

2.2 | Information content of the genetic data

The combination of sample sizes, numbers of microsatellite loci, numbers and frequencies of observed alleles, and the degree of differentiation between individuals and populations, typically quantified by F_{IS} and F_{ST} , establishes the information content and impacts the statistical power to detect genetic structure and differentiation (Ryman et al., 2006). We evaluated it with a simulation approach implemented in POWSIM with observed allele frequencies, 1,000 replicates, 10,000 burn-ins, and 100 batches with 10,000 iterations (Ryman, 2011; Ryman & Palm, 2006). Because the published version is limited to 30 populations, we also used a version adjusted to our data set permitting 90 breeds (Ryman, M. Stephan, S. M. F. Funk, personal communication). To generate expected F_{ST} values, a generation time of 5 years since domestication $\approx 5,500$ KYA was used. Considering that generation time may vary largely and that five to 12 years has been cited (Lorenzen et al., 2011; Sokolov & Orlov, 1986), our choice of five years is conservative and allows drift over 1,100 generations. We selected a range for effective population

size, N_e , that reflects the evolution of modern breeds by varying N_e between 10 and 1,000, resulting in expected F_{ST} values, estimated by POWSIM, between 0.001 and 0.87. This range of simulated F_{ST} values included all observed values of investigated horses. Power was estimated using Fisher's exact and chi-square tests at 95% confidence interval (Ryman & Palm, 2006). Power at $t = 0$ ($F_{ST} = 0$) represents the type I error, the frequency that genetic homogeneity is rejected while it is true. The information content was also evaluated using the clustering program FLOCK (Duchesne, Méthot, & Turgeon, 2013; Duchesne & Turgeon, 2012) which designates an "undecided stopping condition" when an optimal number of population clusters explaining the data cannot be identified. Two underpinning scenarios are possible: lack of "true" genetic structure or insufficient genetic information content. Discriminating these scenarios is part of the interpretation of the FLOCK results described further below in the context of population clustering.

2.3 | Hardy-Weinberg equilibrium and F-statistics and PCoA analysis

GENEPOP 4.5.1 (Rousset, 2008) was used for exact HWE tests for all locus/population combinations employing Markov Chain permutations (Guo & Thompson, 1992) with 10,000 de-memorizations and 400 batches with 10,000 iterations each. Fisher's combined probability test was used to calculate P across loci and across populations, respectively. For the global analysis, P was calculated with and without adjusting for multiple testing by sequential Bonferroni correction (Holm, 1979). F_{IS} was calculated according Weir and Cockerham (1984). Genetic differentiation was evaluated by pairwise weighted mean F_{ST} distances between breeds (Cockerham & Weir, 1993) and the log-likelihood ratio, G , test in FSTAT 2.9.3 (Goudet, 1995) with 10,000 randomizations of genotypes without assuming random mating. Genic differentiation was evaluated by exact G and Fisher's combined probability tests in GENEPOP 4.5.1 (Rousset, 2008). The F_{ST} matrix was summarized by principle coordinate analysis, PCoA, with covariance standardization in GENALEX 6.502 (Peakall & Smouse, 2012). Linkage equilibrium, LE, was not evaluated as it was not rejected for the same loci in several previous studies (e.g., Khanshour et al., 2015).

2.4 | STRUCTURE analysis

STRUCTURE 2.4.4 (Pritchard et al., 2000; Pritchard, Wen, & Falush, 2012) was applied allowing for correlated allele frequencies and admixture permitting mixed ancestry and accounting for recent between-breed gene flow. The locprior model with breed assignment as prior information, which is more sensitive for detecting weak population structure than the nonlocprior model, was used alongside the latter as recommended to check for consistency (Hubisz, Falush, Stephens, & Pritchard, 2009). We denote the models as +LP and -LP henceforth. Gilbert et al. (2012) recommend at least MCMC

100,000 iterations and, ideally, larger runs lengths, but these are often limited by run-time considerations. Three sets of MCMC iterations were run for each of the \pm LP models: 150,000, 750,000, and 1,500,000 of which the first 50,000, 250,000, 500,000, and 1,000,000 iterations, respectively, served as burn-ins. Because of computing time restrictions, the scenario with 3,000,000 MCMCs, including 1,000,000 burn-ins, was restricted to the +LP model. A priori defined clusters, K , were considered from $K = 2$ to $K = 30$ in steps of one and from $K = 35$ to $K = 90$ in steps of five. For each K and each MCMC/LP scenario, one of the replicates was randomly selected for visual inspection of the Dirichlet parameter α , which estimates the degree of admixture, and the log-likelihoods for convergence (Gilbert et al., 2012; Pritchard, Wen, & Falush, 2010). We restricted ourselves to the visual inspection as STRUCTURE has not implemented any formal assessment of convergence nor have any respective statistics such as Gelman and Rubin's statistics (1992) been implemented and tested for STRUCTURE (e.g., Zachos et al., 2016). STRUCTURE reports the posterior probability of the data for a given K for each replicate, denoted as $\Pr(X|K)$ in Bayesian notion and abbreviated as $P(K)$ henceforth. For all MCMC/ \pm LP scenarios, $R = 40$ replicates were used except for $R = 100$ for the 1,500,000 MCMC/-LP scenario. Computations used the SGI UV-2000 INTEL(R) XEON(R) E5-4640 @ 2.40GHz (SANDYBRIDGE) computing platform with 96 nodes, 192 cores, and 1.5 teraflop.

The most likely number of clusters that best describe the given data was first estimated by the visual inspection of the means and standard deviations of $P(K)$ over all investigated K according to Pritchard et al. (2000), Pritchard et al. (2010), henceforth denoted $K_{\text{opt[Pritchard]}}$. Second, STRUCTURE HARVESTER v. 0.6 (Earl & von Holdt, 2012) was used to calculate the optimal K according the Evanno ΔK statistics, henceforth denoted $K_{\text{opt[Evanno]}}$. We evaluated the convergence of the two estimates over r by first partitioning all R replicates into B subsequent blocks of 5 replicates each. $P(K)$, $K_{\text{opt[Evanno]}}$ and $K_{\text{opt[Pritchard]}}$ were then calculated for both, each block separately and accumulated over all preceding blocks.

In order to evaluate the impact of R in smaller data sets on the optimal K selection, a subset of 295 individuals from five Arab breeds, Algerian, Moroccan and Tunisian Barb, Akhal Teke, Caspian, Turkoman and the Przewalski horse was analyzed with $R = 40$. A similar selection of breeds from the same data base was used previously with $R = 10$ and 120,000 MCMCs and 200,000 MCMCs, respectively (A. Khanshour, Conant, et al., 2013). Alongside, data on 1,361 Anacapa deer mice (Ozer, Gellerman, & Ashley, 2011a; Ozer, Gellerman, & Ashley, 2011b) were also re-evaluated with the originally applied STRUCTURE parameters, but with $R = 100$.

Whether all $P(K)$ values resulted in similar or multimodal solutions was evaluated on the CLUMPAK 1.1 webserver (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015, <http://clumpak.tau.ac.il>). All MCMC scenarios were combined for each LP model. Settings utilized the LARGEKGREEDY algorithm (Jakobsson & Rosenberg, 2007) and dynamic threshold determination (Kopelman et al., 2015). We then re-calculated ΔK and $K_{\text{opt[Evanno]}}$ using only the dominant solutions, that is, the "major modes" (Kopelman et al., 2015), and

excluded “minor modes,” rare different solutions when present. The visual presentation of the cluster's estimated membership coefficients for individuals and breeds utilized DISTRUCT 1.1 (Rosenberg, 2003) with replicates from the majority modes.

We identified “ghost clusters” which are spurious clusters that the MCMC search strategy produces. They occur when the a priori K is larger than the optimum with STRUCTURE enforcing suboptimal, not empty clusters. Following Guillot, Mortier, and Estoup (2005), Puechmaille (2016) defines ghost clusters as those with mean membership coefficients <0.5 in any studied population and then subtracts the number of ghost clusters from the ΔK estimate.

2.5 | FLOCK analysis

FLOCK 3.1 (Duchesne et al., 2013; Duchesne & Turgeon, 2012) was applied using 20 re-allocations for the multilocus maximum likelihood procedure with 50 runs for each K and a LLOD threshold score of 0. The optimal number of clusters, $K_{\text{opt[FLOCK]}}$, was determined by ad hoc “stopping” rules (Pierre Duchesne & Turgeon, 2012). Log-likelihood scores, $\text{LLOD}(K)$, averaged over genotypes, and replicates are sorted according their absolute values over all runs. “Plateau lengths” are the counts of scores with the same values and are the core of the “stopping rules.” Hierarchical analysis was applied for each of the identified breed clusters until each cluster could not be further subdivided. The total number of undividable clusters identifies $K_{\text{opt[FLOCK]}}$.

In cases where FLOCK produced an “undecided stopping condition,” the underpinning cause was evaluated in three steps, which are not described in the original publication or the FLOCK manual (Duchesne et al., 2013; Duchesne & Turgeon, 2012). These steps aim to distinguish between insufficiently variable microsatellite data and the level of genetic structure as possible causes. The latter case is produced when relatively many highly admixed individuals and/or breeds are present which FLOCK cannot consistently assign to single clusters because of stochasticity (Duchesne M. Stephan, S. M. F. Funk, personal communication). First, the “Samples likelihood maps” in FLOCK's output, that is, the averaged likelihoods $\text{LLOD}(K)$ in each breed, were visually inspected for clear outlier breeds showing values of a different magnitude than all other breeds. Because of differential effects of stochasticity, $\text{LLOD}(K)$ values among nonoutliers may be deflated when outliers are present, leading to an “undecided stopping condition” (Duchesne, M. Stephan, S. M. F. Funk, personal communication). Outliers and nonoutliers were separated, and then, both sets were separately re-analyzed. Second, when no outliers were identifiable for an “undecided stopping condition,” the “Sample allocation matrix” in FLOCK's output, that is, the allocation of samples to clusters, was evaluated. Breeds with less than 80% of horses allocated to a single cluster were removed from the hierarchical analysis as “admixed.” After reanalysis of the remaining breeds, two outcomes are possible: If an “undecided stopping condition” emerged again, then the genetic information content of the microsatellite markers is insufficient to reveal genetically differentiated clusters (Duchesne, M. Stephan, S. M. F. Funk, personal communication). If a

stopping condition was reached, the previous failure to detect structure indicates that a relatively high number of admixed individuals and/or breeds in the given data set prevented the identification of clusters. Third, the “admixed” breeds identified in the previous step were added stepwise to the reanalysis in order to evaluate whether single breeds or a combination of breeds resulted in the “undecided stopping conditions.”

2.6 | DAPC analysis

The DAPC analysis utilized the ADEGENET 2.0.1 module for the R environment (Jombart et al., 2010; Jombart et al., 2016). We used R 3.3.1 (R Foundation for Statistical Computing, 2016). The first step is to transform all genotypes into noncorrelated variables using a principle component analysis, PCA. For downstream analysis, all principle components, PCs, were retained. After PCA, discriminant analysis, DA, is applied. It partitions the variance within and between a priori defined clusters, K , such that the separation between clusters is maximized. We evaluated K for 2 to 94, the total number of breeds. To prevent overfitting by DAPC, 50 PCs, approximately one-third of the total number of PCs identified by the PCA model was used (Jombart & Collins, 2016). The optimal K , denoted $K_{\text{opt[DAPC]}}$ henceforth, was chosen by visually and statistically identifying sharp changes in the Bayesian information criterion, BIC, over increasing K (Jombart et al., 2016). BIC quantifies the fit of the DAPC model at each K . ADEGENET's option to statistically identify “sharp” changes by Ward's clustering method (Mojena, 2006) was used. DAPC results were visualized as scatter plots in ADEGENET. Individual posterior memberships for cluster assignments were averaged for breeds and plotted analogous to STRUCTURE using DISTRUCT 1.1 (Rosenberg, 2003).

3 | RESULTS

3.1 | HWE and F_{IS}

A total of 23 from 1,410 locus/breed combinations revealed $p = 0$ for the exact HWE tests whereby three breeds had more than one locus involved (Timor Pony, Gotland, and Friesian). Fifteen breeds exhibited Fisher's $p < .0001$ over all loci (Table S1). Excluding the latter, two loci produced $p < .001$ (HTG10 and HTG7), but both their Bonferroni corrected p' values were > 0.05 . Mean $\pm SD F_{IS}$ was 0.03 ± 0.17 over all breeds with a breed F_{IS} range of 0.2 to -0.11 . The only breed with a mean $\pm SD$ range that did not include zero was the Timor Pony, which also exhibited the largest number of loci with $p < .0001$ for the HWE tests.

3.2 | Breed differentiation

The global breed differentiation was highly significant ($p < .0001$, log-likelihood G test). Estimated pairwise F_{ST} values between breeds ranged

from 0.04 (Tushuri Cxeni—Pindos Pony) to 0.45 (Friesian—Abaco Horse) with a mean $F_{ST} \pm SD$ of 0.115 ± 0.064 among the 93 domestic breeds and 0.212 ± 0.037 for Przewalski versus the domestics (Table S2). Three domestics showed mean $F_{ST} > 0.2$ versus all other domestics (Abaco, 0.310 ± 0.041 ; Friesian, 0.259 ± 0.043 ; Sorraia, 0.235 ± 0.045 ; all highly inbred breeds). The 4,370 pairwise comparisons between breeds for genic differentiation revealed Fisher's $p < .000001$ in almost all comparisons except for 12 pairs with $.007 > p > .000001$. Eleven of the latter involved Arab horses bred in Chile and four African (Ethiopian, Moroccan Barb, Nigerian, and Tunisian Barb), two Asian (Pakistani and Kurdish), and four European (Pindos Pony, Tushuri Cxeni, Hanoverian, and Wielkopolski). Within these breed groups, there is shared common ancestry, for example, Hanoverian and Wielkopolski are both warm-blood which have a Thoroughbred cross in their background.

Figure 1 displays the PCoA matrix of pairwise F_{ST} values between breeds for the first two axes explained prearranged by geographic origin. The first two axes explained 18.3% of the total variation and the third axis added 5.4%. Visual inspection of the 1st-versus-2nd and 1st-versus-3rd component plots did not unequivocally separate clusters but there are clear centers of clustering according geographic origin and horse type. No clear overall geographic distribution of scores was seen. The scores on Axis 1 show fairly good correspondence to the breed groups that the breeds best fit rather than geography in most cases. The breeds furthest to the right are primarily the "cold-blood" breeds composed of draft horses and ponies. The center part of Axis 1 is mainly composed by Iberian breeds or Iberian origin including those from the Americas while those on the left side are the Arabian breeds and also breeds with a strong Thoroughbred background such as the Wielkopolski, Hanoverian, Selle Français, and Trakehner. Separated from the others is also an ark formed by African and Asian horses. Axis two appears to best separate the Asian breeds with distribution of scores near the center. The Akhal Teke and Caspian are outliers of this group because of low diversity due to recent bottlenecks. The Abaco horse is the most pronounced outlier reflecting its breed history with a strong recent population bottleneck and a very low population size. The broad pattern seen in Figure 1 reflects well the relations among most of the breeds as known from breed histories.

3.3 | Statistical power and F_{ST}

POWSIM estimated that there was power of 100% for all test scenarios to reject genetic uniformity at $F_{ST} \geq 0.0025$, which is lower than the smallest observed pairwise $F_{ST} = 0.004$ between breeds (Figure S1). The probability to falsely reject uniformity was low ($p < .08$) for all test scenarios, which is higher than 5% but normal for microsatellite markers (Ryman et al., 2006).

3.4 | STRUCTURE analysis

A limiting factor for the choice of R and MCMCs was computing time, which was exceptionally large for the large data set. One replicate of

1,500,000 MCMCs for all K took ≈ 17.5 days for $-LP$ and ≈ 49 days for $+LP$, respectively. Total computing time for all scenarios and R was ≈ 18.5 years, which was only achievable using up to 80 computing cores simultaneously.

Mean $P(K)$ values reached a plateau at around $K = 45$ of each of the MCMC/LP scenarios, indicating an optimal cluster number of $K_{opt[Pritchard]} \approx 45$ (Figure 2). The visual inspection of summary statistics of MCMC runs indicated convergence of the model parameters. For all MCMC/ $\pm LP$ scenarios, plots of mean $\pm SD$ were very similar with three exceptions. First, large variances occurred only for $-LP$ at around $K = 10$ to $K = 15$, indicating that MCMC model failed relatively often to find similar solutions in this region of K values. Second, at $K > 40$, the short MCMC/ $-LP$ scenario produced consistently lower (worse) $P(K)$ values than the three scenarios with more MCMC iterations, whereas the 1,500,000 MCMC/ $+LP$ model consistently produced the highest (best) values. In other words, once the $P(K)$ plateau is reached, the smallest MCMC regime performed worst. Third, the $+LP$ scenario produced consistently higher $P(K)$ values and therefore performed best compared to the $-LP$ scenario. Pearson's pairwise correlations of mean $P(K)$ up to $K = 30$ were all very high ($r > .99$) but low for variances (mean $r \pm SD = 0.29 \pm 0.27$ for six $-LP$ pairs and mean $r = .30, .16$ and $-.04$ for the three $+LP$ pairs, respectively).

The Evanno method did not converge to single maximum ΔK value for any of the investigated scenarios. The consistency of the mean and standard deviations of $P(K)_{MCMC, \pm LP}$ between replicate blocks, b , is shown in Figure 3a. Mean $P(K)_{b, MCMC, LP}$ values are highly correlated between replicate blocks b for all MCMC/LP scenarios (Pearson's $r = .99$ across all 49 pairwise comparisons, $\pm SD = 0.02$). Thus, a single block of 5 replicates is sufficient to consistently estimate mean $P(K)$. In contrast, standard deviations varied largely and the mean $\pm SD$ Pearson's correlation coefficient r across all 49 pairwise comparisons was $r = .14 \pm .24$, demonstrating large impacts of stochasticity on mean $P(K)$ over R (Figure S2). For all scenarios, $K_{opt[Evanno]}$ varied over accumulated blocks b .

The 3,000,000 MCMC/ $-LP$ scenario appeared an exception with an unchanging $K_{opt[Evanno]} = 11$ (Figure 3b) but varied greatly when the sequence of b was randomly iterated. In other words, even with double the number of the currently recommended $R = 20$, the Evanno method did not result in converged $K_{opt[Evanno]}$ estimates. Because there is no difference of membership in major modes of the 1,500,000-MCMC/ $-LP$ and the 3,000,000-MCMC/ $-LP$ scenarios (see below), we used 500,000 MCMCs to evaluate convergence for $R = 100$ (Figure 4). From $b = 9$ ($R = 45$) up to $b = 19$ ($R = 95$), $K_{opt[Evanno]}$ remained stable, but changed at $b = 20$ again. The ratios between maximum and second highest ΔK peaks varied largely even when $K_{opt[Evanno]}$ remained unchanged. This ratio is an indicator for the impact of the variance of $P(K)$ on ΔK estimates and is more sensitive for the variance than $K_{opt[Evanno]}$.

In contrast to joint analysis of all breeds, $K_{opt[Evanno]}$ for the subset of only 11 breeds converged by the second block (Figure 3b). The $K_{opt[Evanno]}$ for re-analyzed Anacapa deer mice (Ozer et al., 2011a) converged by the third repeat block at $K_{opt[Evanno]} = 2$. This

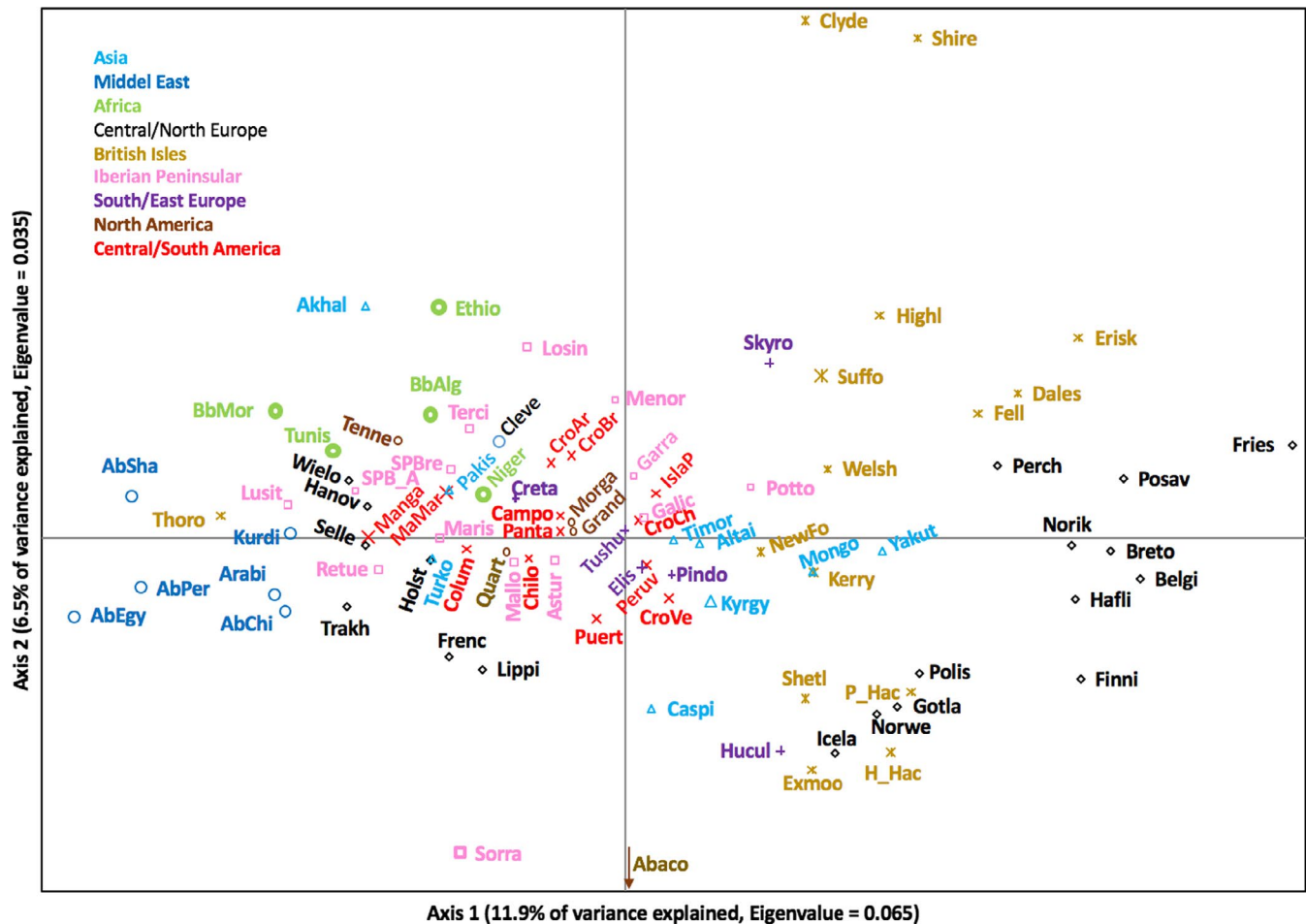


FIGURE 1 Scatter plot of the first three PCoA axes of genetic variation estimated by pairwise F_{ST} between breeds. Colour codes correspond to geographic origin and abbreviations to breeds as defined in Table S1, Supporting information

partition corresponds with the geographic location of the 11 population samples, with one sample for the mainland and 10 samples of three island populations at and after release. Contrasting the large horse data set, the ratios between the highest versus the second highest and third highest ΔK peaks, respectively, and the correlations between subsequent repletion blocks of both cases showed very little variation after their respective stable $K_{opt[Evanno]}$ values were reached.

CLUMPAK identified major and minor modes without a bias toward a particular MCMC/ \pm LP scenario (Figure S3). Only the +LP model at $K = 2$ clustered the 150,000 MCMC results separate from the two other scenarios. Numbers of replicates within major modes correlated strongly between MCMCs for both the -LP model (Pearson's $r = .94$, $r = .97$ and $r = .97$ for 3,000,000 MCMCs versus 150,000 MCMCs, 750,000 MCMCs and 1,500,000 MCMCs, respectively) and the -LP model ($r = .97$ and $.99$ for 1,500,000 MCMCs versus 150,000 MCMCs and 750,000 MCMCs, respectively; excluding $K = 2$). Both LP models produced exclusively major modes when K was relatively large ($K > 17$ and $K > 9$ for -LP and +LP, respectively) but added minor modes at lower K . The re-application of the Evanno method using major modes only did not result in consistent $K_{opt[Evanno]}$ estimates (Figure S4).

While maximum cluster membership coefficients remained high with increasing K , median memberships dropped to approx. 75% and 55% for the -LP and +LP models, respectively (Figure S3). The stronger drop for the +LP model is reflected in the rise of the number of ghost clusters starting at lower K than that for -LP. For the -LP and +LP models, all $K > 26$ and $K > 18$, respectively, exhibited ghost clusters while lower K only rarely had them. A peak of ghost clusters was found at $K = 10$ for +LP, which parallels the large variances of mean $P(K)$ at the same K (Figure 2). Because occurrences of minor modes and ghost clusters were reversely distributed over K , only a relatively small number of K was without both (Figure 2). Nine such clusters without minor modes and ghosts ranged between $K = 13$ and $K = 26$ for -LP. For +LP, the range occurred at lower K , namely eight clusters between $K = 3$ and $K = 18$.

Cluster membership coefficients at various K values are visualized in Figure 4 for the two LP models prearranged by geographic origin. At $K = 3$, the resolution is very low for both models although the +LP model indicates some distinctive and homogeneous patterns including "Cold-blood" breeds, for example, Eriskay Pony and Belgium Draft, and breeds that are associated with the Arab horse, for example, Egyptian Arab and Thoroughbred. Only +LP separates the Przewalski horse and the Abaco, the most inbred

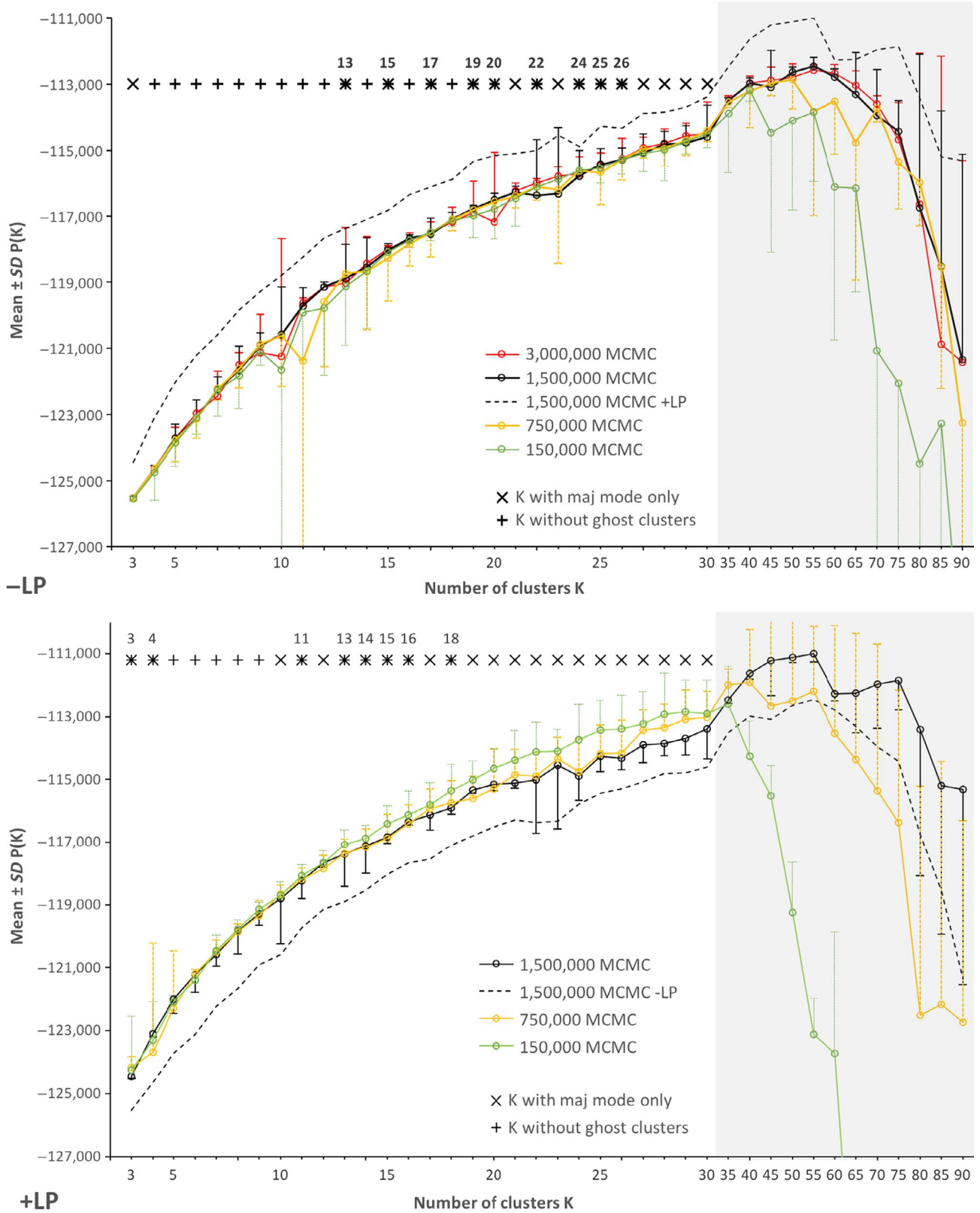


FIGURE 2 Log likelihood $P(K)$ values from the Bayesian clustering approach for explored cluster numbers K for the models with locprior and without locprior. Clusters were explored stepwise from $K=3$ to $K=30$ and then in steps of 5 up to $K=90$ (indicated by the background in grey). Shown are the averaged $\pm SD P(K)$ values for 40 replicates at each MCMC scenario. At the top of each graph those K values, which only exhibited major modes and did not show ghost clusters, are indicated and the K number is shown where both conditions were fulfilled (see also Fig S2, Supporting information)

domestic horse in our sample, from all others. This separation becomes evident by $K = 13$ for $-LP$. With increasing K , several patterns emerge:

First, breeds separate into two types: those which remain relatively homogenous and those which become more and more fragmented into an increasing number of clusters. These highly homogeneous breeds are Abaco, some Iberian breeds (Retuertás, Sorraia), some “Cold-blooded” breeds from the British Isles (Clydesdale, Exmoor, Shire), some Central and North European breeds (Belgian, Breton, Haflinger, French Trotter, Friesian, Icelandic, Norwegian, Trakehner, Wielkopolski), the Arabian breeds, some Asian breeds (Kurdish, Timor), and the Przewalski.

Second, regional trends and genetic similarities between breeds with shared breeding history emerge. By $K = 7$, geographic clustering has emerged in both LP models whereby the American and Iberian breeds tend to cluster jointly. By $K = 13$, European breeds are highly differentiated on a local/ regional scale including two groups of local breeds from the British Isles, three groups in central-North Europe, two groups in Africa and Asia, and several highly admixed breeds in all areas especially in Southeast Europe. This differentiation is stronger for $-LP$ than $+LP$. However, a substantial number of breeds did not show any clear-cut trend of separation, thus reflecting substantial admixture.

Third, differences between the two LP models emerge, for example, the American and Iberian breeds roughly split into North American, South American, and Iberians by $K = 13$ for $-LP$ while no such structuring becomes evident for $+LP$. Iberian and American breeds are more fragmented within breeds for $+LP$ than for $-LP$. Pronounced differences include the Clydesdale, Dales, and Eriskay group, which is tightly linked up to $K \approx 17$ for $-LP$ but disassociated by $K = 13$ for $+LP$, respectively. The Dales and Exmoor pony and the Haflinger are among those breeds which remain genetically homogenous up to $K = 45$ at $-LP$ (excluding many ghost clusters) but clearly fragmented at the same K for $+LP$.

Fourth, adding clusters with smaller membership coefficients does not follow a linear pattern over increasing K for some breeds. For example, Przewalski and Timor Pony cluster separated for $-LP$ at $K = 7$, joint at $K = 13$, separated at $K = 17$, and joint again at $K = 27$. Cleveland Bay fragments up to around $K \approx 13$ and homogenizes at $K \approx 17$ and higher for $-LP$, and Exmoor fragments up to $K \approx 10$ and homogenizes at $K \approx 13$ and higher for both locprior scenarios.

Lastly, the emergence of ghost clusters as shown by the quantitative analysis (Figure S3) becomes visually evident by $K \approx 27$ and is strong at $K \approx 45$ for both LP models.

3.5 | FLOCK

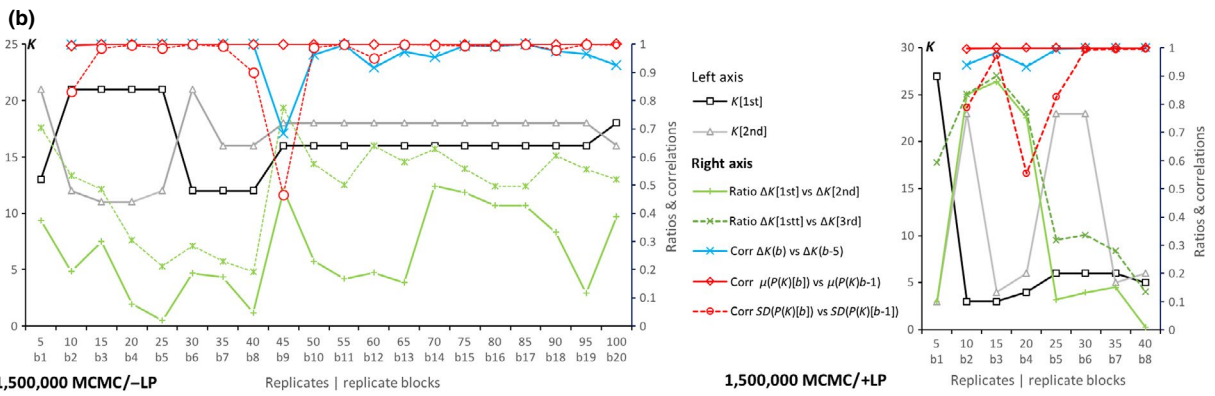
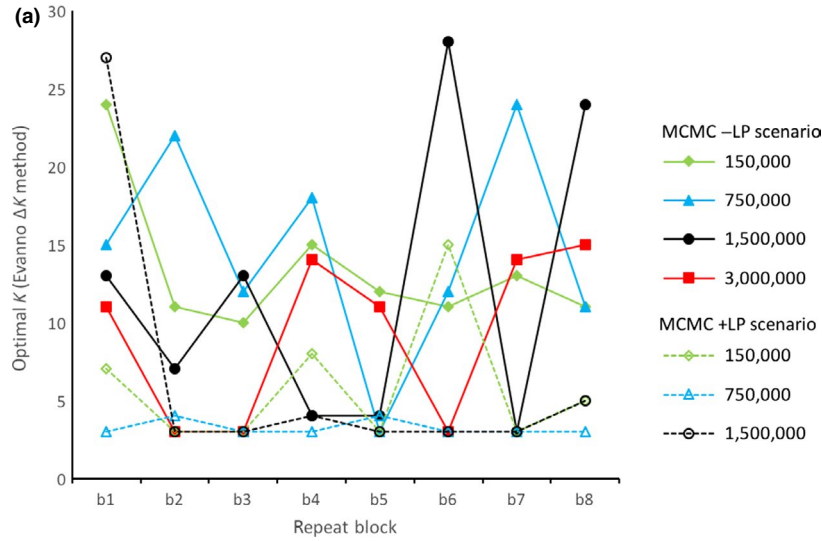
No stopping condition was reached when all breeds were evaluated (Figure S5). The inspection of the $LLOD_{K=2}$ scores indicates the Przewalski horse and the Timor Pony as visually clear outliers. When analyzed separately, each showed a particularly long, and thus well supported, plateau length over 50 runs at $K = 3$. The individuals were

completely separated into three clusters with Przewalski occurring only in cluster C and Timor Ponies in clusters D or E (Figure 5).

After excluding both the Timor and Przewalski, the analysis of the 92 remaining breeds failed to produce a stopping condition again. The inspection of the sample allocation matrix separated 28 admixed breeds. Repeated analysis with the remaining breeds produces a stopping condition with cluster A, representing mainly ponies and the “Cold-blood” group, and cluster B, representing mainly the “Warm-blood” group. In total, six hierarchical levels were identified. Three groups of breeds with admixed ancestry were restricted to the top two levels and included with 40 breeds a substantial proportion (43%) of all investigated breeds. The 54 assigned breeds partitioned into 26 final clusters without any further detected substructure. Partitioning always occurred between breeds except for cluster B2 β d. The three largest of the final clusters include a cluster of mainly ponies ranging from the British Isles to the Carpathian Mountains, A1, a cluster of horses mainly of Iberian origin, B2 α , and an admixed cluster, B1 β a/b, of Arab horses. The successful clustering of 54 breeds highlights that the 40 breeds that resulted in an “undecided stopping condition” are caused not by a global information deficiency of the used microsatellites but by admixture. When adding the admixed breeds one by one to a reanalysis, successful stopping conditions were reached in all cases but three. The total 26 clusters, that could not be further subdivided, correspond to the maximum optimal K , $K_{opt[FLOCK]} = 26$. FLOCK’s separation of admixed and non-admixed breeds at the hierarchical levels 1 and 2, in many cases, did not correspond with the cluster membership coefficients seen for STRUCTURE (Figure 5). This again shows the difficulty of working with a single species differentiated as breeds. Historically, many of these groups were crossed either accidentally or intentionally when horses from different regions were brought into contact by man. As well, many modern breeds, although considered distinct today, were created by crossing two or more breeds.

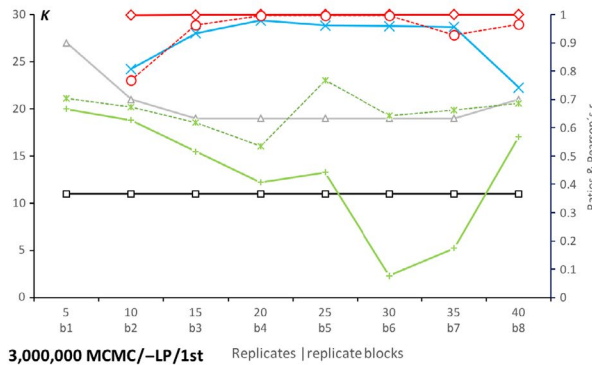
3.6 | DAPC

The Bayesian information criterion (BIC) was minimal between approximately $K = 20$ and $K = 40$ at the “elbow” of BIC values (Figure 6). While this range of K provides a useful parameter space to adequately describe the data, it did not provide a single, clear minimum BIC value at any K . Ward’s clustering method differentiated $K = 27$ as the model with the sharpest decrease of BIC values, indicating $K_{opt[DAPC]} = 27$. Retaining the first 50 principal components and eight discriminants, eleven of the 27 $K_{opt[DAPC]}$ clusters successfully reassigned individuals by DA to their PCA clusters in over 75% of cases. This is mirrored in the DA scatter plots where some genotypic clusters separate clearly along the first four DA axes (Figure 6). The remaining 15 of the 27 $K_{opt[DAPC]}$ clusters had lower proportions of successful DA reassignment (0.34–0.74) and did not separate in the plots. Analyses retaining 75 PCs with 8 and 15 discriminants gave very similar results (not shown). The distribution of the $K = 27$ posterior cluster memberships in breeds is visualized in Figure 4, alongside

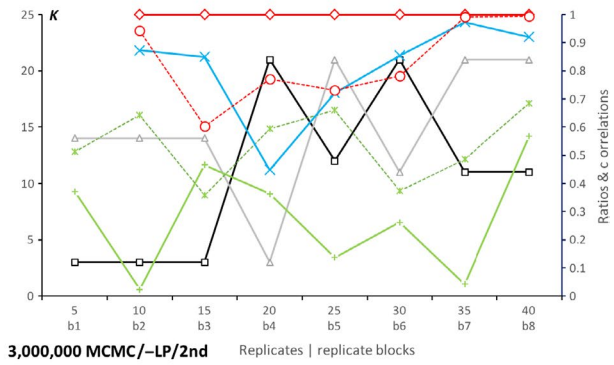


1,500,000 MCMC/-LP

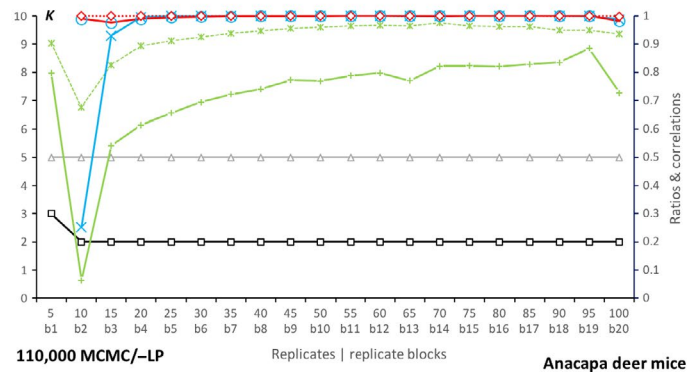
1,500,000 MCMC/+LP



3,000,000 MCMC/-LP/1st

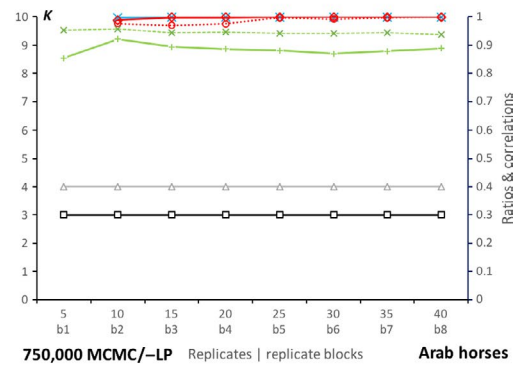


3,000,000 MCMC/-LP/2nd



110,000 MCMC/-LP

Anacapa deer mice



750,000 MCMC/-LP

Arab horses

FIGURE 3 Estimation of the optimal number of clusters $K_{opt[Evanno]}$ using the ΔK method (Evanno et al., 2005) for different numbers of replicates. The log likelihoods $P(K)$ and the ΔK values were calculated for $K=2$ to $K=30$. The number of replicates was subdivided into blocks consisting 5 replicates each. Panel A shows K for each of eight blocks b for different MCMC / \pm LP scenarios which used 40 replicates each. Panel B shows parameters of the ΔK method over accumulated blocks for up to 20 blocks (b1 with 5 replicates up to b20 with 100 replicates). Plotted parameters are the highest, the second and third highest ΔK peaks ($K[1^{st}] = K_{opt[Evanno]}$, $K[2^{nd}]$ and $K[3^{rd}]$), the ratios between the heights of these peaks, the correlations for the mean \pm SD $P(K)$ values over K ($\mu P(K)$ and $SD P(K)$), and the ΔK values between each current b versus the previous $b-1$. For the whole horse data set, the graphs for the 1,500,000 MCMC /-LP scenario, the 1,500,000 MCMC /-LP scenario, and for two iterations of the 3,500,000 MCMC /-LP scenario are shown. For Arab horses, the results for the 750,000 MCMC /-LP scenario and for the Anapaca deer mouse dataset the results for the 110,000 MCMC /-LP set (i.e. the STRUCTURE settings used in the original analysis by Ozer et al., 2011a) are shown. In each case, 40 replicates were conducted except for the 1,500,000 MCMC /-LP scenario with 100 replicates

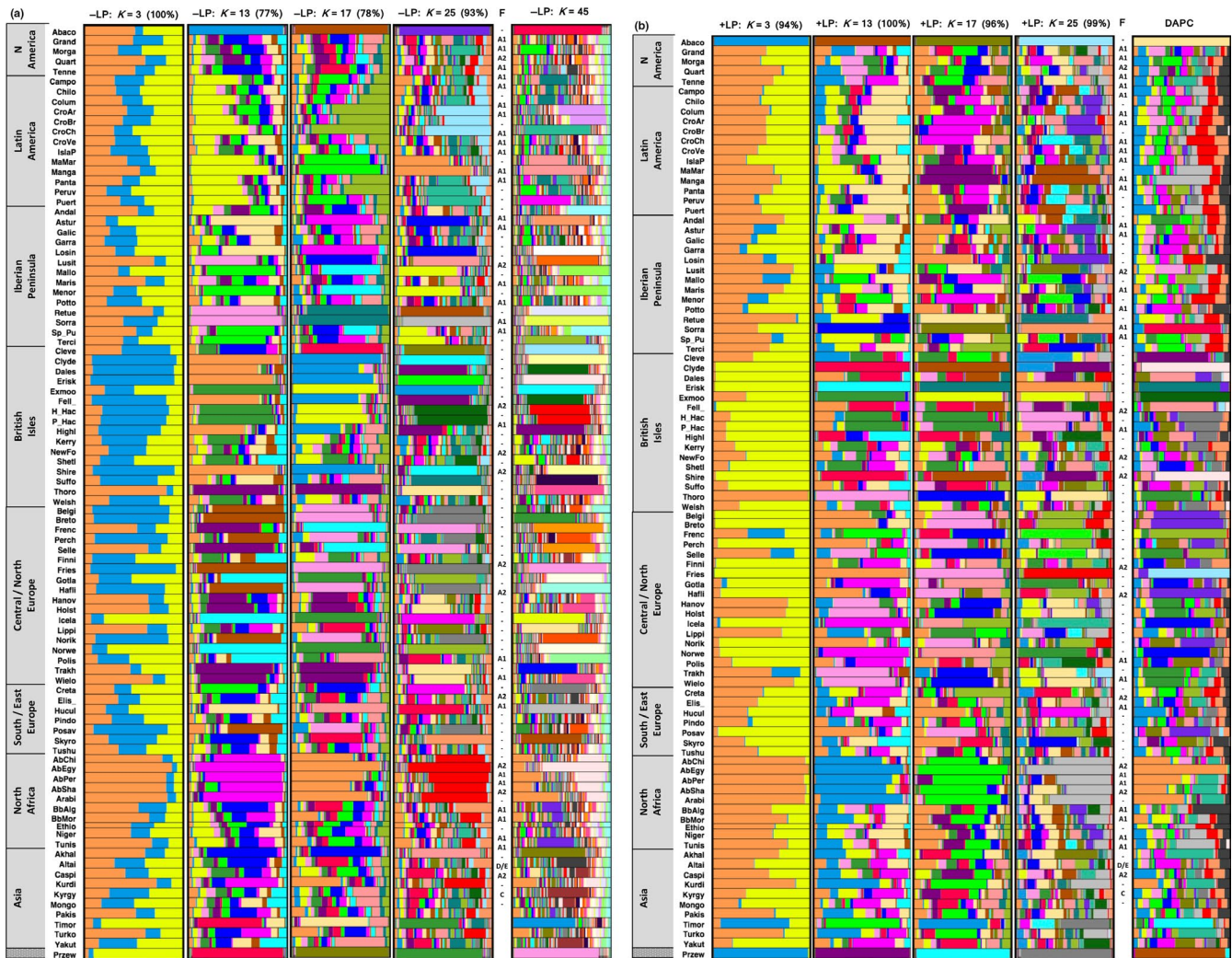


FIGURE 4 Membership coefficients for horse breeds from STRUCTURE at various K and \pm LP models, posterior probabilities for DAPC cluster assignments at the optimal $K_{opt[DAPC]}=27$, and summary of FLOCK breed cluster assignments. Panels for STRUCTURE are specified by the \pm LP models, visualize membership coefficients within each genetic cluster as indicated by different colours. The membership coefficients shown were averaged over all horses within each breed. With the exception for $K=45$, the membership coefficients for each horse are averages for all replicates found in the major modes as estimated by CLUMPAK. CLUMPAK analysed 160 replicates stemming from 40 replicates from each MCMC/-LP scenario and 120 replicates stemming from 40 replicates from each MCMC/+LP scenario. The percentages of replicates within the major clusters are presented in brackets. All these presented panels correspond to K without ghost clusters and with only majority modes (Fig. 2). The exception $K=45$ refers to the estimated optimum, $K_{opt[Pritchard]}=45$, which was inferred by the visual Pritchard method and includes ghost clusters. Major mode could not be calculated as the data volume for $K=45$ could not be processed by CLUMPAK. Therefore, the membership coefficients are visualized for the replicate with the best $P(K=45)$ score. The column for FLOCK (F) refers to the breed assignment at the first hierarchical level and denotes the breed groups as presented in Fig. 5. For visual purposes, column F is shown jointly with the STRUCTURE -LP and +LP panels, respectively. The panel for DAPC cluster assignments are based on refer to the posterior probabilities for the optimal $K_{opt[DAPC]}=27$ (Fig. 6), averaged for each breed

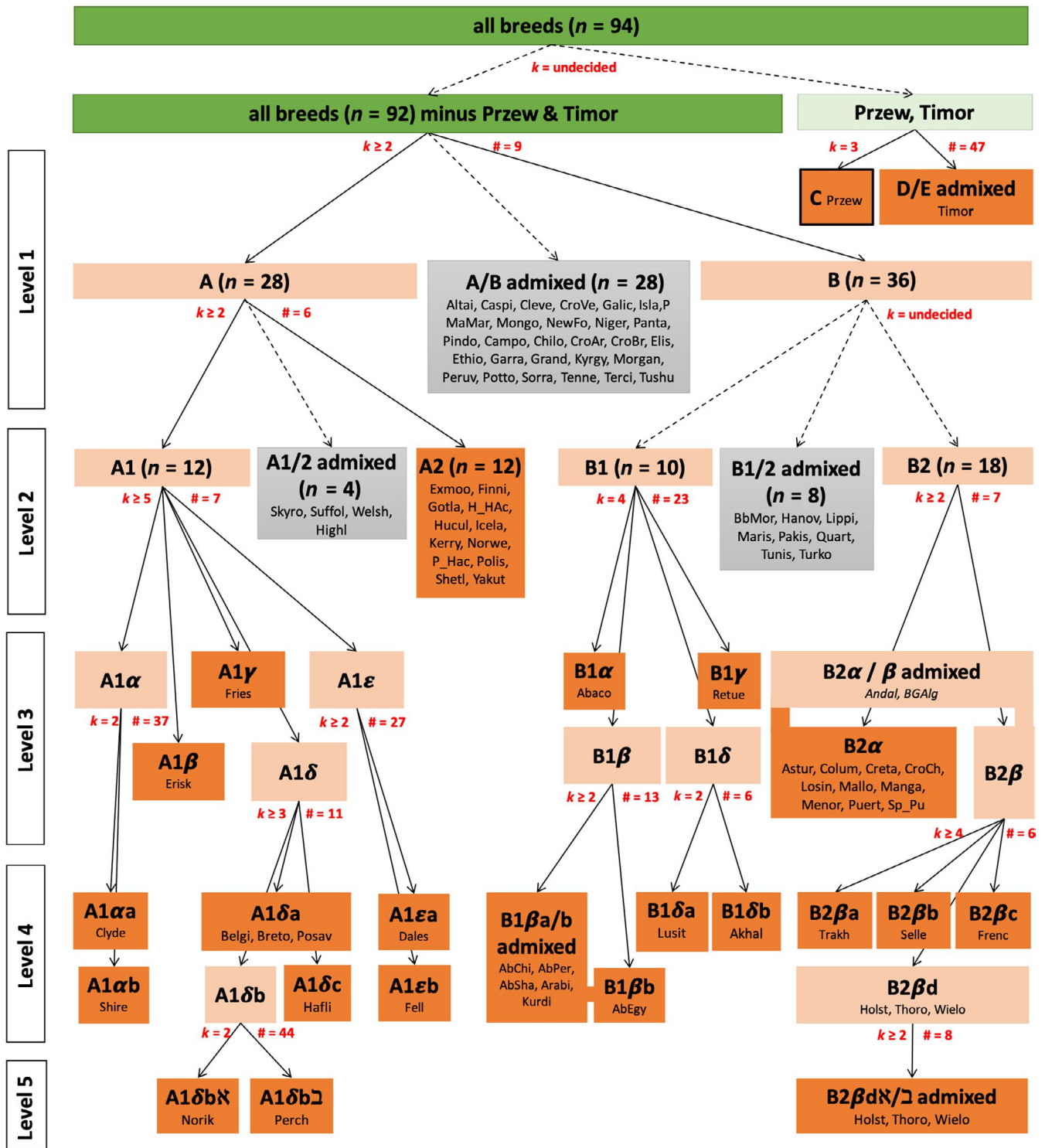


FIGURE 5 Summary of the FLOCK hierarchical analysis. Analyses were performed until there was no evidence for genetic substructure. On the top-level of the hierarchical analysis the breeds were partitioned by identifying outliers from the sample likelihood map (Fig S3, Supporting information; highlighted in green). For all subsequent hierarchical levels, outliers were not observed and not used for decision-making. Light red marks groups of breeds with genetic substructure and dark red marks non-dividable groups without genetic substructure. In each case, when FLOCK produced an “undecided stopping criterion” when all breeds from a group were included but reached a definite stopping criterion after admixed breeds were excluded, these admixed breeds (highlighted with grey and a dashed arrow to the parental breed group) were excluded from subsequent analysis. Partitioning of groups due to reached a definite stopping criterion are indicated by unbroken arrows. The maximum plateau length count, #, and the optimal number of clusters, k , interpreted from the stopping conditions, are shown in red. In the case where admixed breeds were excluded, the # and k values, refer to the repeated FLOCK analysis after the exclusion. Groups of breeds are named according to the hierarchical level of the analysis

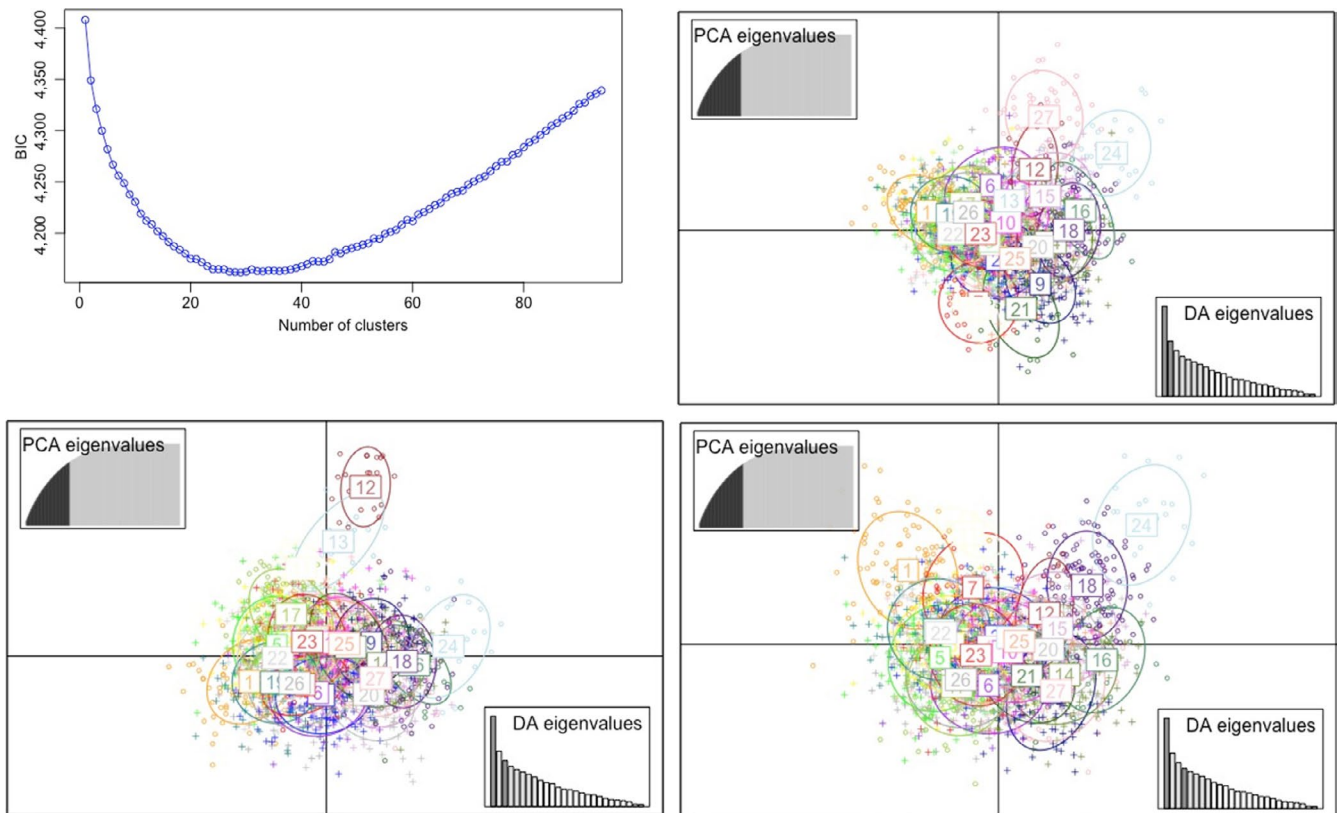


FIGURE 6 DAPC analysis with $K=27$ selected by Ward's clustering method as the most parsimonious model describing the data set of 94 horse breeds. (A) Bayesian Information Criterion, BIC, values for all 94 clusters evaluated with $BIC_{K=27} = 4162.3$ indicated by an arrow. Scatter plots show the DAPC components 1 versus 2 (B), 1 versus 3 (C) and 1 versus 4 (D). Cluster membership of each individuals is depicted by distant colours inside their 95% inertia ellipses. Different symbols represent proportions of successful DA reassignment of individuals to their PCA clusters (circles for $Prop_{cluster} > 0.75$ and crosses for $Prop_{cluster} < 0.75$). The accumulated amount of variance explained by stepwise increasing number PA is shown in the inserted graphs at the top left of the scatter plots; the cut-off point of the 50 PA used for the analysis represents 77.3 % of the total variance (shown in dark grey). DA eigenvalues are shown in the inserted graphs at the bottom right, whereby the numbers of the discriminants plotted against each other are indicated by dark grey and the eight discriminants retained for the analysis in light grey

the STRUCTURE results. The visual pattern corresponds well with STRUCTURE's membership coefficients at $K \approx 13$ for both LP scenarios. However, the posterior cluster memberships are strongly fragmented and reveal only strong genetic overlap between breeds that are known to be similar according to breed history (the Arabian horses, the Belgian/ Breton/ Haflinger group, and Mangalarga/ Mangalarga Marchador, Figure 4b). However, groups of genetically associated breeds, especially the Argentinian, Brazilian, and Chilean Criollo breeds, are not identified as a breed group, which is in contrast to the results of PCoA and STRUCTURE at both LP scenarios, and the known breed histories.

4 | DISCUSSION

Today, there are about 600 populations of horses that could be recognized as breeds (Hendricks, 2007). Analyses of nuclear genetic variation of 36 breeds using SNP markers have shown that most fall into five major breed groups (Petersen et al., 2013) and our results correspond with these categories. These are (a) Oriental breeds of with

the true Arabian breeds as one subgroup and other Asian and Middle Eastern breeds as a separate subgroup; (b) European breeds east of Spain which show strong influence from the English Thoroughbred; (c) Iberian breeds from Spain and Portugal plus American breeds with strong Iberian backgrounds; (d) "cold-blooded" horses of heavy draft breeds and true pony breeds which represent subgroups within this group; and (e) North American breeds which largely represent English and Spanish heritage with components of French, Dutch, and other European ancestry. Within these groups are breeds which do not well fit their group of origin due to severe loss of genetic diversity. Populations such as Abaco and the Sorraia in our study are good examples. However, fine-scale population genetic structure remains unclear.

We used a large set of 94 breeds, covering an enormous amount of differing breed histories, to evaluate the recovery of such fine-scale structure. The analysis of genetic population structure ultimately aims to recover the true, but unknown, underpinning processes leading to observed genetic patterns and, thus, the true but unknown number of population clusters. Besides understanding evolutionary processes, the evaluation is of concern for applied

management as the resources available for conservation of endangered domestic breeds are quite limited. Thus, knowledge that two separate but similar populations are indeed different is critical for proper resource use. However, this is inherently difficult (Evanno et al., 2005; Pritchard et al., 2010). Our large empiric data set produced reliable mean $P(K)$ estimates even with relatively low numbers of MCMC iterations and replicates. However, despite that our number of MCMCs (up to three million) and replicates (up to 100) is among the highest in the literature the ΔK estimator failed to converge, powerfully demonstrating the problems the ΔK statistics may face. How often this problem might have occurred in published studies remains unknown as studies have, in general, not evaluated the issue. The large observed variance between MCMC iterations within each evaluated K for both locprior models is the proximate reason for the poor performance of the ΔK method. Variance over K was as expected with mostly low values up to the plateau of the $P(K)$ (Pritchard et al., 2000). We partitioned the replicated MCMC iterations in replicate blocks to evaluate the variance introduced by the MCMC characteristics. The variance between these replicate blocks was large and correlations between them were low. Consequently, the ratios between the highest ΔK peaks, that is, $K_{\text{opt}[\text{Evanno}]}$, and the second highest peaks changed constantly over accumulated replicate blocks leading often to changes in $K_{\text{opt}[\text{Evanno}]}$ even at the tail end of the up to 100 conducted replicates. Especially worrisome is the observation that $K_{\text{opt}[\text{Evanno}]}$ appeared to have converged over long stretches of replicates in some cases, but then changed again, indicating a random walk characteristics of the MCMC procedure. In another case, $K_{\text{opt}[\text{Evanno}]}$ was stable over 40 replicates, but a recalculation of randomized blocks revealed different $K_{\text{opt}[\text{Evanno}]}$ estimates. Thus, even apparently stable ΔK estimates might be misleading when variance and ratios between the highest ΔK peaks, that is, $K_{\text{opt}[\text{Evanno}]}$ and the second highest peaks are high. Possible factors underpinning the failure to converge on stable ΔK estimates include the general statistical aspects of the estimation procedure per se and how it is being applied, the biological aspects of population history and structure, and the sampling procedure.

Unequal sample sizes between populations, which can skew analyses (Puechmaile, 2016), are unlikely to have contributed to the observed problems of unstable $K_{\text{opt}[\text{Evanno}]}$ estimates as we standardized sample sizes across all breeds. Furthermore, the analysis of the statistical power of the marker set across all test scenarios, which included the range of observed F_{ST} values, confirmed high power, making it unlikely that the marker set was insufficiently large to detect genetic structuring. Failure to reliably identify the main genetic clusters has been attributed to STRUCTURE's inherent feature of forcing genetic components into too few clusters (Kalinowski, 2011; Wang et al., 2007). A CLUMPAK pattern of multimodality indicates insufficient numbers of available clusters for the MCMC searching space to consistently assign genotypes to the same clusters and thus reaching unambiguous results (Wang et al., 2007). The random walk characteristic of MCMC processes may cause convergence at suboptimal solutions in the clustering space of possible genetic membership coefficients. However, the re-application of the ΔK method

for major modes still resulted in instable $K_{\text{opt}[\text{Evanno}]}$ indicating that the presence of suboptimal solutions was unlikely the cause of the observed instability in the empiric data set. STRUCTURE allocates some genetic subsets to spurious ghost clusters instead of leaving them empty, thus increasing the variance of the $P(K)$ likelihoods.

The 15 autosomal microsatellite markers we applied have been in widespread use since they were recommended for diversity studies by ISAG-FAO and the International Society for Animal Genetics (Cothran & Luís, 2005; FAO, 2011). Despite the advance of SNP genotyping, very few studies of horses have used SNPs to date for phylogenetic questions (Petersen et al., 2013) and the use of use of 12 to 17 microsatellites for breed analysis remains popular (e.g., Cosenza, La Rosa, Rosati, & Chiofalo, 2019; Isakova et al., 2019; Khanshour, Hempsey, Juras, & Cothran, 2019; Khaudov et al., 2019; Ustyantseva, Khrabrova, Abramova, & Ryabova, 2019; Zeng et al., 2019). The dendrogram of breed relationships analyzed by SNP typing of 36 breeds (Petersen et al., 2013) is almost an exact match of the results from the 15 microsatellite loci, thus suggesting that the results from our 15 microsatellites are robust. Increase of microsatellite numbers would likely allow higher resolution in the analysis of the genetic architecture of horses as has been observed in numerous other cases (e.g., in chicken, Gärke et al., 2012). Whether a higher number of loci would increase or decrease the performance of the MCMC searches of the parameter space for large sample sets needs to be evaluated.

4.1 | Biological aspects of population history and structure

The evolutionary history and genetic structure of horses might represent a particularly difficult case for STRUCTURE to solve. The relationships among domestic breeds are complex due to a high degree of mixing over generations. Horses were domesticated largely for their transportation abilities, and this ability has been widely used for at least the past 4,000 years. As horses were moved from one place to another, there was interbreeding of the invading horses with the resident horses. Widespread bidirectional gene flow and reticulate events persisted during and after domestication including the Przewalski and an extinct wild, taxonomically undescribed horse population (Der Sarkissian et al., 2015; Pardi & Scornavacca, 2015; Schubert et al., 2014; Warmuth et al., 2012). A strong sex bias during domestication led to differential maternal and paternal contributions to the overall gene pool and characterizes the founding and improvements of modern breeds especially since the formation of the earliest studbook (Lipizzaner in 1580, Galov et al., 2013; Wallner et al., 2017). The complex breed relations are also exemplified by the co-existence of closed breeds without admixture from outside breeds, open breeds with admixture, and different natural and artificial selective pressures. Variation in mtDNA control region sequences gives a clear example of this crossbreeding. Phylogenetic trees typically resulted in low bootstrap values and thus low statistical support irrespective of tree-building and distance algorithms (Conant et al., 2012; Cothran & Luís, 2005;

Pires et al., 2016). An epiphenomenon of the highly complex genetic basis appears to be the lack of a clear separation of breeds at low K and the lack of breed separation of breed membership coefficients at very high K . In contrast, such strong clustering was observed in domestic sheep, which are by far less intensively managed than horses (Grégoire Leroy et al., 2015). Extensive human data sets representing the complex migration and dispersal patterns during and after the peopling of the Americas have also demonstrated that it is difficult to adequately infer genetic structure for complex data (Corander et al., 2008; Wang et al., 2007).

Although STRUCTURE reveals which breeds are largely admixed, it does not reveal which breeds contribute to the large variances of $P(K)$. In contrast, the FLOCK results pointed at the admixed breeds which prevented the algorithm to find solutions for genetic structuring. Excluding them, breed clustering could be identified on a hierarchical level. The FLOCK algorithm has previously been criticized, especially for large data sets, because the reliance on inference rules regarding its “plateau record” is regarded as “not helpful” (Anderson & Barry, 2015). Indeed, the rules appear arbitrary, but the main strength is the important evaluation of the consequences of low information content of the marker system to find reliable solutions (Duchesne & Turgeon, 2016; Duchesne & Turgeon, 2012; Orozco-Wengel, Corander, & Schloetterer, 2011; Putman & Carbone, 2014). In contrast, the quantitative decision-making process of the ΔK method appears statistically elegant, but it always produces a “solution” whether or not the solution is adequate. The additional, so far not utilized, strength is FLOCK’s ability to identify whether failure to identify population structure is based on high levels of admixture.

4.2 | Sample structure

STRUCTURE performs best with a small number of discrete populations (Pritchard et al., 2000), possibly reflecting that MCMC search algorithms face increasing difficulties to find stable solutions when searching complex and large parameter spaces. Although the analysis of a smaller subset of horses and breeds indeed demonstrated convergence at $K_{\text{opt}[\text{Evanno}]}$ values, the number of samples appears not to be the crucial factor per se for consistently finding a solution in the large parameter spaces. This was demonstrated by our reanalysis of the large data set of Anacapa Island mice with eleven population samples and 1,361 individuals, where the increase of the originally published $R = 20$ repetitions to $R = 100$ did not produce any change in the $K_{\text{opt}[\text{Evanno}]}$ estimate. Thus, the crucial factor appears to be the large number of populations and the genetic relationships between them. Large samples of populations can increase the possibility to include complex structures and reticulate events represented in the sample. This is especially true for domestic animals where large variation of breed origin, managed breed relationships, natural and artificial selection, selection of founder animals, population size, and demographic history is common.

Most studies of horses that have utilized STRUCTURE have examined a small number of breeds, often from a specific geographic locality (Barcaccia et al., 2013; Berber et al., 2014; Bömcke, Gengler, & Cothran, 2011; Conant et al., 2012; Cothran et al., 2011; Galov et al., 2013; Janova et al., 2013; Khanshour, Conant, et al., 2013; Khanshour et al., 2015; Koban et al., 2012; Kusza et al., 2013; Lopes et al., 2015; Mackowski, Mucha, Cholewinski, & Cieslak, 2015; Mujica, 2006; Pablo Gómez et al., 2017; Pires et al., 2016; Prystupa, Juras, Cothran, Buchanan, & Plante, 2012; Rendo, Iriondo, Manzano, & Estonba, 2012; Sereno, Sereno, Vega-Pla, Kelly, & Bermejo, 2008; Tozaki et al., 2003; Uzans, Lucas, McLeod, & Frasier, 2015). The ΔK obtained in these studies appeared reasonable for the number of breeds. In some cases, a specific breed was compared to a large number of other breeds in a phylogenetic analysis but the breeds used in their STRUCTURE analysis were a subset of those in the phylogenetic analysis based upon what was found with the phylogenetic tree (Khanshour et al., 2015). In other cases, a small number of closely related breeds were analyzed to reveal fine structure within the group which could indicate differences in the histories of the individual breeds within the group (Khanshour, Conant, et al., 2013; Khanshour, Conant, et al., 2013).

4.3 | Recommendations

First, the possibility of interpretation based on false ΔK estimates is real because ΔK will always give a solution even with a very small number of replicates. Convergence should be evaluated by running many more replicates than the currently recommended $R = 20$. Care must be taken that the random walk characteristics of the MCMC algorithm can falsely suggest convergence even with larger numbers of replicates. The inspection of the ratios between the highest and second highest ΔK peaks will indicate the overall variance in the estimates and guide the choice of R .

Second, the evaluation of convergence of ΔK using replicate blocks is a suitable technique to choose and to justify the final number of replicates. So far, the approach taken in the literature to check for convergence of STRUCTURE parameters focusses on single MCMC iterations and a visual inspection of key parameters, in particular the posterior probability of the data for a given K (Gilbert et al., 2012; Pritchard et al., 2010). However, this inspection is qualitative as there is no definition of when convergence has been achieved and how much variance is acceptable. In the literature, it is rarely applied or reported. In contrast, the monitoring of ΔK over repeats is a more stringent and defined approach, whether ΔK is subsequently being used for decision-making or not.

Third, we suggest to identify a range of K that might feasibly explain the data well instead of using point estimates such as ΔK , corrected ΔK estimators (Puechmaille, 2016), or the Pritchard estimator. The over-reliance on K_{opt} and the enforcement of a specific value K_{opt} irrespective whether ΔK has converged can be inadequate and should be replaced by a qualitative description of clustering over increasing K , which appears scientifically more honest.

This adds strength to the increasing number of studies which forfeit the estimation of an optimal K and describe STRUCTURE results over K in relation to known natural history data from populations or breeds (e.g., Cortés et al., 2017 in horses; Leroy et al., 2015 in sheep). Although these studies do elaborate why point estimators have not been used, the underpinning rationale is to exercise precautionary care. Because the ΔK method identifies the uppermost level of clustering only, an over-reliance on this method might cause the missing of more subtle patterns. Again, this is an indication that subsets of the total breed set may offer a superior method of analysis. When cluster numbers above the maximum ΔK value are being interpreted from the membership plots, then ΔK appears to be of relatively little importance. A combination of those K values that neither produce CLUMPAK minor modes nor ghost clusters points to well-fitting MCMC solutions which neither enforce a too small number of clusters (minor modes) nor too many clusters (ghosts). They can then guide the subsequent interpretation in the context of known information on the populations or breeds. This approach has the major advantage that the overall variance between MCMC estimates is relatively unimportant because these parameters rely on mean $P(K)$ only. The estimates of the means are very stable even at small numbers of repeats and MCMC iterations. Mean $P(K)$ values for the horses converged relatively fast for all MCMC/LP scenario, for example, a correlation of >0.98 compared between 10 and 15 replicates and compared to the other MCMC scenarios. Thus, the computing time can be dramatically reduced from the normally prohibitive extensive times for large data sets. A faster STRUCTURE version is currently available only for SNP data and not microsatellites (Raj, Stephens, & Pritchard, 2014).

Fourth, we recommend to check whether the large-scale analysis clusters populations into smaller units that are consistent with known population processes such as breed histories. If this is the case, we suggest that these small-scale clusters are then analyzed separately. As in our case of Arabian breeds, the analysis of small numbers of sample populations is faster and performs better than the larger sample sets (Pritchard et al., 2000).

Finally, when the data or the known natural history indicate complex population relationships, it is advisable to augment Bayesian analysis with DAPC and FLOCK in order to check for consistencies and to better evaluate the data from different methodological angles. However, it must be noted that DAPC clearly did not characterize population subdivision better than STRUCTURE as evidenced by the failure to recover the similarities of the South American Criollo breeds, where STRUCTURE's results corresponded well with the known historic breed origin. Although DAPC revealed similar qualitative results as PCoA and revealed the same major breed clusters as STRUCTURE, it did not add further information as provided by the other methods. The usefulness of FLOCK for assessing the power of the concretely used genetic markers set has been pointed out previously (Putman & Carbone, 2014) but it is rarely applied (only 30 citations in WEB OF SCIENCE up to June 2017). Its additional major advantage

is that it allows to assess whether failure to successfully cluster populations is caused by the breed relationships themselves. We describe the corresponding strategy here (Duchesne, pers. communication) as it was not described in the publication describing FLOCK or the current manual (Duchesne et al., 2013; Duchesne & Turgeon, 2012).

ACKNOWLEDGMENTS

We thank the many horse owners, breeders, and colleagues (too many to name individually) that contributed the samples used in this study. In many cases, horses were typed for the purpose of parentage verification for entry into a specific breed registry which paid for genotype testing costs. Pierre Duchesne advised for the interpretation of the FLOCK results and Nils Ryman provided a modified POWSIM software version, adapted for our data set. Andres Ignacio Avila Barrera and Raphael Alejandro Verdugo Pincheira relentlessly supported computing. Two anonymous reviewers provided helpful comments. Funding for genotyping came from fees paid by horse owners for diagnostic genetic tests.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

EGC collected the data and provided ideas during analysis and writing of the manuscript. SMF performed the analyses and wrote the manuscript. SG, RJ, AK, FL, CL, AMM, AMM, FM, MMO, LO, YMS and JLPV provided samples. All authors provided ideas and contributed to the manuscript writing.

DATA AVAILABILITY STATEMENT

Microsatellite genotypes are available from Dryad (<https://doi.org/10.5061/dryad.tmpg4f4vh>).

ORCID

Stephan Michael Funk  <https://orcid.org/0000-0001-7992-4115>

Cristina Luis  <https://orcid.org/0000-0001-8005-9624>

Ernest Gus Cothran  <https://orcid.org/0000-0003-2791-4331>

REFERENCES

- Anderson, E. C., & Barry, P. D. (2015). Interpreting the FLOCK algorithm from a statistical perspective. *Molecular Ecology Resources*, 15(5), 1020–1030.
- Barcaccia, G., Felicetti, M., Galla, G., Capomaccio, S., Cappelli, K., Albertini, E., ... Verini Supplizi, A. (2013). Molecular analysis of genetic diversity, population structure and inbreeding level of the Italian Lipizzan horse. *Livestock Science*, 151(2–3), 124–133. <https://doi.org/10.1016/j.livsci.2012.11.022>
- Berber, N., Gaouar, S., Leroy, G., Kdidi, S., Tabet Aouel, N., & Saïdi Mehtar, N. (2014). Molecular characterization and differentiation of five horse breeds raised in Algeria using polymorphic microsatellite markers. *Journal of Animal Breeding and Genetics*, 131(5), 387–394. <https://doi.org/10.1111/jbg.12092>
- Bömcke, E., Gengler, N., & Cothran, E. (2011). Genetic variability in the Skyros pony and its relationship with other Greek and foreign horse breeds. *Genetics and Molecular Biology*, 34(1), 68–76.

- Cockerham, C. C., & Weir, B. S. (1993). Estimation of gene flow from F-statistics. *Evolution*, 47, 855–863.
- Conant, E. K., Juras, R., & Cothran, E. G. (2012). A microsatellite analysis of five Colonial Spanish horse populations of the southeastern United States: A microsatellite analysis of five Colonial Spanish horse populations. *Animal Genetics*, 43(1), 53–62. <https://doi.org/10.1111/j.1365-2052.2011.02210.x>
- Corander, J., Marttinen, P., Sirén, J., & Tang, J. (2008). Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*, 9(1), 539. <https://doi.org/10.1186/1471-2105-9-539>
- Cortés, O., Dunner, S., Gama, L. T., Martínez, A. M., Delgado, J. V., Ginja, C., ... Vega-Pla, J. L. (2017). The legacy of Columbus in American horse populations assessed by microsatellite markers. *Journal of Animal Breeding and Genetics*, 134(4), 340–350. <https://doi.org/10.1111/jbg.12255>
- Cosenza, M., La Rosa, V., Rosati, R., & Chiofalo, V. (2019). Genetic diversity of the Italian thoroughbred horse population. *Italian Journal of Animal Science*, 18(1), 538–545. <https://doi.org/10.1080/1828051X.2018.1547128>
- Cothran, E. G., Canelon, J. L., Luis, C., Conant, E., & Juras, R. (2011). Genetic analysis of the Venezuelan Criollo horse. *Genetics and Molecular Research*, 10(4), 2394–2403. <https://doi.org/10.4238/2011.October.7.1>
- Cothran, E. G., & Luis, C. (2005). Conservation genetics of endangered horse breeds. In I. Bodo, L. Alderson, & B. Langolis (Eds.), *Conservation genetics of endangered horse breeds*. Wageningen, The Netherlands: Academic Publishers.
- Der Sarkissian, C., Ermini, L., Schubert, M., Yang, M. A., Librado, P., Fumagalli, M., ... Orlando, L. (2015). Evolutionary genomics and conservation of the endangered Przewalski's horse. *Current Biology*, 25(19), 2577–2583. <https://doi.org/10.1016/j.cub.2015.08.032>
- Duchesne, P., Méthot, J., & Turgeon, J. (2013). *FLOCK 3.1 User Guide [Software manual]*. Retrieved from Département de biologie, Université Laval website: <https://zenodo.org/record/32466>
- Duchesne, P., & Turgeon, J. (2012). FLOCK provides reliable solutions to the “number of populations” problem. *Journal of Heredity*, 103(5), 734–743.
- Duchesne, P., & Turgeon, J. (2016). Interpreting the FLOCK algorithm: A reply to Anderson & Barry (2015). *Molecular Ecology Resources*, 16(1), 13–16. <https://doi.org/10.1111/1755-0998.12480>
- Earl, D. A., & von Holdt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14, 2611–2620.
- FAO (2011). *Molecular genetic characterization of animal genetic resources* (No. 9). Rome, Italy: FAO.
- Galov, A., Byrne, K., Gomerčić, T., Duras, M., Arbanasić, H., Sindičić, M., ... Funk, S. M. (2013). Genetic structure and admixture between the Posavina and Croatian Coldblood in contrast to Lipizzan horse from Croatia. *Czech Journal of Animal Science*, 58(2), 71–78.
- Gärke, C., Ytornel, F., Bed'hom, B., Gut, I., Lathrop, M., Weigend, S., & Simianer, H. (2012). Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations: Population differentiation: SNPs vs SSRs. *Animal Genetics*, 43(4), 419–428. <https://doi.org/10.1111/j.1365-2052.2011.02284.x>
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472.
- Gilbert, K. J., Andrew, R. L., Bock, D. G., Franklin, M. T., Kane, N. C., Moore, J.-S. (2012). Recommendations for utilizing and reporting population genetic analyses: The reproducibility of genetic clustering using the program structure. *Molecular Ecology*, 21(20), 4925–4930.
- Goudet, J. (1995). FSTAT A program for IBM PC compatibles to calculate Weir and Cockerham's (1984) estimators of F-statistics (version 1.2). *Journal of Heredity*, 86, 485–486.
- Guillot, G., Mortier, F., & Estoup, A. (2005). GENELAND: A computer package for landscape genetics. *Molecular Ecology Resources*, 5(3), 712–715.
- Guo, S. W., & Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, 48(2), 361–372.
- Hendricks, B. L. (2007). *International encyclopedia of horse breeds*. Norman, OK: University of Oklahoma Press.
- Hendrickson, S. L. (2013). A genome wide study of genetic adaptation to high altitude in feral Andean Horses of the páramo. *BMC Evolutionary Biology*, 13(1), 273.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9(5), 1322–1332. <https://doi.org/10.1111/j.1755-0998.2009.02591.x>
- Isakova, Z. T., Toktosunov, B. I., Kipen, V. N., Kalinkova, L. V., Talaibekova, E. T., Aldasheva, N. M., & Abdurasulov, A. H. (2019). Phylogenetic analysis of Kyrgyz horse using 17 microsatellite markers. *Russian Journal of Genetics*, 55(1), 100–104. <https://doi.org/10.1134/S1022795419010071>
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Janova, E., Futas, J., Klumplerova, M., Putnova, L., Vrtkova, I., Vyskocil, M., ... Horin, P. (2013). Genetic diversity and conservation in a small endangered horse population. *Journal of Applied Genetics*, 54(3), 285–292. <https://doi.org/10.1007/s13353-013-0151-3>
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart, T., & Collins, C. (2016). *A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.2*. Retrieved from <http://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf>
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1), 94. <https://doi.org/10.1186/1471-2156-11-94>
- Jombart, T., Kamvar, Z. N., Lustrik, R., Collins, C., Beugin, M.-P., Knaus, B., ... Calboli, F. (2016). *Package 'adegenet' (Version 2.0.1)*.
- Juras, R., Cothran, E. G., & Klimas, R. (2003). Genetic analysis of three Lithuanian native horse breeds. *Acta Agriculturae Scandinavica, Section A - Animal Science*, 53(4), 180–185.
- Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: Simulations and implications for human population structure. *Heredity*, 106(4), 625–632.
- Khanshour, A., Conant, E., Juras, R., & Cothran, E. G. (2013). Microsatellite analysis of genetic diversity and population structure of Arabian horse populations. *Journal of Heredity*, 104(3), 386–398. <https://doi.org/10.1093/jhered/est003>
- Khanshour, A. M., Hempsey, E. K., Juras, R., & Cothran, E. G. (2019). Genetic characterization of Cleveland Bay Horse Breed. *Diversity*, 11(10), 174. <https://doi.org/10.3390/d11100174>
- Khanshour, A., Juras, R., Blackburn, R., & Cothran, E. G. (2015). The legend of the Canadian Horse: Genetic diversity and breed origin. *Journal of Heredity*, 106(1), 37–44. <https://doi.org/10.1093/jhered/esu074>
- Khanshour, A. M., Juras, R., & Cothran, E. G. (2013). Microsatellite analysis of genetic variability in Waler horses from Australia. *Australian Journal of Zoology*, 61(5), 357. <https://doi.org/10.1071/ZO13062>

- Khoudov, A. D., Duduev, A. S., Kokov, Z. A., Amshokov, H. K., Zhekamuhov, M. H., Zaitsev, A. M., & Reissmann, M. (2019). Diversity of Kabardian horses and their genetic relationships with selected breeds in the Russian Federation based on 17 microsatellite loci. *IOP Conference Series: Earth and Environmental Science*, 341, 012072. <https://doi.org/10.1088/1755-1315/341/1/012072>
- Koban, E., Denizci, M., Aslan, O., Aktoprakligil, D., Aksu, S., Bower, M., ... Arat, S. (2012). High microsatellite and mitochondrial diversity in Anatolian native horse breeds shows Anatolia as a genetic conduit between Europe and Asia: High genetic diversity found in Anatolian horses. *Animal Genetics*, 43(4), 401–409. <https://doi.org/10.1111/j.1365-2052.2011.02285.x>
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). CLUMPAK: A program for identifying clustering modes and packaging population structure inferences across. *Molecular Ecology Resources*, 15(5), 1179–1191. <https://doi.org/10.1111/1755-0998.12387>
- Kusza, S., Priskin, K., Ivankovic, A., Jedrzejewska, B., Podgorski, T., Jávora, A., & Mihók, S. (2013). Genetic characterization and population bottleneck in the Hucul horse based on microsatellite and mitochondrial data. *Biological Journal of the Linnean Society*, 109(1), 54–65. <https://doi.org/10.1111/bj.12023>
- Leroy, G., Danchin-Burge, C., Palhière, I., SanCristobal, M., Nédélec, Y., Verrier, E., & Rognon, X. (2015). How do introgression events shape the partitioning of diversity among breeds: A case study in sheep. *Genetics Selection Evolution*, 47(1), <https://doi.org/10.1186/s12711-015-0131-7>
- Leroy, G., Verrier, E., Meriaux, J. C., & Rognon, X. (2009). Genetic diversity of dog breeds: Between-breed diversity, breed assignment and conservation approaches. *Animal Genetics*, 40(3), 333–343. <https://doi.org/10.1111/j.1365-2052.2008.01843.x>
- Levine, M. A. (2005). Domestication and early history of the horse. In D. S. Mills & S. M. McDonnell (Eds.), *The domestic horse: The origins, development and management of its behaviour* (pp. 5–22). Cambridge, UK: Cambridge University Press.
- Librado, P., Fages, A., Gaunitz, C., Leonardi, M., Wagner, S., Khan, N., ... Orlando, L. (2016). The evolutionary origin and genetic makeup of domestic horses. *Genetics*, 204(2), 423–434. <https://doi.org/10.1534/genetics.116.194860>
- Lopes, M. S., Mendonça, D., Rojer, H., Cabral, V. A., Bettencourt, S. A. X., & da Cãmara Machado, A. (2015). Morphological and genetic characterization of an emerging Azorean horse breed: The Terceira Pony. *Frontiers in Genetics*, 6, 62. <https://doi.org/10.3389/fgene.2015.00062>
- Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., ... Willerslev, E. (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, 479(7373), 359–364. <https://doi.org/10.1038/nature10574>
- Luís, C., Cothran, E. G., & Oom, M. D. M. (2007). Inbreeding and genetic structure in the endangered Sorraia horse breed: Implications for its conservation and management. *Journal of Heredity*, 98(3), 232–237.
- Lukoschek, V., Riginos, C., & van Oppen, M. J. H. (2016). Congruent patterns of connectivity can inform management for broadcast spawning corals on the Great Barrier Reef. *Molecular Ecology*, 25(13), 3065–3080. <https://doi.org/10.1111/mec.13649>
- Mackowski, M., Mucha, S., Cholewinski, G., & Cieslak, J. (2015). Genetic diversity in Hucul and Polish primitive horse breeds. *Archives Animal Breeding*, 58(1), 23–31. <https://doi.org/10.5194/aab-58-23-2015>
- Martin-Burriel, I., Rodellar, C., Canon, J., Cortes, O., Dunner, S., Landi, V., ... Delgado, J. V. (2011). Genetic diversity, structure, and breed relationships in Iberian cattle. *Journal of Animal Science*, 89(4), 893–906. <https://doi.org/10.2527/jas.2010-3338>
- Martínez, A. M., Gama, L. T., Cañón, J., Ginja, C., Delgado, J. V., Dunner, S., ... Zaragoza, P. (2012). Genetic Footprints of Iberian Cattle in America 500 Years after the Arrival of Columbus. *PLoS ONE*, 7(11), e49066. <https://doi.org/10.1371/journal.pone.0049066>
- Mojena, R. (2006). Ward's clustering algorithm. *Encyclopedia of Statistical Sciences*, 15. <https://doi.org/10.1002/0471667196.ess2887.pub2>
- Mujica, F. (2006). *Diversidad, Conservación y Utilización de los Recursos Genéticos Animales en Chile*. Retrieved from <http://biblioteca.inia.cl/link.cgi/Catalogo/Boletines/>
- Orozco-terWengel, P., Corander, J., & Schloetterer, C. (2011). Genealogical lineage sorting leads to significant, but incorrect Bayesian multilocus inference of population structure. *Molecular Ecology*, 20(6), 1108–1121.
- Ozer, F., Gellerman, H., & Ashley, M. V. (2011a). Data from: Genetic impacts of Anacapa deer mice reintroductions following rat eradication. *Dryad Digital Repository*, <https://doi.org/10.5061/dryad.778c9>
- Ozer, F., Gellerman, H., & Ashley, M. V. (2011b). Genetic impacts of Anacapa deer mice reintroductions following rat eradication: Anacapa Island Mouse Reintroductions. *Molecular Ecology*, 20(17), 3525–3539. <https://doi.org/10.1111/j.1365-294X.2011.05165.x>
- Pablo Gómez, M., Landi, V., Martínez, A. M., Gómez Carpio, M., Nogales Baena, S., Delgado Bermejo, J. V., ... Vega-Pla, J. L. (2017). Genetic diversity of the semi-feral Marisemeño horse breed assessed with microsatellites. *Italian Journal of Animal Science*, 16(1), 14–21. <https://doi.org/10.1080/1828051X.2016.1241132>
- Paetkau, D., Calvert, W., Stirling, I., & Strobeck, C. (1995). Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, 4, 347–354.
- Pardi, F., & Scornavacca, C. (2015). Reconstructible phylogenetic networks: Do not distinguish the indistinguishable. *PLoS Computational Biology*, 11(4), e1004135. <https://doi.org/10.1371/journal.pcbi.1004135>
- Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28(19), 2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Petersen, J. L., Mickelson, J. R., Cothran, E. G., Andersson, L. S., Axelsson, J., Bailey, E., ... McCue, M. E. (2013). Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS ONE*, 8(1), e54997. <https://doi.org/10.1371/journal.pone.0054997>
- Pires, D. A. F., Coelho, E. G. A., Melo, J. B., Oliveira, D. A. A., Ribeiro, M. N., Cothran, G. E., & Juras, R. (2016). Genetic relationship between the Nordestino horse and national and international horse breeds. *Genetics and Molecular Research*, 15(2), 1–8. <https://doi.org/10.4238/gmr.15027881>
- Pires, D., Coelho, E., Melo, J. B., Oliveira, D., Ribeiro, M. N., Gus Cothran, E., ... Khanshour, A. (2014). Genetic diversity and population structure in remnant subpopulations of Nordestino horse breed. *Archivos de Zootecnia*, 63(242), 349–358.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Pritchard, J. K., Wen, X., & Falush, D. (2010). *Documentation for structure software Version 2.3 [Software manual]*. Retrieved from <http://pritch.bsd.uchicago.edu/structure.html>
- Pritchard, J. K., Wen, X., & Falush, D. (2012). *Structure for Mac OS X without front end (Version 2.3.4)*. Retrieved from http://pritchardlab.stanford.edu/structure_software/release_versions/v2.3.4/html/structure.html
- Prystupa, J. M., Juras, R., Cothran, E. G., Buchanan, F. C., & Plante, Y. (2012). Genetic diversity and admixture among Canadian, Mountain and Moorland and Nordic pony populations. *Animal*, 6(01), 19–30. <https://doi.org/10.1017/S1751731111001212>
- Puechmaile, S. J. (2016). The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the

- problem. *Molecular Ecology Resources*, 16(3), 608–627. <https://doi.org/10.1111/1755-0998.12512>
- Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, 4, 4399–4428. <https://doi.org/10.1002/ece3.1305>
- R Foundation for Statistical Computing (2016). *R (Version 3.3.1 GUI 1.68 Mavericks build)*. Retrieved from <https://www.r-project.org>
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, 197(2), 573–589. <https://doi.org/10.1534/genetics.114.164350>
- Rendo, F., Iriondo, M., Manzano, C., & Estroba, A. (2012). Effects of a 10-year conservation programme on the genetic diversity of the Pottoka pony – new clues regarding their origin: Genetic diversity and origin of Pottoka pony. *Journal of Animal Breeding and Genetics*, 129(3), 234–243. <https://doi.org/10.1111/j.1439-0388.2011.00955.x>
- Rosenberg, N. A. (2003). DISTRUCT: A program for the graphical display of population structure: PROGRAM NOTE. *Molecular Ecology Notes*, 4(1), 137–138. <https://doi.org/10.1046/j.1471-8286.2003.00566.x>
- Rousset, F. (2008). GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Ryman, N. (2011). POWSIM: A computer program for assessing statistical power when testing for genetic differentiation [Software manual]. Retrieved from <http://www.zoologi.su.se/~ryman/>
- Ryman, N., & Palm, S. (2006). POWSIM: A computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes*, 6(3), 600–602. <https://doi.org/10.1111/j.1471-8286.2006.01378.x>
- Ryman, N., Palm, S., André, C., Carvalho, G. R., Dahlgren, T. G., Jorde, P. E., ... Ruzzante, D. E. (2006). Power for detecting genetic divergence: Differences between statistical methods and marker loci: POWER FOR DETECTING DIVERGENCE. *Molecular Ecology*, 15(8), 2031–2045. <https://doi.org/10.1111/j.1365-294X.2006.02839.x>
- Schubert, M., Jónsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., ... Orlando, L. (2014). Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Sciences of the United States of America*, 111(52), E5661–E5669. <https://doi.org/10.1073/pnas.1416991111>
- Sereno, F. T. P. D. S., Sereno, J. R. B., Vega-Pla, J. L., Kelly, L., & Bermejo, J. V. D. (2008). Genetic diversity of Brazilian Pantaneiro horse and relationships among horse breeds. *Pesquisa Agropecuária Brasileira*, 43(5), 595–604.
- Sokolov, V. E., & Orlov, V. N. (1986). *The Przewalski horse and restoration to its natural habitat in Mongolia*. Rome, Italy: FAO, Agriculture and Consumer Protection Department.
- Tozaki, T., Takezaki, N., Hasegawa, T., Ishida, N., Kurosawa, M., Tomita, M., ... Mukoyama, H. (2003). Microsatellite variation in Japanese and Asian horses and their phylogenetic relationship using a European horse outgroup. *Journal of Heredity*, 94(5), 374–380.
- Ustyantseva, A. V., Khrabrova, L. A., Abramova, N. V., & Ryabova, T. N. (2019). Genetic characterization of Akhal-Teke horse subpopulations using 17 microsatellite loci. *IOP Conference Series: Earth and Environmental Science*, 341, 012070. <https://doi.org/10.1088/1755-1315/341/1/012070>
- Uzans, A. J., Lucas, Z., McLeod, B. A., & Frasier, T. R. (2015). Small N_e of the Isolated and Unmanaged Horse Population on Sable Island. *Journal of Heredity*, 106(5), 660–665. <https://doi.org/10.1093/jhered/esv051>
- van de Goor, L. H. P., van Haeringen, W. A., & Lenstra, J. A. (2011). Population studies of 17 equine STR for forensic and phylogenetic analysis: Equine population data. *Animal Genetics*, 42(6), 627–633. <https://doi.org/10.1111/j.1365-2052.2011.02194.x>
- Wallner, B., Palmieri, N., Vogl, C., Rigler, D., Bozlak, E., Druml, T., ... Brem, G. (2017). Y chromosome uncovers the recent oriental origin of modern stallions. *Current Biology*, 27(13), 2029–2035.e5. <https://doi.org/10.1016/j.cub.2017.05.086>
- Wang, S., Lewis, C. M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., ... Ruiz-Linares, A. (2007). Genetic variation and population structure in Native Americans. *PLoS Genetics*, 3(11), e185. <https://doi.org/10.1371/journal.pgen.0030185>
- Warmuth, V., Eriksson, A., Bower, M. A., Barker, G., Barrett, E., Hanks, B. K., ... Manica, A., (2012). Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proceedings of the National Academy of Sciences of the United States of America*, 109(21), 8202–8206. <https://doi.org/10.1073/pnas.111122109>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358–1370.
- Zachos, F. E., Frantz, A. C., Kuehn, R., Bertouille, S., Colyn, M., Niedziałkowska, M., ... Flamand, M.-C. (2016). Genetic structure and effective population sizes in European Red Deer (*Cervus elaphus*) at a Continental Scale: Insights from microsatellite DNA. *Journal of Heredity*, 107(4), 318–326. <https://doi.org/10.1093/jhered/esw011>
- Zeng, L., Chen, N., Yao, Y., Dang, R., Chen, H., & Lei, C. (2019). Analysis of genetic diversity and structure of Guanzhong Horse using microsatellite markers. *Animal Biotechnology*, 30(1), 95–98. <https://doi.org/10.1080/10495398.2017.1416392>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Funk SM, Guedaoura S, Juras R, et al. Major inconsistencies of inferred population genetic structure estimated in a large set of domestic horse breeds using microsatellites. *Ecol Evol*. 2020;10:4261–4279. <https://doi.org/10.1002/ece3.6195>