

Genetic analysis and management in small populations: the Asturcon pony as an example

Susana Dunner^{a*}, Maria L. Checa^a, Juan P. Gutierrez^a,
Juan P. Martin^b, Javier Cañon^a

^a Laboratorio de Genética Molecular, Departamento de Producción Animal,
Facultad de Veterinaria, 28040 Madrid, Spain

^b Departamento de Biología, Escuela Técnica Superior de Ingenieros Agrónomos,
28040 Madrid, Spain

(Received 24 March 1998; accepted 28 May 1998)

Abstract – Geneticists are faced with various problems when managing small natural populations (e.g. high inbreeding, loss of economic value). We propose here the management of a small population through the example of the Asturcon (a Celtic pony population) by examining two sources of information: a studbook created in 1981 and the polymorphism of ten microsatellite markers chosen according to the recommendations of ISAG (International Society of Animal Genetics). This information allows us to estimate several genetic parameters useful in assessing the genetic situation of the population in order to propose conservation strategies. Results show the reliability of molecular information in populations where no studbook exists. Overall inbreeding value (F) and fixation index (F_{IT}) are moderate ($F = 0.027$; $F_{IT} = 0.056$), effective number of founders is small ($n = 22$), and the population is divided into three distinct groups ($F_{ST} = 0.078$; $P < 0.001$). The molecular heterozygosity ($H_M = 71.2\%$) computed in a random sample gives an accurate vision of the real inbreeding. These parameters and the application of the concept of average relatedness allow us to recommend to the breeders the choice of the best matings to control the inbreeding level while maintaining a low paternity error rate. © Inra/Elsevier, Paris

genetic management / demographic parameters / microsatellite / equine

Résumé – Analyse génétique et gestion des petites populations : l'exemple du poney Asturcon. Les généticiens sont confrontés à plusieurs problèmes quand ils ont à gérer des petites populations animales, comme une consanguinité élevée et une perte d'intérêt économique. Ici on traite l'exemple du poney Asturcon à partir de deux sources d'information : le livre généalogique créé en 1981 et le polymorphisme de dix

* Correspondence and reprints
E-mail: Dunner@eucmax.sim.ucm.es

marqueurs de type microsatellite. Elles fournissent plusieurs paramètres génétiques utiles aux stratégies de conservation. Les résultats montrent l'intérêt de l'information moléculaire. Le coefficient de consanguinité global (F) et l'index de fixation (F_{IT}) sont modérés ($F = 0,027$; $F_{IT} = 0,056$). L'effectif efficace de fondateurs est petit ($n = 22$) et la population est divisée en trois groupes distincts ($F_{IT} = 0,078$). Le taux d'hétérozygotie moléculaire ($H_M = 71,2\%$) donne une image plus précise du taux réel de consanguinité. Ces paramètres associés à l'utilisation du concept de parenté moyenne permettent de définir les accouplements pour contrôler le taux de consanguinité et limiter les erreurs de paternité. © Inra/Elsevier, Paris

gestion génétique / paramètres démographiques / microsatellites / équins

1. INTRODUCTION

Small natural populations raise several problems when faced with their conservation: they have lost most of their economic value, they usually show a high inbreeding level which threatens their long term maintenance, and the conservation of the biodiversity they represent makes unsuitable the introduction of individuals of other populations. On these grounds, genetic variation with the goal of its maintenance is the first point to examine for conservation of a small population.

The use of genetic information based on microsatellite variation is based on the assumption that the level of variation detected at marker loci directly reflects the level of variation that influences future adaptation. The addition of the demographic history information (e.g. inbreeding, effective population size and population subdivision) contributes to the knowledge of a population for conservation purposes [16].

The Asturcon is a pony breed of the Asturias region in the north of Spain. Animals of this breed are elipometric with a black coat in different tones, long hair and an average height of 1.22 m. This breed was brought by the Celtic populations who colonised Asturias in the VIII century BC, and has been used in the last centuries mainly as a military horse, and as a work animal. Both activities have been abandoned because of their evident lack of interest nowadays. The Asturcon pony is used today as a riding horse due to its gentleness and to its particular amblegait ('ambladura', that is, both legs of the same side are extended together at the same time) making it a very comfortable animal to ride. After going through a major bottleneck at the beginning of this century, the population has stabilised, although the breed is still threatened. There are now 451 individuals, with a studbook started in 1981, and there is a need for a breeding program to provide a better management of the population dynamics.

In this paper, we make inferences about genetic diversity parameters using two sources of information on the Asturcon pony breed: pedigree studbook information and allele frequency distributions at ten microsatellite loci, and we propose mating strategies based on a parameter called average relatedness in an attempt to reduce the increase in inbreeding over time, with a goal of managing the future genetic diversity of a small population.

2. MATERIALS AND METHODS

2.1. Analysis of the studbook information

The pedigree completeness level was computed taking all the ancestors known per generation. Ancestors with no known parent were considered as founders (generation 0) and the number of known generations was computed as those separating the offspring of its furthest known ancestor in either path. Malécot [14] defined the coefficient of coancestry between two animals as the probability that a randomly chosen allele in one individual is identical by descent to a randomly chosen allele at the same locus in the other. Average relatedness (AR) could be defined as twice the probability that two random alleles, one from the animal and the other from the population in the pedigree (including the animal), are identical by descent and can then be interpreted as the representation of the animal in the whole pedigree regardless of the knowledge of its own pedigree. A vector containing the AR coefficients for all animals in a pedigree can be obtained by $\mathbf{c}' = (1/n)\mathbf{1}'\mathbf{A}$, where \mathbf{c}' is a row vector where c_i is the average of the coefficients in the row of individual i in the numerator relationship matrix, \mathbf{A} , of dimension n . In founder individuals, AR can be obtained assigning to each individual a value of 1 for its belonging to the population, 1/2 for each offspring the animal has in this population, 1/4 for each grandson and so on, and weighting by the size of the population, in such a way that AR will indicate its genetic contribution to the population.

The effective number of founders in a pedigree is defined as the number of individuals contributing equally to generate the population, given the unbalanced representation of the present number of founders. It was calculated as:

$$E_{nf} = \frac{1}{nb \sum_{i=1}^{nb} c_i^2}$$

where nb is the number of individuals in the founder population, given that AR in a founder individual explains the rate of population it contributes to. When a population is made up of an unequal contribution of founder animals, this parameter is very interesting since it could be increased if the chosen breeding animals are those with minimum AR values, regardless of any other parameter. Inbreeding coefficients (F) were computed for all animals [27]. As the population is divided into three subpopulations, inbreeding and AR were also computed for each group. The effective size per generation (N_e) is computed following Falconer and McKay [4] and is the inverse of twice the increase in inbreeding.

All inbreeding values [27] were computed (starting from zero in generation 0), assuming an ideal state of the population at generation 0. As this assumption is not met, molecular heterozygosity values (H_M) obtained with microsatellite loci at generation 0 were used, computing the heterozygosities in the later generations based on this initial value (H_P).

2.2. Animals

A total of 451 individuals (218 males and 233 females) were included in the studbook. Blood samples were collected from individuals belonging to different groups which compose the population: a sample of 25 individuals from the *founder population* ($n = 60$), 50 *random* sampled individuals (25 males and 25 females), and, according to geographic criteria, a sample of 40 individuals from the *Borines* subpopulation ($n = 82$), 18 individuals from the *LaVita* subpopulation ($n = 60$) and a sample of 60 individuals from the *Icona* subpopulation ($n = 114$) were taken to complete the sampling of the entire population which had 451 individuals included in the studbook at the time of the study in 1996.

2.3. Microsatellite amplification

DNA was extracted according to standard procedures. Ten equine microsatellites were chosen according to the ISAG (Comparison Tests, 1996): HTG4 and HTG6 [3], HTG8 and HTG10 [15], VHL20 [24] and HMS2, HMS3, HMS6, HMS7 [8], ASB2 (GenBank Accession no. X93516) were amplified using the polymerase chain reaction [19]. PCR products were separated by electrophoresis in 8 % polyacrylamide gels under denaturing conditions, followed by silver staining according to the procedure of Bassam et al. [1].

2.4. Analysis of microsatellite polymorphism

Microsatellite data were analysed using the BIOSYS-1 computer package [21] and F-Statistics (F_{IS} , F_{IT} , F_{ST} ; Wright [28]) were computed using the FSTAT version 1.2 computer program [7] which computes Weir and Cockerham [26] estimators. Permutations were used to test the significance of fixation indices over all loci and their confidence intervals were computed by bootstrapping [25]. Heterogeneity of allelic frequencies among subpopulations was tested using a chi-square test for each locus independently. To test the deviation of frequencies from Hardy-Weinberg equilibrium, the usual Chi-square test was performed using observed genotype frequencies and those expected under H-W equilibrium. The molecular heterozygosity (H_M) was computed per generation using all individuals (with blood samples available) identified in the studbook.

3. RESULTS AND DISCUSSION

The information generated from the Asturcon pony population originates from two sources: genetic parameters from the studbook which has incomplete pedigrees, and those derived from the use of molecular markers.

The first block of information has been analysed to compute inbreeding values (overall and by subpopulations), number of known generations and effective number of founders and of parents per generation (*tables I and II*). The second block of information is used to compute the proportion of heterozygotes present in the population as well as the existence of population structuring.

Although the overall inbreeding mean value is low ($F = 2.7\%$; *table I*) when only animals leaving offspring and with more than one known generation

Table I. Inbreeding coefficient by subpopulations (F^*) (in %), mean known generations and average relatedness (AR) for Asturcon Pony subpopulations. In parenthesis are the values found after parentage corrections.

| Subpopulation | N | Generations | F^* | AR |
|-----------------|-----|-------------|------------|------------|
| Icona | 114 | 2.68 (2.63) | 6.1 (5.3) | 10.9 (9.8) |
| Borines | 82 | 1.51 (1.43) | 0.8 (0.61) | 2.6 (2.3) |
| La Vita | 60 | 1.20 (1.20) | 1.3 (1.3) | 3.3 (3.3) |
| Rest of Animals | 195 | 2.86 (2.85) | 1.9 (1.8) | 7.4 (6.9) |
| | 451 | 2.35 (2.31) | 2.7 (2.4) | 6.8 (6.3) |

Table II. Inbreeding coefficient (F) (in %), average relatedness (AR) and effective population size (N_e) per generation, expected heterozygosity (H_P) and molecular heterozygosity (H_M).

| Generation | N | F | AR | N_e | n | H_P | H_M |
|------------|-----|-----------|------|-------------|-----|-------------|--------------------|
| 0 | 60 | 0 | 1.4 | – | 25 | 76.4 | 76.4 (± 3.5) |
| 1 | 56 | 0 | 3.7 | – | 13 | 76.4 | 76.5 (± 4.3) |
| 2 | 92 | 1.9 (1.8) | 6.2 | 26.3 (28) | 37 | 74.9 (75) | 73.1 (± 4.8) |
| 3 | 160 | 3.6 (3.2) | 8.8 | 29.4 (35.6) | 36 | 73.6 (74) | 68.8 (± 3.5) |
| 4 | 76 | 5.3 (4.9) | 9.8 | 27.9 (28) | 15 | 72.3 (72.7) | 59.3 (± 6.7) |
| 5 | 7 | 7.6 (7.1) | 10.5 | 20.7 (21.2) | 1 | 70.6 (71) | – |
| | 451 | 2.7 (2.4) | 6.8 | | 126 | 74.3 (74.2) | 71.2 (± 3.0) |

N refers to studbook sample size and n to blood sample size. Molecular heterozygosity (H_M) with standard error is computed for the n individuals in each generation. H_P is the expected heterozygosity when values in generations 1 and 2 (these coefficients of inbreeding were assumed to be zero) start with the molecular heterozygosity computed with microsatellites. In parenthesis are the values resulting after parentage correcting.

are considered, this value increases (circa 10 %), and is critically high when compared with other populations, e.g. 3 % in the Arab [18], 6 % in the Italian Haflinger [5] or the Norwegian Standardbred [10], 8 % in the Spanish breed [9]. The value of inbreeding by subpopulations (F^* , table I) is very high for Icona. F_{IS} (table III) is the average within-population inbreeding coefficient (measuring the extent of non-random mating) and gives values not different from 0, which means that no appreciable inbreeding is present in the subpopulations. This result is contradictory to that found when using studbook information: this means that molecular markers fail to detect the inbreeding level of the subpopulations in this case. Although we corrected the parentages computed in the studbook using molecular typing (finding nearly 10 % incorrect paternities which result in a lowering of the inbreeding level – $F_{ICONA}^* = 5.3\%$ – table I), the rate of inbreeding still remains relatively high.

Inbreeding increases the number of homozygotes and whenever no other factor modifies their expected frequency (all loci but HTG10 for Icona subpopulation were consistent with Hardy-Weinberg proportions), F_{IT} is a good indicator of the inbreeding coefficient of the global population [28]. An excess of homozygotes of 5.6 % seems to be in agreement with the inbreeding estimation of F .

Table III. F statistics (F_{ST} , F_{IS} , F_{IT}) estimated by subpopulation for the Asturcon Pony breed.

| Subpopulations | F_{ST} | | | F_{IS} | F_{IT} |
|----------------|----------|----------------|--------|-----------------|----------------|
| | ICONA | BORINES | LAVITA | | |
| ICONA | – | 0.095* | 0.066* | – 0.047 | |
| BORINES | | – | 0.050* | – 0.018 | |
| LAVITA | | | – | 0.031 | |
| TOTAL | | 0.078* ± 0.012 | | – 0.024 ± 0.018 | 0.056* ± 0.023 |

* Significant; $P < 0.01$.

It would be wise to point out that founders are assumed genetically unrelated, so inbreeding during the first generations is underestimated, leading to smaller values than in a representative sample of the population. To overcome this gap, we replaced the population heterozygosity (H_P) (table II) at generation 0 with the molecular heterozygosity (H_M) obtained with molecular marker information, expecting to take into consideration the relationship of the founders. However, H_P decreases over generations slower than H_M (table II) which was expected as this approach does not completely avoid the problem. In most population studies (e.g. [12, 17, 29]) sampling is based on unrelated individuals (or is not even mentioned) but when the goal of a study is the estimation of genetic parameters, random sampling should give unbiased estimates of these parameters in a population under study. We sampled 50 individuals on a random basis (25 from each sex) and the results ($H_M = 71.1 \pm 4.2$) allow us to infer that, in the case of the Asturcon pony population, the molecular heterozygosity of a random sample should give an accurate vision of the real inbreeding of a population for genetic management purposes.

Molecular marker information can also be used to analyse the distribution of genetic variability within and between subpopulations, allowing us to check the existence of geographical structures. The calculation of F_{ST} detects that nearly 8 % of the total genetic variability in the Asturcon is due to population differences (table III) possibly caused by different mating or selection strategies within the three subpopulations. Such an inference is reasonable since rates of gene flow ($N_e m$: effective number of individual exchange between populations per generation [22]) found between those populations are great enough (> 1) to attenuate the genetic differentiation between subpopulations by genetic drift. Pairwise F_{ST} values as well as heterogeneity of allele frequencies (data not shown) indicate a significant level of genetic differentiation between all subpopulations, but mostly between Icona and Borines whose members show a strong and significant divergence of circa 10 % (table III). This suggests that geographically separate populations are both demographically and genetically distinct.

Whatever the source of information used, genetic variability depends on the founder population size and a natural wastage of genetic material occurs as a

result of unequal founder contributions. Effective number of founders is small (22) relative to the actual number of founders present in the studbook (60) indicating the excessive use of some individuals as parents. It should be noted that after parentage verification this number increases to 24. Subdivision exists in this population (as F_{ST} values show above), and is a result of the mating of animals within subpopulations producing an increase in the inbreeding coefficients which can be lowered using a particular mating policy. For example, restricted matings obtained by linear programming [23] minimise the average coancestry coefficients but only in the first generation, having a negative effect in those following. The probability of gene origin [11] or founder equivalent [13, 20] is useful to describe a population structure after a small number of generations in order to characterise a breeding policy or to detect recent changes in the breeding strategy. Boichard et al. [2] have recently defined an effective number of ancestors accounting for the potential bottlenecks that could have occurred in the pedigree. All these concepts are based on a population under study, which are useful basically for description purposes. The effective number of founders in a pedigree defined in the present paper is equivalent to that of Rochambeau et al. [20] and Lacy [13] if all the animals in a pedigree were included in the present population. We proposed to the Breeder Association (ACPRA) the use of AR (see *tables I and II*) as a good criterion to maintain the genetic variability by maintaining the balance of the representation of the founder ancestors using the whole pedigree and not only the present population, permitting us to identify and use animals with the lowest AR coefficient, while describing the situation of the population and making use of all the potential genetic stock. Following this concept, a less represented animal (smaller AR value) will be preferred as parent for the next generation, resulting in a better maintenance of genetic variability and thus lower inbreeding coefficients in the long term. That means that all individual contributions in the population can be balanced using this coefficient and this allows the animal breeders to make matings in such a way as to preserve the genetic variability of the population. In practice, after the expected progeny size of the next generation is established, the average relatedness coefficients of all individuals are recomputed assuming an offspring resulting from the mating of the two lower AR (stallion and mare). This step is repeated until the progeny number is reached. The parents chosen during this process are then mated following the minimum coancestry strategy. Thus, the effective number of founders will grow, this increase in inbreeding will be minimised in the short and in the long term and as a result the initial genic diversity is conserved. Nevertheless, other reasons justify the use of average relatedness: this coefficient can also be used to define the influence of each founder animal in the whole population; mean subpopulation AR values show the degree of inbreeding and coancestry in each subpopulation when considered as a component of the whole population and if relatively high AR values are found, the introduction of new individuals is then indicated, although most matings occur within the subpopulation. ACPRA is at the moment using a program where every individual contribution can be controlled (Gutiérrez, pers. comm.) and mating recommendations are made to the breeders.

This study contributes as a first approach to the practical understanding of the genetic management of a small semi-feral population. The use of the incomplete herdbook data is optimised with the calculation of the AR value of

each individual for mating purposes. The information provided by the molecular markers also has other advantages. DNA microsatellites are efficiently used to determine incorrect paternity attribution which can be very high (e.g. 4–23 % of misidentification in German milk cattle, Gelderman et al. [6]; 9.6 % in this study).

In the special case of the Asturcon pony, all individuals born in the last 3 years are checked by genotyping giving the possibility of obtaining population information but also to contrast the parentages involved, changing the computed values (see *tables I and II*, values in parenthesis). Moreover, molecular marker information gives us a good idea of a population structure enabling the breeders association to better understand and manage the relationships between subpopulations. As a third advantage, we have seen above that the level of heterozygotes measured in the population as a whole (F_{IT}) can eventually allow us to compute the population inbreeding, which means that in those populations where pedigree information is not available, the use of molecular information based on an adequate sampling procedure should lead to the same conclusions.

ACKNOWLEDGEMENTS

The financial support of the Comisión Interministerial de Ciencia y Tecnología (CICYT): (Grant no. AGF95-064), of ACPRA (Asociación de Criadores de Ponis de Raza Asturcón) and Caja Asturias is greatly acknowledged. We are indebted to J. Martínez for personally providing the blood samples.

REFERENCES

- [1] Bassam B.J., Caetano-Anollés G., Gresshoff P.M., Fast and sensitive silver staining of DNA in polyacrylamide gels, *Analyt. Biochem.* 80 (1991) 81–84.
- [2] Boichard D., Maignel L., Verrier E., The value of using probabilities of gene origin to measure genetic variability in a population, *Genet. Sel. Evol.* 29 (1997) 5–23.
- [3] Ellegren H., Johansson M., Sandberg K., Andersson L., Cloning of highly polymorphic microsatellites in the horse, *Anim. Genet.* 23 (1992) 133–142.
- [4] Falconer D.S., McKay T.F.C., *Introduction to Quantitative Genetics*, 4th ed., Longman, Essex, 1996.
- [5] Gandini G.C., Bagnato A., Miglior F., Pagnacco G., Inbreeding in the Italian Haflinger horse, *J. Anim. Breed. Genet.* 109 (1992) 433–443.
- [6] Gelderman H., Pieper U., Weber W.E., Effect of misidentification on the estimation of breeding value and heritability in cattle, *J. Anim. Sci.* 63 (1986) 1759.
- [7] Goudet J., FSTAT (version 1.2): A Computer program to calculate F-Statistics, *J. Hered.* 86 (1995) 485–486.
- [8] Guérin G., Bertaud M., Amigues Y., Characterization of seven new horse microsatellites, *Anim. Genet.* 25 (1994) 62.
- [9] Gutiérrez J.P., Cañón J., Reducción de la variabilidad genética en el caballo de pura raza español, in: *Equino: Aspectos de Cría y Clínica*, Ciencias Veterinarias Volume XVIII, 1997, pp. 43–59.
- [10] Klemetsdal G., Demographic parameters and inbreeding in the Norwegian Trotter, *Acta Agric. Scand.* 43 (1993) 1–8.
- [11] James J.W., Computation of genetic contributions from pedigrees, *Theor. Appl. Genet.* 42 (1972) 272–273.

- [12] Jordana J., Piedrafita J., Sánchez A., Genetic relationships in Spanish dog breeds. II. The analysis of biochemical polymorphism, *Genet. Sel. Evol.* 24 (1992) 245–263.
- [13] Lacy L.C., Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalents, *Zool. Biol.* 8 (1989) 111–123.
- [14] Malécot G., *Les mathématiques de l'hérédité*, Masson, Paris, 1948.
- [15] Marklund S., Ellegren H., Eriksson S., Sandberg K., Andersson L., Parentage testing and linkage analysis in the horse using a set of highly polymorphic microsatellites, *Anim. Genet.* 25 (1994) 19–23.
- [16] Milligan B.G., Leebens-Mack J., Strand A.E., Conservation genetics: beyond the maintenance of marker diversity, *Mol. Ecol.* 3 (1994) 423–435.
- [17] Moazami-Goudarzi K., Laloë D., Furet J.P., Grosclaude F., Analysis of genetic relationships between 10 cattle breeds with 17 microsatellites, *Anim. Genet.* 28 (1997) 338–345.
- [18] Moureaux S., Verrier E., Ricard A., Mériaux J.C., Genetic variability within French race and riding horse breeds from genealogical data and blood marker polymorphisms, *Genet. Sel. Evol.* 28 (1996) 83–102.
- [19] Mullis K.B., Falbona F., Sharf S.J., Saiki R.K., Horn G.T., Erlich H.A., Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction, *Cold Spring Harbor Symp. Quant. Biol.* 51 (1986) 263–273.
- [20] Rochambeau H. de, La Fuente L.F. de, Rouvier R., Ouhayoun J., Sélection sur la vitesse de croissance post-sevrage chez le lapin, *Genet. Sel. Evol.* 21 (1989) 527–546.
- [21] Swofford D.L., Selander R.B., BIOSYS-1 User's Manual, Release 1.7. A computer program for the analysis of allelic variation in population genetics and biochemical systematics, Illinois Natural History Survey, Champaign, Illinois, 1989.
- [22] Takahata N., Gene identity and genetic differentiation of populations in the finite island model, *Genetics* 104 (1983) 497–512.
- [23] Toro M.A., Nieto B., Salgado C., A note on minimization of inbreeding in small-scale selection programs, *Livest. Prod. Sci.* 20 (1988) 317–323.
- [24] Van Haeringen H., Bowling A.T., Stott M.L., Lenstra J.A., Zwaagstra K.A., A highly polymorphic horse microsatellite locus VHL20, *Anim. Genet.* 25 (1994) 207.
- [25] Weir B.S., *Genetic Data Analysis II*, Sinauer, Sunderland, Massachusetts, 1996.
- [26] Weir B.S., Cockerham C.C., Estimating F-statistics for the analysis of population structure, *Evolution* 38 (1984) 1358–1370.
- [27] Wright S., Coefficients of inbreeding and relationship, *Am. Nat.* 56 (1922) 330–338.
- [28] Wright S., *Evolution and the Genetics of Populations*, Vol. IV. Variability within and among Natural Populations, University of Chicago Press, Chicago, 1969.
- [29] Zajc Y., Mellersh C., Sampson J., Variability of canine microsatellites within and between different dog breeds, *Mamm. Gen.* 8 (1997) 182–185.