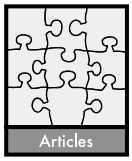


J. Avian Biol. 41: 34–49, 2010
doi: 10.1111/j.1600-048X.2009.04730.x

Received 29 September 2008, accepted 9 March 2009

Microsatellites and single nucleotide polymorphisms in avian hybrid identification: a comparative case study

Ülo Väli, Pauli Saag, Valery Dombrovski, Bernd-Ulrich Meyburg, Grzegorz Maciorowski, Tadeusz Mizera, Rimgaudas Treinys and Sofie Fagerberg

Ü. Väli, (correspondence) Dept. of Evol. Biol., Evol. Biol. Centre, Uppsala Univ., Norbyvägen 18D, SE-75236 Uppsala, Sweden, and Inst. of Agricult. and Environm. Sci., Estonian Univ. of Life Sciences, Riia 181, EE-51014 Tartu, Estonia. E-mail: ulo.vali@emu.ee. – P. Saag, Inst. of Zool. and Hydrobiol., Univ. of Tartu, Vanemuise 46, 51014 Tartu, Estonia. – V. Dombrovski, Inst. of Zool. NAN B, Akademichnaya 27, 220072 Minsk, Belarus. – B.-U. Meyburg, World Working Grp. of Birds of Prey, Wangenheimstraße 32, D-14193 Berlin, Germany. – G. Maciorowski and T. Mizera, Zool. Dept., Poznan Univ. of Life Sci., Wojska Polskiego 71 c, PL 60-625 Poznan, Poland. – R. Treinys, Lab. of Avian Ecol., Inst. of Ecol., Vilnius University, Akademijos 2, LT-08412 Vilnius, Lithuania. – S. Fagerberg, Dept. of Evol. Biol., Evol. Biol. Centre, Uppsala Univ., Norbyvägen 18D, SE-75236 Uppsala, Sweden.

The correct identification of hybrids is essential in avian hybridisation studies, but selection of the appropriate set of genetic markers for this purpose is at times complicated. Microsatellites and single nucleotide polymorphisms (SNPs) are currently the most commonly used markers in this field. We compare the efficiency of these two marker types, and their combination, in the identification of the threatened avian species, the greater spotted eagle and the lesser spotted eagle, as well as hybrids between the two species. We developed novel SNP markers from genome-wide distributed 122 candidate introns using only sympatric samples, and tested these markers successfully in 60 sympatric and allopatric spotted eagles using Bayesian model-based approaches. Comparatively, only one out of twelve previously described avian nuclear intron markers showed significant species-specific allele frequency difference, thus stressing the importance of selecting the proper markers. Twenty microsatellites outperformed selected nine SNPs in species identification, but were poorer in hybrid detection, whereas the resolution power of ten microsatellites remained too low for correct assignment. A combination of SNPs and microsatellites resulted in the most efficient and accurate identification of all individuals. Our study shows that the use of various sets of markers could lead to strikingly different assignment results, hybridisation studies may have been affected by too low a resolution power of used markers, and that an appropriate set of markers is essential for successful hybrid identification.

Interspecific hybridisation is widespread among birds. Only two decades ago, it was believed that almost one tenth of all bird species interbreed in nature and produce hybrid offspring (Panov 1989, Grant and Grant 1992), the estimated proportion today is even higher (McCarthy 2006). Studies of interspecific hybridisation are therefore a highly important component of avian research where they enable various aspects to be understood. Hybridisation studies provide new insights into species' biology and help in our understanding of evolutionary processes, particularly speciation. Hybridisation itself is known to be source of new material for evolution; the existence of hybrids and introgression raises questions about the validity of previously defined taxa (Barton and Hewitt 1989, Arnold 1992, Grant and Grant 1994, Dowling and Secor 1997, Grant and Grant 1997). Identification of hybrids is the first step in all hybridisation studies. This is often based simply on checking morphological variables considered to have specific values in both parent species and which show intermediate values in hybrids. However, not all hybrids can be identified by morphological investigation and the probability of identifying backcrosses, which may resemble F1 hybrids or parental species, is lower still, if not impossible (Allendorf et al. 2001, Gaubert et al. 2005, Chan et al. 2006). Moreover, morphological analysis requires usually the trapping of individuals, which is not appropriate in studies of threatened species. Genetic methods, on the other hand, can be performed non-invasively for example by using moulted feathers (Lõhmus and Väli 2004, Rudnick et al. 2005), and relevant marker sets give reliable identification of hybrids and backcrosses.

Although many potential molecular methods are available for hybridisation studies, easily interpretable co-dominant markers are preferred nowadays. Microsatellite genotyping was probably the most popular method over the past two decades, and it is usually still the first choice of researchers (Avise 2004, Schlötterer 2004). The success of these markers is based on their high polymorphism rate, which leads to allele frequency differences between species and thus to recognition of hybrids. During the last decade

however, the popularity of single nucleotide polymorphisms (SNPs) has increased remarkably (Morin et al. 2004). In the simplest case, a single SNP with two fixed alleles is sufficient to assign individuals to a species, as well as to identify F1 hybrids (Bureš et al. 2002, Primmer et al. 2002). This is possible in cases where either two species have been separated for a long period of time, or hybridisation has begun only recently and the introgression rate is low. However, when species have diverged only recently, or introgression has existed over a longer period, fixed differences are more difficult to identify and reliance should be made on simultaneous analysis of a large number of markers with species-specific allele frequencies (Allendorf et al. 2001, Randi et al. 2001, Vilà et al. 2003, Lecis et al. 2006, Verardi et al. 2006, Nittinger et al. 2007). A larger numbers of markers is also needed for determination of backcrosses and later-generation hybrids (Boecklen and Howard 1997, Allendorf et al. 2001, Vähä and Primmer 2006). There are several assignment techniques now available, and these methods identify hybrids efficiently by using allele frequency information (Manel et al. 2005). Moreover, these approaches do not necessarily require data from allopatric reference populations (Vähä and Primmer 2006), but can rely on the analysis of mixed sympatric populations alone. This is good news for conservationists, as some rare species lack pure allopatric populations or it is not possible for researchers to have access to them. However, as it is generally considered that hybridisation surveys require the development of markers using allopatric populations, empirical studies confirming the theoretical redundancy of allopatric references are almost completely lacking.

The greater spotted eagle *Aquila clanga* and the lesser spotted eagle *A. pomarina* are two palaearctic raptor species of conservation concern that, according to estimates using mitochondrial sequences, diverged about one million years ago (Seibold et al. 1996). The two species differ in several aspects of their morphology and ecology, but there is still no complete reproductive barrier between the species. Several species-specific morphological characters in spotted eagles have been described, but all of them vary to a great extent within both species (Bergmanis 1996, Forsman 1999, Dombrovski 2006, Dombrovski and Demongin 2006), and variation among hybrids is anyway hard to predict. Although a multivariable morphological analysis, using several characteristics simultaneously, permits a reasonably effective identification of the two species and their hybrids (Väli and Lõhmus 2004), some bias may be introduced due to geographical variability across regions and differences in investigation techniques among researchers, and morphological characteristics are probably not sufficient to identify backcrosses.

There are two well-segregated mitochondrial DNA lineages among spotted eagles (Seibold et al. 1996, Väli 2002, Helbig et al. 2005b), and this species-specific genetic marker has been used as an additional variable in hybrid identification (Väli and Lõhmus 2004, Helbig et al. 2005b). However, as the analysis of mitochondrial DNA only provides information on the maternal lineage, nuclear markers must also be used. Of the nuclear markers available, only amplified fragment length polymorphism (AFLP) markers have been applied so far to the study of genetic differentiation and gene flow in spotted eagles (Helbig et al. 2005b). AFLPs are a good choice for the analysis of species with limited genetic information, and a large amount of data can be obtained with relatively little expense of time and effort (Mueller and Wolfenbarger 1999, Bensch and Åkesson 2005). Nevertheless several drawbacks, such as dominance and difficulties in interpretation, limit the use of this method (Avise 2004). Moreover, although AFLPs reveal the differentiation between the two spotted eagle species, it appears that the resolution power of AFLPs is still inadequate for reliable assignment of individuals, since many birds are impossible to assign unequivocally to a species, or to identify as hybrids (see Figs in Helbig et al. 2005b).

In the current study we: (1) searched for simple SNP markers that distinguish spotted eagle species and which can be used for hybrid identification, (2) studied the performance of different type and number of genetic markers by comparing their assignment efficiency (i.e. the proportion of individuals in a group that were correctly identified) and accuracy (the proportion of an identified group that truly belong to that category), and (3) tested whether sympatric samples are adequate for development of markers in a hybridisation study. We first tested twelve nuclear intron markers that have been used previously for taxonomic and hybridisation studies in many avian species. As these markers appeared to be ineffective for spotted eagles, we tested a large number of genome-wide distributed intron markers, developed recently by Backström et al. (2008), in a small sympatric sample of spotted eagles. We selected the small set of unlinked SNPs that best distinguish one species from another, and tested these markers in a larger sample collected from both sympatric and allopatric individuals. Subsequently we applied different numbers of microsatellite markers to assign the same individuals: ten microsatellites, reflecting the number of markers commonly used in similar studies to date, and twenty microsatellites, which supposedly increase the power of analysis. Finally, we checked whether the combination of microsatellites and SNPs developed by us produces the best result in hybrid identification.

Material and methods

Samples

Altogether we used 60 blood and feather samples, collected during field studies between 2001 and 2006. Fifty three (21 *A. clanga*, 16 *A. pomarina*, 16 putative hybrids) were obtained from those parts of the distribution ranges in Europe where both species have been recorded as breeders (Estonia, Lithuania, Germany, Poland, Belarus, European Russia (Upper Volga region)), and seven from strictly allopatric populations (six *A. pomarina* from Greece, Romania and Slovakia, and one *A. clanga* from Asian Russia, near Lake Baikal; see Appendix 1). Sympatric samples were collected from nestlings or from trapped adults, which were always examined in the hand, whereas allopatric samples were obtained from feathers moulted by breeding adults. Eagles are long-lived birds and occupy the same nest in consecutive years, therefore only one sample

from each nest site was included in the analysis (only on one occasion did we include an adult female in the *A. clanga* and its nestling in the hybrid group). Prior to any genetic analysis, each bird was identified by methodical in hand inspection using a comprehensive complex of morphological characteristics (Väli and Lõhmus 2004). The field identification of parent birds (according to the characters described by Forsman 1999) was also taken into account, as well as mitochondrial pseudo-control region sequence that differs by some 4% between species (Väli 2002). None of the *A. clanga* or *A. pomarina* individuals selected had a mitochondrial DNA typical of the other species (Appendix 1). Only individuals morphologically assigned to a species, or identified as a F1 hybrid, were included, whereas all those with limited morphological data, as well as probable backcrosses, were excluded. However, we cannot completely rule out that our sample includes some backcrosses since such nestlings, as well as adult F1 hybrids, are difficult to identify. In later genetic analysis no morphological pre-assignment information was used.

Selection of single nucleotide markers

We used initially 12 published primer sequences, β-fibrinogene (FIB) intron 7 (Prychitko and Moore 1997), B-creatine kinase (BC-K), tumour growth factor β (TGF) intron 2 (Bureš et al. 2002), amino levulinate synthetase (ALASY) intron 8, neural glucoprotein CEPUS intron 1, Lamin A intron 3 (LAMA), lamine receptor precursor protein 40 (LRPP40) intron 5, ornitine decarboxylase (ODC) introns 6 and 7, rhodopsine (RDPSN) intron 1 (Primmer et al. 2002), lactate dehydrogenase (LDH3) intron 3, adenylate kinase (AK) intron 5 and recombination activating gene RAG-1 (Helbig et al. 2005a) to amplify and sequence these markers in 5–20 individuals from both species. We detected no SNPs that were fixed within species among the published intron markers, and a SNP with allele frequency difference greater than 0.5 (threshold value used also by Primmer et al. 2002) was found only in FIB (0.55), while lower differentiation was recorded among those found from the LDH (0.42), AK-1 (0.25) or RAG-1 (0.20). In BC-K, LAMA, LRPP and ODC no potential SNPs were detected, and from TGFβ2, ALASY-8, CEPUS and RDPSN no proper amplification products, or sequences, were obtained. Subsequently we retained only FIB marker for the further analysis.

Second, we used 122 primer pairs developed and tested in several species by Backström et al. (2008) in order to amplify introns (with a length of 350–950 bp) distributed across the avian genome and separated from each other by some 10 Mb. Amplification of all introns was first tested in a high-quality *A. clanga* sample. Testing of 122 loci resulted in 87 single-band amplification products. In other cases we recorded either several products or no DNA-band at all. Forty eight strongest single-band PCR products were sequenced in eleven individuals (5 *A. clanga*, 5 *A. pomarina*, 1 F1 hybrid) and altogether 52 SNPs present in more than one individual (i.e. at least in 18% of samples used) were detected in 32 loci. The SNPs found were ranked according to differences in allele frequencies between two species (hybrid individuals were excluded). Sixteen SNPs from 12 loci showed an interspecific difference 0.5 or higher. We chose nine SNPs for further analysis: two SNPs (from loci 5/17367 and 6/13446) with the highest interspecific allele frequency difference, and seven SNPs having a unique restriction site in one of two detected alleles (Table 1, Appendix 2). We also added the SNP from the FIB locus (chromosome 4 in the chicken genome) from the first approach. The ten markers selected were analysed in 60 individuals but, as the selected SNP in locus 6/13446 appeared to have very low allele frequency difference in a large sample, we excluded that marker and performed the final analysis with only nine SNP markers.

Selection of microsatellites

We used microsatellites developed previously from three other raptor species and used successfully in cross-species amplification in golden eagles *Aquila chrysaetos* (Burke and Dawson 2006). Thirty two microsatellites resulted in single-band PCR products in spotted eagles: 16 markers developed from Spanish imperial eagle *Aquila adalberti* (Aa02, Aa04, Aa11, Aa12, Aa15, Aa26, Aa27, Aa35, Aa39, Aa43, Aa49, Aa50, Aa51, Aa53, Aa56, Aa57; Martinez-Cruz et al. 2002), 6 from the eastern imperial eagle *A. heliaca* (IEAAAG04, IEAAAG11, IEAAAG12, IEAAAG13, IEAAAG14 and IEAAAG15) (Busch et al. 2005), and 10 from the white-tailed eagle *Haliaeetus albicilla* (Hal1, Hal2, Hal3, Hal4, Hal7, Hal9, Hal10, Hal13, Hal14 and Hal18) (Hailer et al. 2005). We excluded 8 monomorphic markers (Aa04, Aa11, Aa50, Aa56, Aa51, Hal2, Hal3, Hal14), as well as poorly interpretable Hal10 (which also shows linkage disequilibrium with Aa26; Burke and Dawson 2006), and IEAAAG11. To match the power of SNPs and microsatellites, we also excluded Aa49 and Hal9, which did not significantly segregate two species (test for significance with 999 permutations by GenAlEx (Peakall and Smouse 2006): $P > 0.05$; $n = 44$). The final set contained 20 microsatellite markers (Aa02, Aa26, Aa35, Aa39, Aa43, Aa12, Aa15, Aa27, Aa53, Aa57, IEAAAG04, IEAAAG12, IEAAAG13, IEAAAG14, IEAAAG15; Hal1, Hal4, Hal7, Hal13, Hal18) that had on average 6.0 (2–13) alleles in 44 spotted eagle individuals (Table 1). Hybridisation studies often use a much smaller number of microsatellites. In order to test a smaller set of microsatellites, we chose randomly ten of them simply by selecting every second marker from the list. The estimated differentiation between species F_{st} (Weir and Cockerham 1984) was similar either it was calculated by the 10 included markers ($F_{st} = 0.19$; 95%CI = 0.11–0.28), or by the 10 excluded markers ($F_{st} = 0.21$; 95%CI = 0.11–0.35). No allelic dropouts, null alleles or stuttering were detected by Micro-Checker 2.2.3 (Van Oosterhout et al. 2004) and, after applying Bonferroni correction, there was no indication of linkage disequilibrium within species. [you have changed a lot - e.g. where is Backstrom et al. 2008????)

Laboratory analyses

Amplification by polymerase chain reaction (PCR) was performed in 10 μl (microsatellite and restriction analysis) or 20 μl (sequencing) volume containing 25–50 ng DNA,

Table 1. Descriptive statistics for genetic markers used in the current study.

Locus[a]	Unit[b]	Size[c]	No. of alleles				Expected heterozygosity				Obs. heterozygosity				Fst[d]	Ref[e]
			Ac	Ap	Ac+Ap	hybrids	Ac	Ap	Ac+Ap	hybrids	Ac	Ap	Ac+Ap	hybrids	Ac-Ap	
1/26928	C/G	704 (562)	2	2	2	2	0.45	0.09	0.46	0.38	0.41	0.09	0.25	0.38	0.57	1
4/12303	A/G	913 (745)	2	2	2	2	0.20	0.20	0.50	0.50	0.23	0.23	0.23	0.69	0.74	1
4/β-FIB	A/G	1089 (905)	2	2	2	2	0.30	0.38	0.50	0.47	0.26	0.39	0.32	0.50	0.47	2
5/15691	A/G	669 (520)	2	2	2	2	0.17	0.04	0.50	0.49	0.18	0.05	0.11	0.75	0.88	1
5/17367	A/G	1127 (951)	2	2	2	2	0.24	0.06	0.50	0.49	0.17	0.06	0.11	0.71	0.81	1
7/04557	C/G	1124 (960)	2	2	2	2	0.38	0.49	0.48	0.38	0.14	0.41	0.27	0.38	0.16	1
8/17388	A/G	704 (603)	2	2	2	2	0.38	0.40	0.50	0.50	0.41	0.27	0.34	0.69	0.35	1
13/12260	A/G	1224 (1033)	2	2	2	2	0.17	0.18	0.50	0.50	0.19	0.20	0.20	0.73	0.78	1
17/14657	A/G	774 (631)	2	2	2	2	0.24	0.29	0.50	0.50	0.18	0.35	0.26	0.75	0.64	1
SNPs total			18	18	18	18	0.28	0.24	0.49	0.47	0.24	0.23	0.23	0.62	0.64	
Aa02	2	131–135	2	2	3	2	0.35	0.13	0.25	0.22	0.15	0.14	0.14	0.25	0.12	3
Aa12	2	125–133	4	3	4	3	0.18	0.54	0.61	0.65	0.19	0.24	0.21	0.73	0.61	3
Aa15	2	197–209	5	6	6	6	0.64	0.60	0.75	0.78	0.59	0.52	0.56	0.88	0.28	3
Aa26	2	146–168	7	6	10	6	0.67	0.68	0.70	0.69	0.77	0.64	0.70	0.88	0.05	3
Aa27	2	93–99	3	3	3	3	0.39	0.64	0.55	0.65	0.48	0.71	0.58	0.71	0.14	3
Aa35	2	249–271	9	8	10	6	0.74	0.53	0.76	0.74	0.77	0.52	0.65	1.00	0.26	3
Aa39	2	165–179	6	4	7	5	0.69	0.52	0.64	0.63	0.73	0.59	0.66	0.69	0.07	3
Aa43	2	110–136	8	9	12	7	0.80	0.76	0.84	0.76	0.85	0.73	0.79	0.75	0.11	3
Aa53	2	130–144	7	5	7	4	0.78	0.58	0.80	0.72	0.62	0.65	0.63	0.93	0.22	3
Aa57	2	112–120	4	2	4	3	0.38	0.06	0.25	0.42	0.27	0.06	0.18	0.40	0.08	3
Hal1	4	116–120	2	2	2	2	0.47	0.09	0.44	0.43	0.32	0.09	0.20	0.50	0.52	4
Hal4	2	148–180	10	9	13	10	0.79	0.84	0.86	0.87	0.71	0.82	0.77	0.94	0.08	4
Hal7	2	143–153	4	3	5	5	0.55	0.34	0.47	0.45	0.57	0.32	0.44	0.50	0.10	4
Hal13	2	150–154	3	3	3	3	0.39	0.41	0.54	0.53	0.35	0.43	0.39	0.56	0.40	4
Hal18	2	141–151	3	3	4	3	0.48	0.52	0.50	0.50	0.52	0.59	0.55	0.60	0.01	4
IEAAAG04	4	210–226	4	3	5	4	0.67	0.35	0.58	0.61	0.77	0.38	0.58	0.75	0.20	5
IEAAAG12	4	130–146	5	5	5	5	0.68	0.54	0.70	0.65	0.68	0.64	0.66	0.63	0.21	5
IEAAAG13	4	242–270	5	5	8	7	0.56	0.72	0.68	0.75	0.59	0.90	0.74	0.93	0.09	5
IEAAAG14	4	204–212	3	2	3	3	0.53	0.36	0.48	0.54	0.48	0.35	0.42	0.64	0.07	5
IEAAAG15	4	122–142	5	4	6	5	0.66	0.62	0.68	0.64	0.76	0.57	0.67	0.50	0.10	5
Microsatellites total		99	87	120	92	0.57	0.49	0.61	0.61	0.56	0.49	0.53	0.69	0.20		

[a] In SNPs, chromosome/locus.
[b] Substitutions in SNPs and repeat unit lengths in microsatellites.
[c] Amplified fragment length (intron length) in SNP markers and size range in microsatellites.
[d] F_{st} – differentiation between *A. clanga* and *A. pomarina*.
[e] References. 1-(Backström et al. 2008); 2-(Prychitko and Moore 1997); (Martinez-Cruz et al. 2002); 4-(Hailer et al. 2005); 5-(Busch et al. 2005).

0.25 U AmpliTaq Gold polymerase with 1 × Amplitaq Gold PCR buffer (Applied Biosystems), 2.5 mM $MgCl_2$, 0.5 μM of each primer and 0.2 mM dNTP. When the initial set of 122 markers was tested, the PCR profile included an initial heating at 95° C for 5 min, followed by 35 cycles of 95° C for 30 s, 60° C to 50° C for 30 s and 72° C for 1 min, and a final extension at 72° C for 10 min. During first ten cycles, an annealing temperature was decreased by 1° C for every cycle, whereas for the last 25 cycles 50° C was used. In the later analysis of nine selected primers, we used a constant annealing temperature (Appendix 2), HotStar Taq polymerase with appropriate buffer (Qiagen) and initial heating for 15 min. It was possible to use standard Restriction Fragment Length Polymorphism analysis (RFLP) to genotype individuals for eight out of nine selected SNPs by incubating 10 μl of PCR product with 0.5 – 1U of suitable restrictase (Fermentas) overnight at 37°C. PCR and restriction products were analysed on 2% or 3% agarose gel.

Microsatellites were amplified by FAM-, HEX- and TET-labelled primers, and the resulting PCR products were diluted 25- to 100-fold before genotyping. PCR fragments for sequencing were purified by exonuclease I and shrimp alkaline phosphatase (USB) treatment at 37° C for 15 min, followed by denaturation at 80° C for 15 min. Sequencing was performed in both directions in 10 μl (5 μl of purified PCR product, 1 μl amplification primer and 4 μl of DYEnamic ET terminator sequencing reagent premix) using 29 cycles of the following profile: 95° C for 20 s, 50° C for 15 s and 60° C for 1 min; the sequencing products were purified in AutoSeq96 plates (both by Amersham Biosciences). Sequencing and microsatellite genotyping were performed using MegaBACE 1000 automated capillary sequencer (Amersham Biosciences) according to the manufacturer's recommendations. Microsatellite alleles were detected by Genetic Profiler 2.2. Sequences were checked and corrected using Sequencher 4.1.4 (Gene Codes Corp.) and aligned by ClustalW (Thompson et al. 1994). Polymorphisms were detected by visual inspection of sequences and always re-checked from chromatograms.

Data analysis

The number of alleles, as well as expected and observed heterozygosities, were calculated by GenAlEx 6 (Peakall and Smouse 2006) for both species separately and together, as well as for hybrids.

Clustering of all individuals was first checked by Fact. Correspond. Anal. (Belkhir et al. 2004), where allele composition of each individual is transformed into a point in multidimensional factorial space. Only the results of the two first factors, covering the largest proportion of variation, are presented in this study. The third axis was checked as well, but was not used as it did not provide any additional information.

Thereafter, individuals were assigned to groups using two Bayesian model-based MCMC-simulation approaches, whereas no morphological identification results were involved in assignments. We used the programmes Structure (Pritchard et al. 2000) and NewHybrids (Anderson and Thompson 2002), which both detect population structuring as linkage disequilibrium between unlinked loci and departures of allele frequencies from Hardy-Weinberg proportions. All MCMC simulations were run in three replicates, and assignment results differed from each other usually by some 0.1% only.

In "Structure", we first assessed the most probable number of different populations. The presumed number ranged from one to five, replicating situations from a single panmictic population to highly differentiated groups of species, F1 hybrids and backcrosses, or structuring between intraspecific geographical populations. According to the minimum plateau-reaching value of log-likelihood probabilities (Pritchard et al. 2000), support for two populations was always the highest and this logical presumption (K = 2) was used in assignment tests. As the occurrence of two species was presumed within our sample, we used the admixture model without preliminary information of species ancestry (to avoid any potential influencing of results), and the correlated allele frequency models assuming gene flow between species. Comparatively, all analyses were also run using the independent allele frequency model; but no significant differences between results were detected. The programme was always run for 500,000 iterations following 100,000 burn-in iterations. For each individual, the probability of belonging to one of two species q_s (denoting here probability q by Structure) was calculated. In order to assign an individual to a group, we used the error rate 0.1 from the expected value, which is most commonly used in hybridisation studies. This resulted in the following ranges (indicated as a probability of recent *A. clanga* ancestry): *A. pomarina* $q_s = 0$–0.1; backcross F1 × *A. pomarina* $q_s = 0.15$–0.35; F1 $q_s = 0.4$–0.6; backcross F1 × *A. clanga* $q_s = 0.65$–0.85; *A. clanga* $q_s = 0.9$–1.0. Individuals having probabilities intermediate between the target group and the adjacent groups were treated as unidentified according to this analysis.

NewHybrids is developed specifically for identification of hybrids and it is designed to assign each individual into a group such as pure species, F1 hybrids, later generation hybrids or backcrosses to parental species. We assumed that the occurrence of F2 or later generation hybrids and their backcrosses is very unlikely in the wild, since the meeting and mating of two hybrids is probably rare. As power analysis with simulated genotypes (see below) also showed that none of our marker sets is able to identify F2 hybrids correctly (data not shown), we included only five potential classes in our analyses – *A. clanga*, *A. pomarina*, F1 hybrids, backcross *F1* × *A. clanga*, backcross F1 × *A. pomarina*. We used NewHybrids' assignment posterior probability (q_n) threshold 0.5, meaning that individuals with at least one q_n value above 0.5 were assigned into that particular group, while those with all probabilities below that limit remained unidentified. Each run lasted at least 25,000 sweeps in burn-in and 100,000 sweeps in analysis step. Both uniform and Jeffreys-like priors were used to detect unreliable prior-dependent results (Anderson and Thompson 2002), but no great differences were obtained. We present here only assignment results obtained by uniform-priors.

In order to test the assignment power of markers used, we generated synthetic genotypes of F1 hybrids, as well as

backcrosses to both species (20 individuals in each class), by Hybridlab 1.0 (Nielsen et al. 2006) using the allele composition in our 44 field-collected samples of pure species as source data. All abovementioned analyses conducted with natural samples were repeated using 44 natural samples from two species and 60 synthetic genotypes (i.e. approximately the equal number of individuals in each class). Possible sample size effects were later reduced by comparative power tests using 15 hybrid samples (five individuals from each hybrid class). Only NewHybrids was used for power analysis in order to ease the interpretation of results. Q_n values are presented as average $\pm$ SD of individuals crossing the threshold 0.5.

Results

Marker performance

There were more microsatellite alleles in *A. clanga*, and less in *A. pomarina*, than in hybrids but differences were not large. However, expected heterozygosity was both in *A. clanga* and *A. pomarina* smaller than in hybrids or when the two species were combined; observed heterozygosity was much higher in hybrids than in the two species either separately or combined (Table 1).

Although the discriminative power was variable among markers (Table 1, Appendix 3), the two species were significantly differentiated according to 10 microsatellites ($F_{st} = 0.19$), 20 microsatellites ($F_{st} = 0.20$), nine SNPs ($F_{st} = 0.64$), or all markers combined ($F_{st} = 0.34$).

Hybrid identification by SNPs

The FCA plot separated the two species and hybrids from each other (Fig. 1C). Structure correctly assigned 73% of *A. pomarina*, 75% of hybrids and 45% of *A. clanga* individuals; 23% of *A. pomarina*, 18% of *A. clanga* and 19% of hybrids remained unidentified, while 5% of *A. pomarina*, 6% of hybrids and 37% of *A. clanga* were classified incorrectly, mostly as backcrosses (Fig. 2C). NewHybrids identified all *A. pomarina* and 86% of *A. clanga* individuals with three *A. clanga* individuals remaining unidentified (Fig. 3C). Power analysis with 60 simulated hybrids suggested that NewHybrids using nine SNPs could identify all F1 individuals, though with relatively low q_n values ($q_n = 0.76 \pm 0.11$ in correctly assigned individuals), but only 20% of *A. clanga* $\times$ F1 ($q_n = 0.54 \pm 0.03$) and 40% of *A. pomarina* $\times$ F1 ($q_n = 0.54 \pm 0.03$). When 15 simulated genotypes were analysed, all F1, two (40%) *A. clanga* $\times$ F1 but no A. *pomarina* $\times$ F1 was identified.

Hybrid identification by microsatellites

FCA plotted assigned 60 eagles into three slightly overlapping clusters: *A. clanga*, *A. pomarina* and hybrids (Fig. 1B). Structure assigned 95% of *A. pomarina* and 86% of *A. clanga*, but only 31% of hybrid individuals as expected (Fig. 2B). Five percent of *A. clanga* and 19% of hybrids remained unidentified, while 5% of *A. pomarina*, 9% of *A. clanga* and altogether 44% of hybrids were assigned as backcrosses. Results of NewHybrids were similar in respect of the two species, but only 37% of hybrids were identified as F1 (Fig. 3B). Other hybrids remained mostly unidentified, showing equal probabilities of F1 and backcross classes; but 12% of them were assigned to the *A. pomarina* group.

Ten selected microsatellites gave poorer assignment results. Using Structure, 9% of *A. clanga* and 9% of *A. pomarina* remained unidentified, whereas 18% of *A. clanga* and 18% of *A. pomarina* were misidentified as backcrosses (Fig. 2A). Altogether 63% of putative F1 hybrids were identified as backcrosses and 13% as *A. pomarina*. NewHybrids correctly assigned 86% of *A. pomarina*, but only 50% of *A. clanga*, whereas others were identified as backcrosses to respective species (Fig. 3A). None of the hybrids were assigned to the F1 group. Power

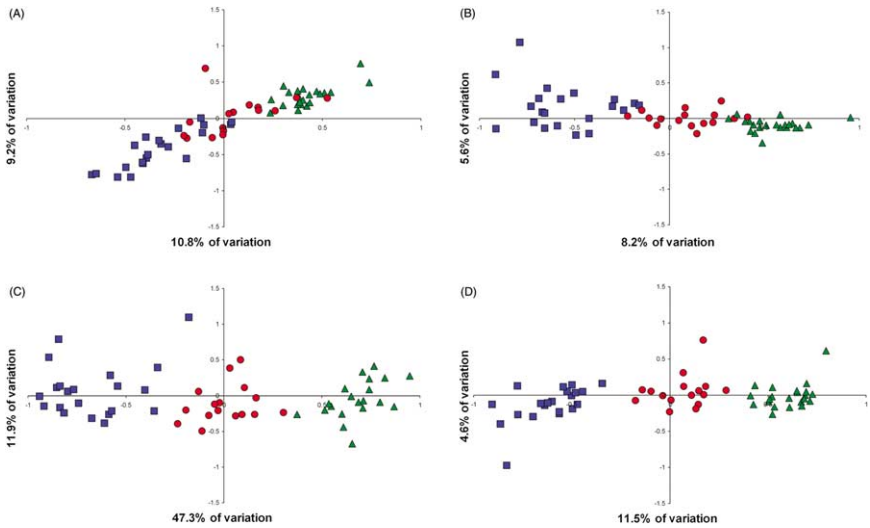


Figure 1. Distribution of the studied spotted eagle individuals in the two-dimensional space of factorial correspondence analysis, according to the allele composition of 10 microsatellites (A), 20 microsatellites (B), 9 SNPs (C), 9 SNPs and 20 microsatellites (D). *A. clanga* individuals are plotted as blue boxes, *A. pomarina* as green triangles and hybrids as red dots.

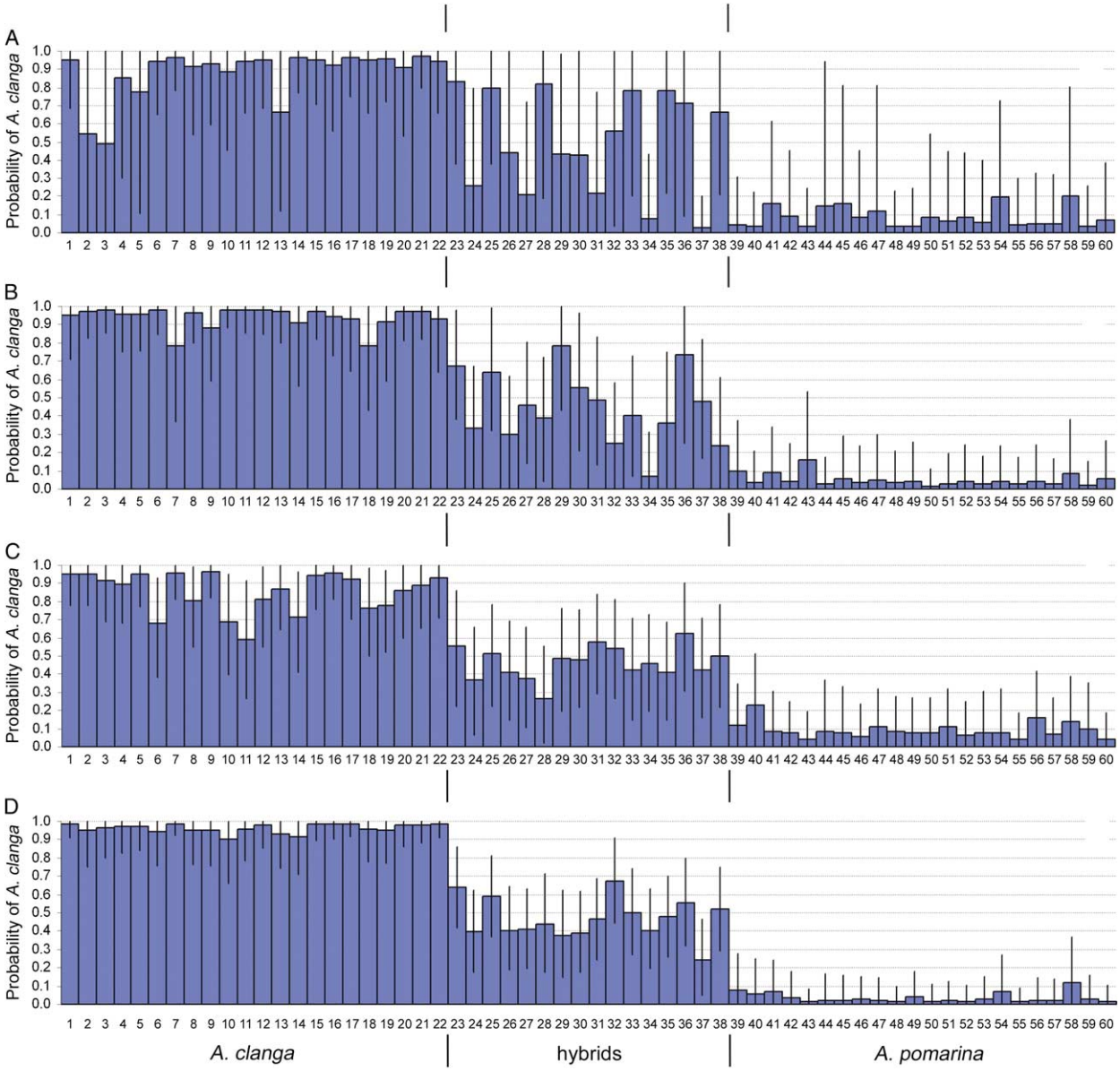


Figure 2. Probabilities ($\pm$95% intervals) that studied individuals have recent *A. clanga* ancestry obtained by "STUCTURE" (see Methods) using 10 microsatellites (A), 20 microsatellites (B), 9 SNPs (C), 9 SNPs and 20 microsatellites (D). Individuals are listed as in the Appendix 1.

analysis by 60 simulated hybrid genotypes suggested that NewHybrids with 20 microsatellites were able to correctly identify 95% of F1 hybrids (average $q_n = 0.78 \pm 0.11$), but only 20% ($q_n = 0.54 \pm 0.03$) and 5% ($q_n = 0.52$) of backcrosses to *A clanga* and *A. pomarina*, respectively. Ten microsatellites were able to identify 50% of F1 hybrids with low probabilities ($q_n = 0.64 \pm 0.07$), but no backcrosses. When 15 simulated hybrids were analysed, 20 microsatellites correctly assigned only one F1 hybrid, but no backcrosses; 10 microsatellites were not able to identify any of the hybrids.

Hybrid identification by SNPs and microsatellites

Finally, we analysed our samples with all markers, nine SNPs and 20 microsatellites, simultaneously. FCA plot separated two species and hybrids from each other well (Fig. 1D). Structure correctly assigned all *A. clanga* and 95% of *A. pomarina* individuals, while one *A. pomarina* individual was marginally above the limit ($q_s = 0.11$). Sixty three percent of hybrids were identified as F1 and 12% as backcrosses, while 25% were assigned between the F1 and backcross groups (Fig. 2D). NewHybrids identified all *A. pomarina* and *A. clanga* individuals, as well as 12 putative F1 hybrids, although four of them showed rather low support for F1 ($q_n < 0.6$; Fig. 3D). Three of the remaining hybrids were assigned to *A. pomarina* $\times$ F1 and one to the *A. clanga* $\times$ F1 group (Fig. 3D). Power analysis by 60 artificial hybrids showed that NewHybrids was able to correctly assign all F1 hybrids ($q_n = 0.93 \pm 0.07$), 55% of *A. clanga* $\times$ F1 ($q_n = 0.75 \pm 0.12$), and 85% ($q_n = 0.86 \pm 0.11$) of *A. pomarina* $\times$ F1. When only 15 simulated

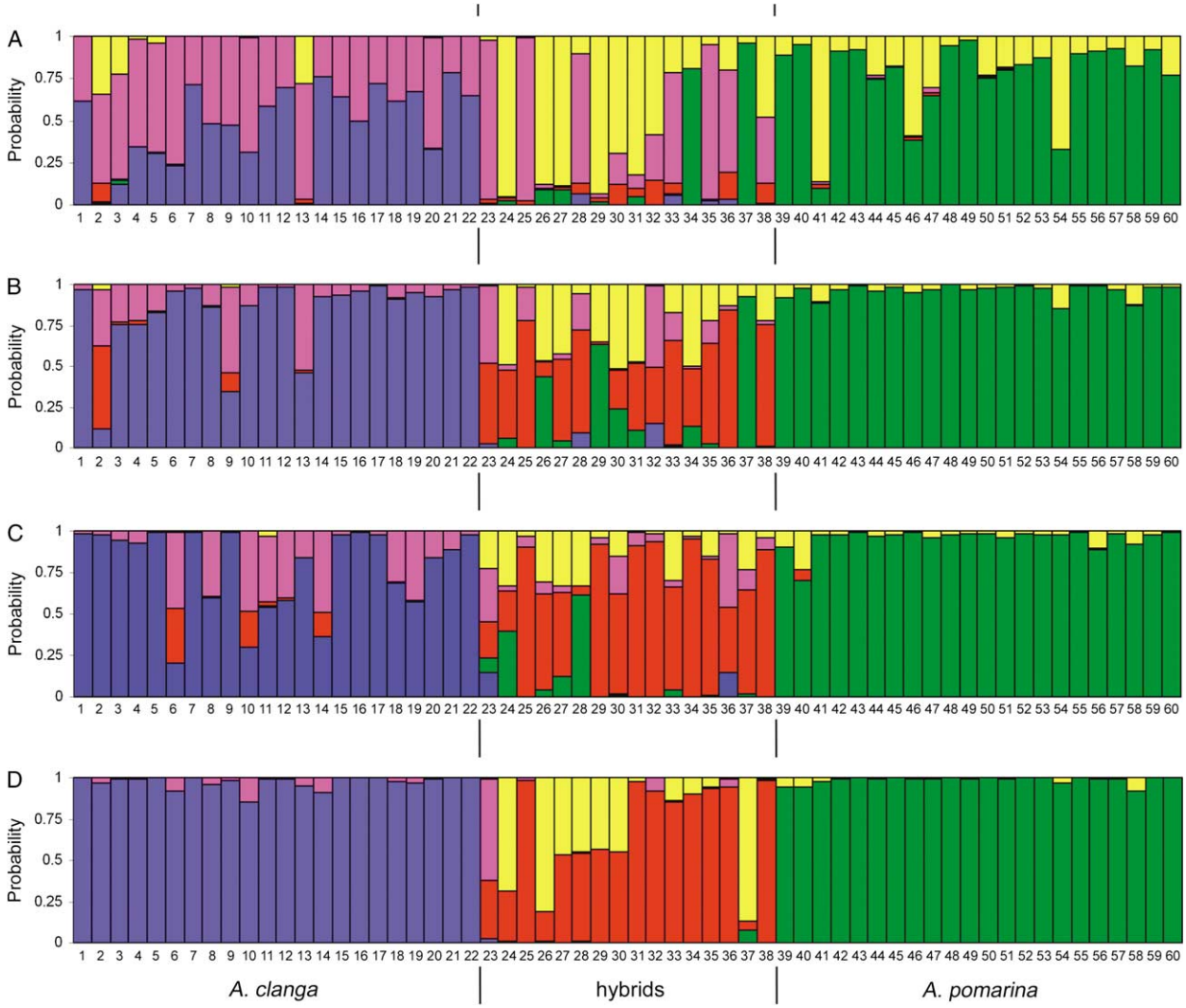


Figure 3. Probabilities that studied individuals belong to *A. clanga* (blue), *A. pomarina* (green), F1 hybrids (red), *A. clanga* × F1 (purple) and *A. pomarina* × F1 (yellow) obtained by NewHybrids using 10 microsatellites (A), 20 microsatellites (B) 9 SNPs (C), 9 SNPs and 20 microsatellites (D). Individuals are listed as in the Fig. 2 and Appendix 1.

hybrids were studied, all F1 and 40% of *A. pomarina* × F1 were correctly assigned, but all *A. clanga* × F1 backcrosses were incorrectly identified as *A. clanga*.

Discussion

Our study confirms that an appropriate set of markers is essential for successful hybrid identification; a marker set with insufficient resolution power increases uncertainty in results however and also leads to incorrect assignments. If the number of markers is limited, a careful check should be made to see whether the markers are powerful enough to discriminate species correctly. It should be borne in mind for instance that biallelic markers always indicate the existence of two genetic lineages, but that these are not necessarily the same as phylogenetic lineages of species. The use of more markers is always better, and the large number of markers is essential when less-specific (randomly chosen) markers are used. We discuss below the strengths and shortcomings of both marker types, and their combination, in respect of hybridisation between spotted eagles.

The standard practice in the use of SNPs in hybridisation studies has been the exploitation of a few markers from autosomal genes often used in taxonomic studies, and making conclusions based on their inheritance. Several nuclear genes have been used in phylogenetic studies of eagles (Helbig et al. 2005a, Lerner and Mindell 2005). Our data show that these markers cannot be used to assign individuals to species (at least in sympatry) as both alleles of these biallelic markers may occur with high frequency in both species. Remarkably, only one out of twelve such markers added power to species identification. New markers had to be developed therefore for our study, based on the assumption that the proportion of differentiated loci could be small. Indeed, only eight of 48 sequenced introns added significantly to the identification of spotted eagle species and differentiation of hybrids, whereas our extensive genome-wide survey did not result in any strictly species-specific markers that could be used on their own to identify

hybrids (see also Helbig et al. 2005b for similar results by AFLPs). As our complete marker set eventually identified all groups well, and genetic results showed a high agreement with pre-identified classes, we are confident that extensive allele sharing does not arise from flaws in the initial assignment. Since eight out of nine SNPs selected by us could be analysed using a simple PCR and a restriction analysis, these markers may be useful for large-scale low-cost monitoring using very basic laboratory equipment. However, as there was some uncertainty and even some errors associated with using the selected SNPs, the addition of a few additional SNPs is recommended.

In many cases a comparatively low number of microsatellites was considered adequate enough for hybrid identification, especially when there was no need to distinguish advanced backcrosses (Crochet et al. 2003, Shriver et al. 2005, Berthier et al. 2006, Vigfúsdóttir et al. 2008). The efficiency of such markers depends mainly on their interspecies differences of allele frequencies. As the common practice is the employment of randomly selected or mostly polymorphic (usually within a species of origin) microsatellites, large number of markers should be used in order to include also those with high discriminative power. Although ten microsatellites may be adequate for successful identification of hybrids in other species, the resolution power of the set employed by us was evidently too low for correct assignment. Twenty microsatellites were more effective, and outperformed SNPs in species identification, but most of the hybrids were incorrectly assigned. A set of such 'naïve' multilocus microsatellite markers, not developed for distinguishing between the two spotted eagle species, obviously described a much smaller proportion of variation compared with the specially selected biallelic SNPs (Fig. 1–3), since a large proportion of variation is required to describe intra-specific differences. If we select only the most discriminative microsatellites with largest inter-species allele frequency differences (see Appendix 3), results were more similar to those obtained by SNPs. Although we used microsatellites developed from other species, we do not believe that a low allele number necessarily results from cross-species use, since the allele number of particular microsatellites was even lower in the original species. On the other hand, cross-species use of microsatellites helps to avoid identification bias, which could occur if markers are developed from only one of the interbreeding species. A higher degree of polymorphism is often found in species from which markers are developed (Avise 2004) and that could affect, for instance, the estimation of introgression rate. In our study, no such polymorphism difference between species was expected since all markers came from other species.

Combination of SNP and microsatellite data has also been used successfully in earlier hybridisation studies (e.g. Muñoz-Fuentes et al. 2007). The simultaneous use of two marker types, i.e. including all the possible information, also gave the best result in our study, showing that two marker sets complement each other, and re-emphasises that an adequate number of markers is essential for correct conclusions. We consider the efficiency of final assignment a proof of the correctness of our initial morphological preassignment and the insufficient analysis power of a smaller marker set. It appears that phenotypic identification is a reasonably good means of identifying a genotype in spotted eagles, which is not always the case (Allendorf et al. 2001, Gaubert et al. 2005, Chan et al. 2006, Gay et al. 2009, Haas et al. 2009). However, even when markers were used in combination, this did not always lead to unequivocal assignment of spotted eagle hybrids to the F1 group as expected. We agree that four individuals eventually assigned as backcrosses (Fig. 3D) could indeed be backcrosses, since these types are very hard to distinguish from F1 hybrids morphologically; a parent of one such backcross (Ind. 26) was later genetically assigned to *A. pomarina* (Ü. Väli, unpubl. data). However, four individuals from Estonia showed almost equal probabilities of being F1 hybrids or backcrosses to *A. pomarina* (Fig. 3D). At least three of them are F1 hybrids and definitely not backcrosses to *A. pomarina*, as their mothers were later unequivocally assigned to *A. clanga* by similar genetic analysis, and the fathers identified as *A. pomarina* in the field (Ü. Väli, unpubl. data). Increasing the number of markers would probably result in more accurate assignment, but as Vähä and Primmer (2006) have shown, some backcrosses could remain unidentified even when a large number of markers are used. According to the power analysis, a larger sample size would also increase the efficiency of assigning hybrids (see also Morin et al. 2009), but some backcrosses could still remain unidentified.

Modern assignment techniques clearly enable us to identify hybrids, and separate parental species, more efficiently than earlier distance or ordination based methods (Randi and Lucchini 2002). Compared to the likelihood-based assignments, Bayesian analyses do not require knowledge of parental (allopatric) populations (Vähä and Primmer 2006). Even though selected SNPs and microsatellites gave better results than AFLPs in spotted eagle hybrid identification, the re-analysis of data by Helbig et al. (2005b) using a modern Bayesian model-based method would probably result in better discrimination of species and identification of hybrids. Although both methods used by us resulted in the identification of hybrids, the basic idea behind the analysis is somewhat different. Structure assumes that a hybrid has inherited some fraction of its genome from both ancestral species, thus being intermediate in some degree between two parental species. NewHybrids provides a posterior probability to reflect the level of certainty that an individual belongs to a certain hybrid group. Although NewHybrids seems thus to give a more direct answer on hybrid identity, it may have problems in covering the complete genetic variation between the two species, when introgression and weak selection against hybrids result in the occurrence of later-generation hybrids and backcrosses. In that case, Structure seems to be more reliable in that it treats the probability of ancestry as a continuous variable rather than attempting to classify discrete groups. Our results suggest, however, that in practice an adequate number and proper selection of markers is far more important than analytical differences between the methods. This is clearly illustrated by the increasing similarity of results, and decreasing amplitude of probability intervals, when more markers were used. The simultaneous use of two or more assignment methods is therefore always justified, but becomes essential when the number of markers used is limited.

For the detection of hybrids and parental species the selection of the threshold is crucial, and there is always a trade off between assignment efficiency and accuracy. We used the error threshold 0.1 in our analyses, but also included backcross classes and permitted some uncertainty in assignments. If we permitted a higher error rate, of say 0.2 (i.e. $q_s = 0.8$ for parental species), which may give the best performance in respect of both identification accuracy and efficiency (Vähä and Primmer 2006), and included only three presumed groups, we would increase the efficiency of detection, but decrease its accuracy. The threshold in Structure is sometimes applied only to separate pure species, whereas all individuals below the given probability indicate genetic admixture (e.g. Barilani et al. 2007). This would lead to higher efficiency but lower accuracy of hybrid identification. In summary, the threshold selection should be based on the purpose of the study (Vähä and Primmer 2006), as well as on the properties of particular samples and markers. However, one should always check probability intervals as the assignments could be meaningless due to the low assignment power of markers (see Fig. 2).

SNPs found in introns are diagnostic probably not directly due to natural selection, but at least some such alleles may be 'hitchhiking' along with those located within exons (Maynard Smith and Haigh 1974). In the future, screening of such markers could result in the identification of genes responsible for speciation (Schlötterer 2003). On the other hand, the sharing of SNP alleles could be caused simply by incomplete lineage sorting or back-mutations, whereas the size of homoplasy could be responsible for the similarity of microsatellites. Historical hybridisation events may also lead to the same pattern. Based on the AFLP data, Helbig et al. (2005b) estimated that gene flow between two species may be one migrant per generation (but the estimation was ten times smaller for mtDNA). They also suggested that hybridisation of spotted eagles began recently, but did not exclude the possibility of interbreeding during earlier interglacial periods. Our data of extensive allele sharing among species also indicates that the spotted eagle species which have shared the same continent for a long period of time, and which are capable of colonising new areas relatively rapidly, must have had several secondary contacts throughout their history leading to genetic introgression. Despite such probable repeated hybridisation periods, the existence of pure species nowadays should not be called into question (sensu Helbig et al. 2002), since the differentiation between species is clear, individuals can be successfully assigned to species, and interspecific hybrids can be identified using the appropriate set of markers. As efficient genetic markers and assignment methods have now become available, a more comprehensive analysis of hybridisation between spotted eagles can be performed. This would add to the present knowledge on the biology and conservation of spotted eagles, as hybridisation appears to be an important threat to the *A. clanga* in Europe (Väli 2005). Furthermore, as regular hybridisation among raptors is rather rare in nature (Panov 1989), analysis of hybridisation between spotted eagles would provide interesting data on hybrid zones in general.

Acknowledgements – Our thanks go to Szilard Daroczi, Miroslav Dravecky, Marina Dzmitranok, Mikhail Ivanov, Roman Kiselev, Anton Makarov, Joachim Matthes, Vladimir Melnikov, Gennady Mindlin, Kostas Poirazidis, Svetlana Romanova, Vitaliy Ryabtsev, Wolfgang Scheller, Dmitriy Zhuravlev, Eet and Aarne Tuule, Nikolay Yakovets and the members of the Estonian Eagle Club for their assistance with the collection of samples. Discussions with Frank Hailer, Jennifer Leonard, Violeta Muñoz-Fuentes, Carles Vilà, as well as comments from two anonymous referees helped to improve the manuscript. Financial support was provided by APB-BirdLife Belarus, the Estonian Environm. Investm. Centre, the Estonian Sci. Found. (grants no. 6050 and 7593), the Frankfurt Zool. Society, the Marie Curie Intra-European Fellowship for ÜV within the 6th European Comm. Framework programme, the Royal Soc. for the Protection of Birds and the Visby programme of the Swedish Inst.

References

Allendorf, F. W., Leary, R. F., Spruell, P. and Wenburg, J. K. 2001. The problems with hybrids: setting conservation guidelines. – Trends Ecol. Evol. 16: 613–622.

Anderson, E. C. and Thompson, E. A. 2002. A model-based method for identifying species hybrids using multilocus genetic data. – Genetics 160: 1217–1229.

Arnold, M. L. 1992. Natural hybridization as an evolutionary process. – Annu. Rev. Ecol. Syst. 23: 237–261.

Avise, J. C. 2004. Molecular markers, natural history, and evolution. – Sinauer Ass., Sunderland.

Backström, N., Fagerberg, S. and Ellegren, H. 2008. Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. – Mol. Ecol. 17: 964–980.

Barilani, M., Sfougaris, A., Giannakopoulos, A., Mucci, N., Tabarroni, C. and Randi, E. 2007. Detecting introgressive hybridisation in rock partridge populations (*Alectoris graeca*) in Greece through Bayesian admixture analyses of multilocus genotypes. – Conserv. Genet. 8: 343–354.

Barton, N. H. and Hewitt, G. M. 1989. Adaptation, speciation and hybrid zones. – Nature 341: 497–503.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. and Bonhomme, F. 2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. – Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

Bensch, S. and Åkesson, M. 2005. Ten years of AFLP in ecology and evolution: why so few animals? – Mol. Ecol. 14: 2899–2914.

Bergmanis, U. 1996. On the taxonomy of the lesser spotted eagle *Aquila pomarina* and greater spotted eagle *A. clanga*. – In: Meyburg, B.-U. and Chancellor, R. D. (eds). Eagle studies. World working grp. on birds of prey and owls, Berlin, London, Paris, pp. 199–207.

Berthier, P., Excoffier, L. and Ruedi, M. 2006. Recurrent replacement of mtDNA and cryptic hybridization between two sibling bat species *Myotis myotis* and *Myotis blythii*. – Proc. R. Soc. B Sci. 273: 3101–3109.

Boecklen, W. J. and Howard, D. J. 1997. Genetic analysis of hybrid zones: numbers of markers and power of resolution. – Ecology 78: 2611–2616.

Bureš, S., Nadvorník, P. and Sætre, G. P. 2002. Hybridization and apparent hybridization between meadow pipit (*Anthus pratensis*) and water pipit (*A. spinoletta*). – Hereditas 136: 254–256.

Burke, B. P. and Dawson, D. A. 2006. Fifteen microsatellite loci characterized in the golden eagle *Aquila chrysaetos* (Accipitridae, Aves). – Mol. Ecol. Notes 6: 1047–1050.

Busch, J. D., Katzner, T. E., Bragin, E. and Keim, P. 2005. Tetranucleotide microsatellites for aquila and haliaeetus eagles. – Mol. Ecol. Notes 5: 39–41.

Chan, C. H., Ballantyne, K. N., Aikman, H., Fastier, D., Daugherty, C. H. and Chambers, G. K. 2006. Genetic analysis of interspecific hybridisation in the world's only Forbes' parakeet (*Cyanoramphus forbesi*) natural population. – Conserv. Genet. 7: 493–506.

Crochet, P. A., Chen, J. J. Z., Pons, J. M., Lebreton, J. D., Hebert, P. D. N. and Bonhomme, F. 2003. Genetic differentiation at nuclear and mitochondrial loci among large white-headed gulls: sex-biased interspecific gene flow? – Evolution 57: 2865–2878.

Dombrovski, V. C. 2006. Morphological characteristics and diagnostic features of the greater spotted (*Aquila clanga*), lesser spotted (*A. pomarina*) eagles, and their hybrids. – Ornithologia 33: 29–41.

Dombrovski, V. C. and Demongin, L. 2006. On geographic variability of some diagnostic characteristics of the greater spotted eagle (*Aquila clanga*). – Ornithologia 33: 42–52.

Dowling, T. E. and Secor, C. L. 1997. The role of hybridization and introgression in the diversification of animals. – Annu. Rev. Ecol. Syst. 28: 593–619.

Forsman, D. 1999. The raptors of Europe and the Middle East: a handbook of field identification. – T. & A. D. Poyser, London.

Gaubert, P., Taylor, P. J., Fernandes, C. A., Bruford, M. W. and Veron, G. 2005. Patterns of cryptic hybridization revealed using an integrative approach: a case study on genets (*Carnivora*, *Viverridae*, *Genetta* spp.) from the southern African subregion. – Biol. J. Linn. Soc. 86: 11–33.

Gay, L., Neubauer, G., Zagalska-Neubauer, M., Pons, J.-M., Bell, D. and Crochet, P.-A. 2009. Speciation with gene flow in the large white-headed gulls: does selection counterbalance introgression? – Heredity 102: 133–146.

Grant, P. R. and Grant, B. R. 1994. Phenotypic and genetic effects of hybridization in Darwin's finches. – Evolution 48: 297–316.

Grant, P. R. and Grant, B. R. 1992. Hybridization of bird species. – Science 256: 193–197.

Grant, P. R. and Grant, B. R. 1997. Hybridization, sexual imprinting and mate choice. – Am. Nat. 149: 1–28.

Haas, F., Pointer, M. A., Saino, N., Brodin, A., Mundy, N. I. and Hansson, B. 2009. An analysis of population genetic differentiation and genotype-phenotype association across the hybrid zone of carrion and hooded crows using microsatellites and MC1R. – Mol. Ecol. 18: 294–305.

Hailer, F., Gautschi, B. and Helander, B. 2005. Development and multiplex PCR amplification of novel microsatellite markers in the white-tailed sea eagle, *Haliaeetus albicilla* (*Aves*: *Falconiformes*, *Accipitridae*). – Mol. Ecol. Notes 5: 938–940.

Helbig, A. J., Knox, A. G., Parkin, D. T., Sangster, G. and Collinson, M. 2002. Guidelines for assigning species rank. – Ibis 144: 518–525.

Helbig, A. J., Kocum, A., Seibold, I. and Braun, M. J. 2005a. A multi-gene phylogeny of aquiline eagles (*Aves*: *Accipitriformes*) reveals extensive paraphyly at the genus level. – Mol. Phylogenet. Evol. 35: 147–164.

Helbig, A. J., Seibold, I., Kocum, A., Liebers, D., Irwin, J., Bergmanis, U., Meyburg, B. U., Scheller, W., Stubbe, M. and Bensch, S. 2005b. Genetic differentiation and hybridization between greater and lesser spotted eagles (*Accipitriformes*: *Aquila clanga*, *A. pomarina*). – J. Ornithol. 146: 226–234.

Lecis, R., Pierpaoli, M., Biro, Z. S., Szemethy, L., Ragni, B., Vercillo, F. and Randi, E. 2006. Bayesian analyses of admixture in wild and domestic cats (*Felis silvestris*) using linked microsatellite loci. – Mol. Ecol. 15: 119–131.

Lerner, H. R. L. and Mindell, D. P. 2005. Phylogeny of eagles, Old World vultures, and other *Accipitridae* based on nuclear and mitochondrial DNA. – Mol. Phylogenet. Evol. 37: 327–346.

Lõhmus, A. and Väli, Ü. 2004. The effects of habitat quality and female size on the productivity of the lesser spotted eagle *Aquila pomarina* in the light of the alternative prey hypothesis. – J. Avian Biol. 35: 455–464.

Manel, S., Gaggiotti, O. E. and Waples, R. S. 2005. Assignment methods: matching biological questions techniques with appropriate. – Trends Ecol. Evol. 20: 136–142.

Martinez-Cruz, B., David, V. A., Godoy, J. A., Negro, J. J., O'brien, S. J. and Johnson, W. E. 2002. Eighteen polymorphic microsatellite markers for the highly endangered Spanish imperial eagle (*Aquila adalberti*) and related species. – Mol. Ecol. Notes 2: 323–326.

Maynard Smith, J. and Haigh, J. 1974. The hitch-hiking effect of a favourable gene. – Gen. Res. 23: 23–35.

McCarthy, E. M. 2006. Handbook of avian hybrids of the world. – Oxford Univ. Press.

Morin, P. A., Luikart, G., Wayne, R. K. and Group, S. W. 2004. SNPs in ecology, evolution and conservation. – Trends Ecol. Evol. 19: 208–216.

Morin, P. A., Martien, K. K. and Taylor, B. L. 2009. Assessing statistical power of SNPs for population structure and conservation studies. – Mol. Ecol. Res. 9: 66–73.

Mueller, U. G. and Wolfenbarger, L. L. 1999. AFLP genotyping and fingerprinting. – Trends Ecol. Evol. 14: 389–394.

Muñoz-Fuentes, V., Vilà, C., Green, A. J., Negro, J. J. and Sorenson, M. D. 2007. Hybridization between white-headed ducks and introduced ruddy ducks in Spain. – Mol. Ecol. 16: 629–638.

Nielsen, E. E. G., Bach, L. A. and Kotlicki, P. 2006. HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. – Mol. Ecol. Notes 6: 971–973.

Nittinger, F., Gamauf, A., Pinsker, W., Wink, M. and Haring, E. 2007. Phylogeography and population structure of the saker falcon (*Falco cherrug*) and the influence of hybridization: mitochondrial and microsatellite data. – Mol. Ecol. 16: 1497–1517.

Panov, E. N. 1989. Hybridization and ecological isolation in birds. – Nauka, Moscow.

Peakall, R. and Smouse, P. E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. – Mol. Ecol. Notes 6: 288–295.

Primmer, C. R., Borge, T., Lindell, J. and Saetre, G. P. 2002. Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. – Mol. Ecol. 11: 603–612.

Pritchard, J. K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. – Genetics 155: 945–959.

Prychitko, T. M. and Moore, W. S. 1997. The utility of NA sequences of an intron from the beta-fibrinogen gene in phylogenetic analysis of woodpeckers (*Aves*: *Picidae*). – Mol. Phylogenet. Evol. 8: 193–204.

Randi, E., Pierpaoli, M., Beaumont, M., Ragni, B. and Sforzi, A. 2001. Genetic identification of wild and domestic cats (*Felis silvestris*) and their hybrids using Bayesian clustering methods. – Mol. Biol. Evol. 18: 1679–1693.

Randi, E. and Lucchini, V. 2002. Detecting rare introgression of domestic dog genes into wild wolf (Canis lupus) populations by Bayesian admixture analyses of microsatellite variation. – Conserv. Genet. 3: 31–45.

Rudnick, J. A., Katzner, T. E., Bragin, E. A., Rhodes, O. E. and Dewoody, J. A. 2005. Using naturally shed feathers for individual identification, genetic parentage analyses, and population monitoring in an endangered eastern imperial eagle (*Aquila heliaca*) population from Kazakhstan. – Mol. Ecol. 14: 2959–2967.

Schlötterer, C. 2003. Hitchhiking mapping-functional genomics from the population genetics perspective. – Trends Gen. 19: 32–38.

Schlötterer, C. 2004. The evolution of molecular markers-just a matter of fashion? – Nat. Rev. Genet. 5: 63–69.

Seibold, I., Helbig, A. J., Meyburg, B. U., Negro, J. J. and Wink, M. 1996. Genetic differentiation and molecular phylogeny of European *Aquila* eagles according to cytochrome b nucleotide sequences. – In: Meyburg, B.-U. and Chancellor, R. D. (eds). Eagle studies. WWGBP, pp. 1–15.

Shriver, W. G., Gibbs, J. P., Vickery, P. D., Gibbs, H. L., Hodgman, T. P., Jones, P. T. and Jacques, C. N. 2005. Concordance between morphological and molecular markers in assessing hybridization between sharp-tailed sparrows in New England. – Auk 122: 94–107.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. 1994. Clustal-W-Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. – Nucleic Acids Res. 22: 4673–4680.

Vähä, J. P. and Primmer, C. R. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. – Mol. Ecol. 15: 63–72.

Väli, Ü. 2002. Mitochondrial pseudo-control region in old world eagles (genus *Aquila*). – Mol. Ecol. 11: 2189–2194.

Väli, Ü. 2005. Hybridisation: a threat to the European Greater Spotted Eagle *Aquila clanga* population. – In: Mizera, T. and Meyburg, B.-U. (eds). International meeting on Spotted Eagles (*Aquila clanga*, *A. pomarina* and *A. hastata*) – research and conservation. Biebrza National Park, Osowiec, pp. 103–114.

Väli, Ü. and Lõhmus, A. 2004. Nestling characteristics and identification of the lesser spotted eagle *Aquila pomarina*, greater spotted eagle *A. clanga*, and their hybrids. – J. Ornithol. 145: 256–263.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. and Shipley, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. – Mol. Ecol. Notes 4: 535–538.

Verardi, A., Lucchini, V. and Randi, E. 2006. Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (*Canis lupus*) by admixture linkage disequilibrium analysis. – Mol. Ecol. 15: 2845–2855.

Vigfúsdóttir, F., Pálsson, S. and Ingólfsson, A. 2008. Hybridization of glaucous gull (*Larus hyperboreus*) and herring gull (*Larus argentatus*) in Iceland: mitochondrial and microsatellite data. – Philos. Trans. R. Soc. B. 363: 2851–2860.

Vilà, C., Walker, C., Sundqvist, A. K., Flagstad, O., Andersone, Z., Casulli, A., Kojola, I., Valdmann, H., Halverson, J. and Ellegren, H. 2003. Combined use of maternal, paternal and bi-parental genetic markers for the identification of wolf-dog hybrids. – Heredity 90: 17–24.

Weir, B. S. and Cockerham, C. C. 1984. Estimating F-statistics for the analysis of population structure – Evolution 38: 1358–1370.

Appendix 1. Spotted eagle individuals examined in the current study. In addition to the sample group, age (nestling 'pull.' or adult 'ad'), sampling locality and maternal lineage type are shown.

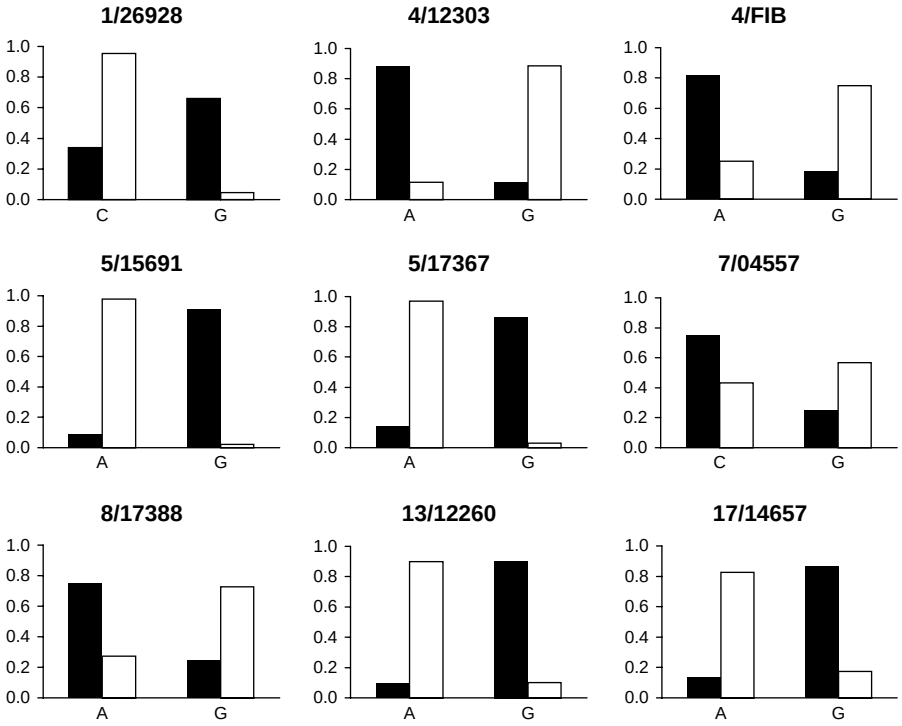
No.	Gr	Age	Population	MtDNA	No.	Grp.	Age	Population	MtDNA
1.	A. cla	pull.	Russia, Upper Volga	A. cla	31.	Hybrid	pull.	Estonia	A. cla
2.	A. cla	pull.	Russia, Upper Volga	A. cla	32.	Hybrid	pull.	Estonia	A. cla
3.	A. cla	pull.	Russia, Upper Volga	A. cla	33.	Hybrid	pull.	Estonia	A. cla
4.	A. cla	pull.	Russia, Upper Volga	A. cla	34.	Hybrid	pull.	Estonia	A. cla
5.	A. cla	ad.	Estonia	A. cla	35.	Hybrid	pull.	Estonia	A. cla
6.	A. cla	pull.	Estonia	A. cla	36.	Hybrid	pull.	Lithuania	A. pom
7.	A. cla	pull.	Estonia	A. cla	37.	Hybrid	pull.	Poland	A. pom
8.	A. cla	pull.	Estonia	A. cla	38.	Hybrid	pull.	Poland	A. cla
9.	A. cla	pull.	Estonia	A. cla	39.	A. pom	ad.	Germany	A. pom
10.	A. cla	pull.	Estonia	A. cla	40.	A. pom	pull.	Germany	A. pom
11.	A. cla	pull.	Belarus	A. cla	41.	A. pom	pull.	Germany	A. pom
12.	A. cla	pull.	Belarus	A. cla	42.	A. pom	pull.	Germany	A. pom
13.	A. cla	pull.	Belarus	A. cla	43.	A. pom	pull.	Germany	A. pom
14.	A. cla	pull.	Belarus	A. cla	44.	A. pom	ad.	Greece	A. pom
15.	A. cla	pull.	Belarus	A. cla	45.	A. pom	ad.	Greece	A. pom
16.	A. cla	pull.	Belarus	A. cla	46.	A. pom	pull.	Estonia	A. pom
17.	A. cla	pull.	Belarus	A. cla	47.	A. pom	pull.	Estonia	A. pom
18.	A. cla	pull.	Belarus	A. cla	48.	A. pom	pull.	Estonia	A. pom
19.	A. cla	ad.	Poland	A. cla	49.	A. pom	pull.	Estonia	A. pom
20.	A. cla	pull.	Poland	A. cla	50.	A. pom	pull.	Estonia	A. pom
21.	A. cla	ad.	Poland	A. cla	51.	A. pom	pull.	Estonia	A. pom
22.	A. cla	ad.	Russia, Lake Baikal	A. cla	52.	A. pom	ad.	Poland	A. pom
23.	Hybrid	pull.	Russia, Upper Volga	A. pom	53.	A. pom	ad.	Romania	A. pom
24.	Hybrid	pull.	Russia, Upper Volga	A. cla	54.	A. pom	ad.	Romania	A. pom
25.	Hybrid	pull.	Germany	A. cla	55.	A. pom	ad.	Slovakia	A. pom
26.	Hybrid	pull.	Estonia	A. pom	56.	A. pom	pull.	Lithuania	A. pom
27.	Hybrid	pull.	Estonia	A. cla	57.	A. pom	pull.	Lithuania	A. pom
28.	Hybrid	pull.	Germany	A. cla	58.	A. pom	pull.	Lithuania	A. pom
29.	Hybrid	pull.	Estonia	A. cla	59.	A. pom	pull.	Lithuania	A. pom
30.	Hybrid	pull.	Estonia	A. cla	60.	A. pom	ad.	Slovakia	A. pom

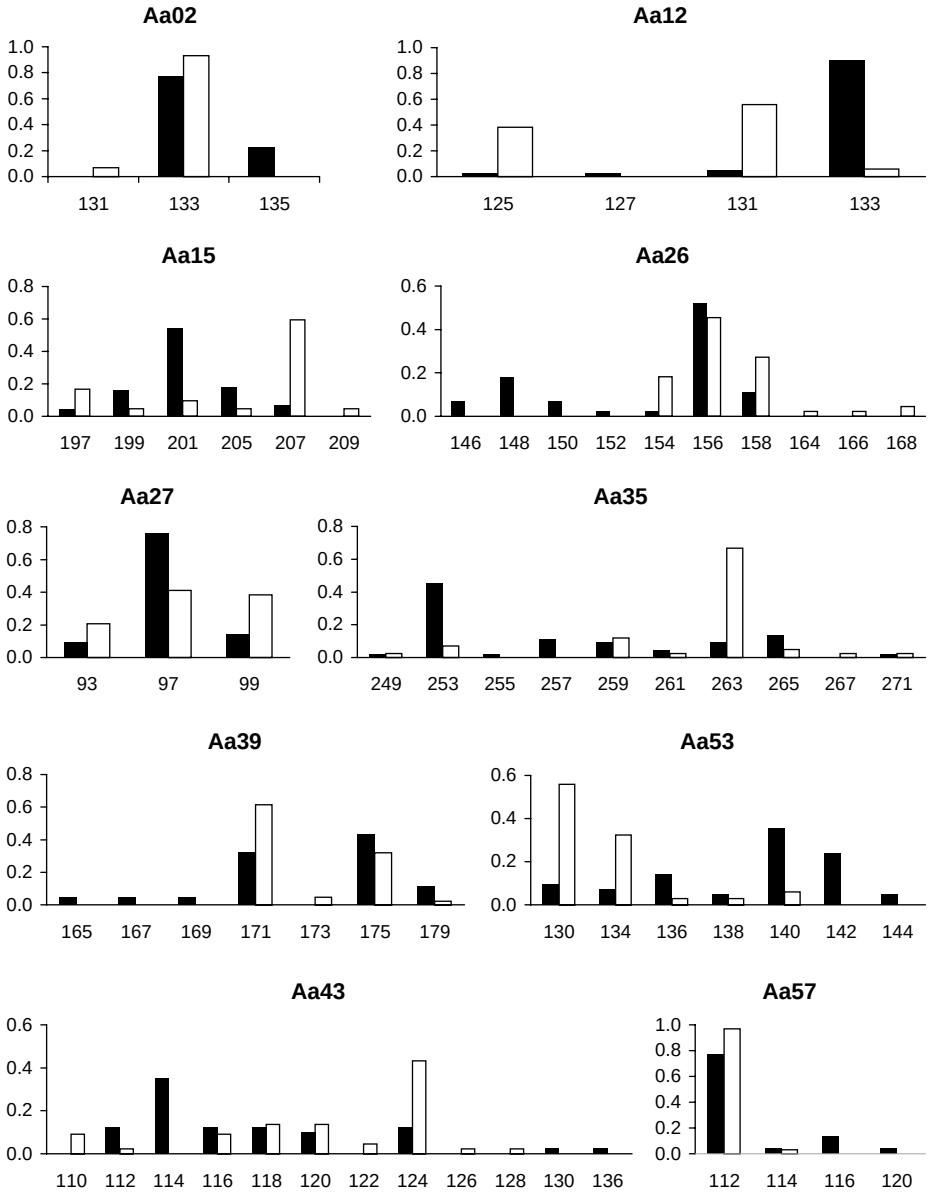
Appendix 2. Nine selected SNP markers and their amplification primer sequences, annealing temperatures, position of the SNP in the intron and appropriate restriction enzymes.

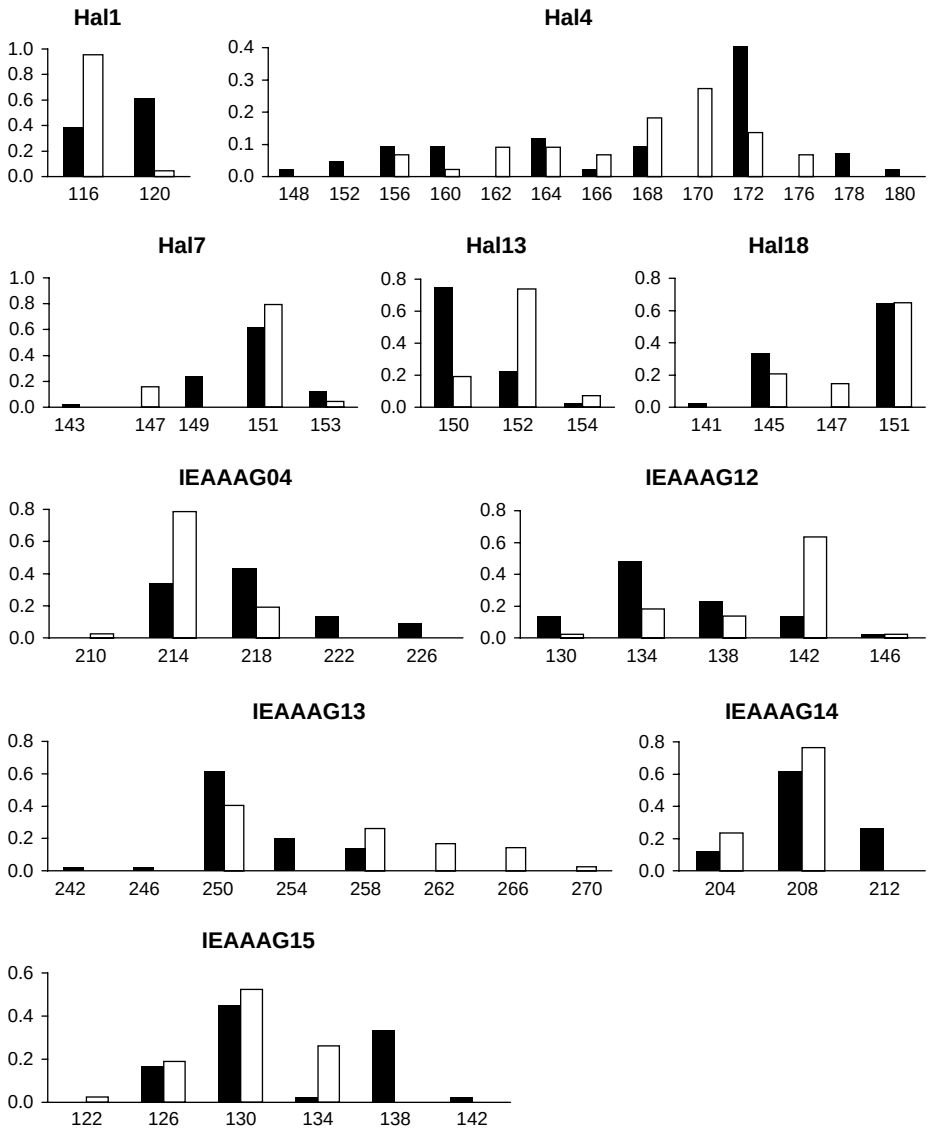
Chr	Locus	Gene	Amplification primers (5′–3′)	Ann. temp. (°C)	SNP	Restr. enz.
1	26928	Unknown protein	F 5′-GACCTTCCAGAAGCTATTGC-3′ R 5′-TCGTGAAGAACACGTGAAAG-3′	48	91	Bsp1286I [b]
4	12303	Unknown protein	F 5′-AATGTCAGCATGAAGAATGC-3′ R 5′-TTCTTTACTGTCATTGCCGC-3′	57	245	MboII [c]
4	FIB	β-fibrinogene	F 5′-GGAGAAAACAGGACAATGACAATTCAC-3′ R 5′-TCCCCAGTAGTATCTGCCATTAGGGTT-3′	52	446	MboII [c]
5	15691	Unknown protein	F 5′-CTCCAGATGAAATCTTCTGG-3′ R 5′-GGATCACTGGTGTTGACAAC-3′	55	370	NmuCI [c]
5	17367	26S protease regulatory subunit 4	F 5′-GGAGTCCTGATGGATGACAC-3′ R 5′-GGTGGCTTTATACCCATCTC-3′	58	346	–
7	04557	Acyl-coenzymw A dehydrogenase	F 5′-CCCTACATTGCAAACTATGG-3′ R 5′-AGAATCCAGTCACTTCCATC-3′	53	331	StyI [b]
8	17388	MAGO-NASHI homologue	F 5′-GACGAGCACATCTCCTTTAC-3′ R 5′-TGGACCAGGTAGTAGAACAC-3′	53	209	NheI [b]
13	12260	Dihydropyrimidinase-like 3	F 5′-AAGCAGAAGCTGTCTTCCG-3′ R 5′-CAGTTCTTGCTCCAGTAGTG-3′	52	84	BauI [c]
17	14657	Endothelial differentiation-related factor 1 (EDF1)	F 5′-CAGAAGAGCTGCACCATG-3′ R 5′-TTCTGATTCGTAGTCAGC-3′	51	149	BsmAI [c]

[a] according to the Structure analysis of 22 *A. clanga* and 22 *A. pomarina*.
[b] cuts *A. pomarina*-specific allele.
[c] cuts *A. clanga*-specific allele.

Appendix 3. Allele frequencies of used markers in the greater (black bars) and in the lesser spotted eagle (white bars).




Aa02
1.0
0.8
0.6
0.4
0.2
0.0
131
133
135
Aa12
1.0
0.8
0.6
0.4
0.2
0.0
125
127
131
133
Aa15
0.8
0.6
0.4
0.2
0.0
197
199
201
205
207
209
Aa26
0.6
0.4
0.2
0.0
146
148
150
152
154
156
158
164
166
168
Aa27
0.8
0.6
0.4
0.2
0.0
93
97
99
Aa35
0.8
0.6
0.4
0.2
0.0
249
253
255
257
259
261
263
265
267
271
Aa39
0.8
0.6
0.4
0.2
0.0
165
167
169
171
173
175
179
Aa53
0.6
0.4
0.2
0.0
130
134
136
138
140
142
144
Aa43
0.6
0.4
0.2
0.0
110
112
114
116
118
120
122
124
126
128
130
136
Aa57
1.0
0.8
0.6
0.4
0.2
0.0
112
114
116
120


Hal1
1.0
0.8
0.6
0.4
0.2
0.0
116
120
Hal4
0.4
0.3
0.2
0.1
0.0
148
152
156
160
162
164
166
168
170
172
176
178
180
Hal7
1.0
0.8
0.6
0.4
0.2
0.0
143
147
149
151
153
Hal13
0.8
0.6
0.4
0.2
0.0
150
152
154
Hal18
0.8
0.6
0.4
0.2
0.0
141
145
147
151
IEAAAG04
0.8
0.6
0.4
0.2
0.0
210
214
218
222
226
IEAAAG12
0.8
0.6
0.4
0.2
0.0
130
134
138
142
146
IEAAAG13
0.8
0.6
0.4
0.2
0.0
242
246
250
254
258
262
266
270
IEAAAG14
0.8
0.6
0.4
0.2
0.0
204
208
212
IEAAAG15
0.6
0.4
0.2
0.0
122
126
130
134
138
142