Supplementary information for

Genetic diversity and spatial genetic structure in two sympatric woodpecker species support the specialist-generalist variation hypothesis

Gilberto Pasinelli

Swiss Ornithological Institute, Sempach, Switzerland and University of Zurich, Zurich, Switzerland, gilberto.pasinelli@vogelwarte.ch

Table of contents

Journal name: ***Conservation Genetics***

Table S1 Markers used in this study and multiplex settings

Locus	Source	Genebank	Repeat	Label F	Forward Primer	Reverse Primer	Concentration (µM)	Multiplex Panel	Annealing temperature (°C)
DMA2	Vila et al. 2008	EF592034	(CA)12	NED	CTCTTCCAACCTTGGTGATAC	TTGCATGGCTCAATCATATAG	0.2	5‡	60
DMA4	Vila et al. 2008	EF592035	(CA)8	VIC	GGGATGAACCATTTGTGG	GCTCTTTGTGGGAACTGG	0.1	1	59-55†
DMA120	Vila et al. 2008	EF592036	(GT)17	FAM	ACATGGATAGCAATAGTTCTG	AATACGACCCTCCAAAAG	0.2	1	59-55†
DMB119	Vila et al. 2008	EF592037	(ATC)25	PET	CCTGTCCAGCAAGCCTATC	CCTGAGCATGTGGTCAGTG	0.2	1	59-55†
DMC1	Vila et al. 2008	EF592038	(CATC)12	PET	TGCTGCACCTTACCATTTCTAG	ATGGAAGGCAGTGGTTTATACC	0.2	5	60
DMC8	Vila et al. 2008	EF592039	(GATG)15	FAM	GGTTGGACGTGATGATCTTG	CTGACCTTGAACCCACTCAG	0.2	5	60
DMC111	Vila et al. 2008	EF592040	(CATC)10	VIC	CGTATGGACCAGAACATAATG	TGGGCTTTTAAGTCTTGTTG	0.1	1	59-55†
DMC115	Vila et al. 2008	EF592041	(ATCC)10	FAM	AGGATGGTTTCTGTATTGTGTC	CTTTAAGCCATTTCTGAATGAC	0.2	2	59-55†
DMD7	Vila et al. 2008	EF592042	(TCTA)13	VIC	TGGAATCAAGTGTGGATACAC	CAGGGCTAATATCTCATGTTTC	0.2	5	60
DMD9	Vila et al. 2008	EF592043	(TAGA)14	VIC	CAAAATGTTCTGGCAGAGTG	TATTCAGGGCTATGTGTCTGTC	0.2	5	60
DMD112	Vila et al. 2008	EF592044	(CTAT)10	PET	TGTGGCATTAACATAGATATGC	AAGAACCTTGCACAAAACAC	0.2	2	59-55†
DMD118	Vila et al. 2008	EF592045	(TGGA)10(TAGA)13	FAM	CCCATATCCAGAGTTAGTTCTG	TCCTAGAGTCTTCAACCTGATC	0.2	2	59-55†
Dlu1	Ellegren et al. 1999	AF081910	(TG)17	VIC	CACACTGAACATACCATGTG	TAAAGACCCTAAACTTGCACA	0.1	2	59-55†
Dlu4	Ellegren et al. 1999	AF081913	(GT)36	NED	TTAAGAAGTTTCCAGTTCACC	ACAGATCTGTATGTTGATGAGA	0.2	1	59-55†
Dlu6	Ellegren et al. 1999	AF081915	(GT)11(GC)4	FAM	GGAACATGAATTTGCCCAAGT	CAGATTAGTTGAGTCATAGTT	0.2	2	59-55†
Denmaj_001481	this study	JX196852	(AC)15	FAM	TGAGAAGATAGGCTCACACAAG	TCTGCTGTAGCACAGTTGAAG	0.2	3	56

Denmaj_002164	this study	JX196853	(TCTA)14	VIC	CAATCACAGAGCAGGCAAGG	GAGCCCAGGAGACATGAGAC	0.2	4	62
Denmaj_002693	this study	JX196854	(AG)11	NED	GGAGCCACCTGAAAAAGCTG	CAGCCGTGTTTGCTTTGTTG	0.2	3	56
Denmaj_005890	this study	JX196855	(CA)13	PET	TGGGAACAGCATTTCCATAACC	ACCCTCTTCATCTTCCTCTTGG	0.2	3	56
Denmaj_006508	this study	JX196856	(AC)13	VIC	AGCCCACATACACTGGAAAATG	ACCAGCTAAAATAGGTTGATGCAG	0.2	3	56
Denmaj_006540	this study	JX196857	(CA)11	FAM	CATGATGTGCTGGAAGCAGG	GTGTCAGCATCCGTGTCTTG	0.2	4	62
Denmaj_007242	this study	JX196858	(CA)13	NED	CACAAGGCTACCCCTCCG	CTGGCACAACGAGATGTGTG	0.2	4	62

† touchdown protocol -0.2°C per cycle

‡ run in singleplex and pooled for fragment analysis

Table S2 Origin and number of available and working microsatellite markers used in this study. Origin refers to the species, for which the markers had originally been developed

Origin	Available	Working	
		Msw	Gsw
Middle spotted woodpecker[†]	12	12	6[¶]
White-backed woodpecker[‡]	6	3[#]	3[#]
Great spotted woodpecker[§]	7	-[°]	7
Total	25	15	16

[†] Vila et al. (2008), [‡] Ellegren et al. (1999), [§] developed by Ecogenics GmbH
[¶] Only 7 of the 12 available middle spotted woodpecker markers tested with samples from great spotted woodpecker.
[#] All six white-backed woodpecker markers tested in both species.
[°] Great spotted woodpecker markers not tested with middle spotted woodpecker samples.

Table S3 Genetic diversity per locus across seven local populations of the middle spotted woodpecker (Msw) and the great spotted woodpecker (Gsw). N = number of individuals typed, A = total number of alleles, H_o = mean observed heterozygosity, H_e = mean expected heterozygosity (SMM, Nei 1987), PIC = polymorphic information content, a measure of locus informativeness, HW E= p-values for Hardy-Weinberg probability tests (bold = significant after sequential Bonferroni correction) and population significantly out of HWE per locus, null alleles = populations showing signs of null alleles. Origin refers to the species for which the markers had originally been developed. Calculations were done in CERVUS 3.07 (Kalinowski et al. 2007), MICRO-CHECKER 2.2.3 (van Oosterhout et al. 2004) and GENEPOP (Raymond and Rousset 1995) on the web (http://genepop.curtin.edu.au/)

Species	Locus	N	A	H_o	H_e	PIC	HWE / Pop	Null alleles†	Origin‡
Msw	DMA120	115	13	0.843	0.849	0.828	0.861		Msw
	DMA2	116	6	0.534	0.589	0.537	0.425	HE	Msw
	DMA4	115	3	0.678	0.550	0.445	0.208		Msw
	DMB119	115	14	0.852	0.889	0.874	0.396		Msw
	DMC1	116	9	0.595	0.806	0.777	**0.001 / BL**	AG, BL	Msw
	DMC111	114	6	0.596	0.665	0.604	0.749		Msw
	DMC115	116	7	0.655	0.685	0.630	0.708		Msw
	DMC8	116	9	0.853	0.855	0.834	0.131		Msw
	DMD112	116	7	0.638	0.688	0.650	0.064		Msw
	DMD118	116	8	0.819	0.785	0.753	0.739		Msw
	DMD7	116	6	0.603	0.749	0.704	**0.002 / BL**	BL	Msw
	DMD9	116	9	0.767	0.824	0.797	0.430	NE	Msw
	Dlu1	116	2	0.009	0.042	0.041	0.010		Wbw
	Dlu4	111	21	0.586	0.904	0.893	nd	AG, BL, HE, TG	Wbw
	Dlu6	115	2	0.052	0.051	0.050	na		Wbw
Gsw	DMA1481	146	10	0.829	0.793	0.762	0.9		Gsw
	DMA2164	145	31	0.855	0.904	0.893	0.109	TG	Gsw
	DMA2693	147	8	0.639	0.689	0.650	0.098	SH	Gsw
	DMA5890	147	19	0.796	0.860	0.846	**0.001 / HE**	HE	Gsw
	DMA6508	145	10	0.297	0.321	0.310	0.618		Gsw
	DMA6540	146	8	0.473	0.497	0.447	0.294		Gsw
	DMA7242	145	4	0.283	0.466	0.427	nd	AG, BL, NE, TG	Gsw
	DMA4	143	1	0	0	0	na		Msw

DMB119	145	9	0.690	0.738	0.710	0.009		Msw
DMC111	145	10	0.683	0.745	0.711	0.146	ZH	Msw
DMC115	142	24	0.761	0.806	0.790	0.711	HE	Msw
DMD112	144	16	0.833	0.871	0.857	0.343		Msw
DMD118	145	15	0.807	0.828	0.803	0.609		Msw
Dlu1	145	8	0.290	0.708	0.671	nd	AG, BL, HE, NE, SH, TG, ZH	Wbw
Dlu4	139	43	0.763	0.948	0.943	nd	AG, HE, SH, TG, ZH	Wbw
Dlu6	146	5	0.274	0.302	0.285	0.478	AG	Wbw

† AG = Aargau, BL = Basel, HE = Hesse (Germany), NE = Neuchâtel, SH = Schaffhausen, TG = Thurgau, ZH = Zürich

‡ Msw: Vila et al. (2008), Wbw: Ellegren et al. (1999), Gsw: developed by Ecogenics GmbH

Table S4 Summary of STRUCTURE analyses conducted. See STRUCTURE manual for the meaning of ancestry prior, initial alpha and alpha for all populations

Location information	Ancestry prior	Initial alpha	Alpha for all pops	Analysis code
not used	Uniform (default)	1	same	noloc_1
not used	Uniform (default)	1	separate	noloc_2
not used	Alternative (gamma)	0.143	same	noloc_3
not used	Alternative (gamma)	0.143	separate	noloc_4
used	Uniform (default)	1	same	loc_1
used	Uniform (default)	1	separate	loc_2
used	Alternative (gamma)	0.143	same	loc_3
used	Alternative (gamma)	0.143	separate	loc_4

Table S5 Results of bottleneck tests for seven populations of two sympatric woodpecker species. The two phase model (TPM, Di Rienzo et al. 1994) and the stepwise mutation model (SMM, Ohta and Kimura 1973) were assumed as mutation models for microsatellite loci. Shown are one-sided *P* values of Wilcoxon tests for heterozygosity excess calculated with BOTTLENECK. For the TPM, two analyses were run, with values for the proportion of single-step mutations being 0.95 (Piry et al. 1999) and 0.78 (Peery et al. 2012), respectively, and a variance in the mean size of multi-repeat mutations of 12 in both analyses (Peery et al. 2012; Piry et al. 1999). MS indicates presence (+) or absence (-) of mode-shift distortion of allele frequencies (see figure S3)

	Middle spotted woodpecker				Great spotted woodpecker			
Population	TPM 0.95	TPM 0.78	SMM	MS	TPM 0.95	TPM 0.78	SMM	MS
AG	0.034	0.002	0.040	+	0.998	0.954	0.999	-
BL	0.005	0.001	0.055	-	0.974	0.867	0.998	-
HE	0.227	0.169	0.420	-	0.997	0.961	0.999	-
NE	0.088	0.004	0.190	+	0.998	0.983	0.999	-
SH	0.047	0.011	0.073	-	0.998	0.979	0.999	-
TG	0.580	0.368	0.729	-	0.999	0.995	0.999	-
ZH	0.055	0.017	0.055	-	0.974	0.974	0.996	-

Table S8 Number of middle and great spotted woodpecker individuals assigned to genetic clusters (C) inferred from DAPC with prior information on and in relation to the sampling locations. Individuals were assigned to the cluster for which their posterior probability was >0.5. Column UA gives the number of unassigned individuals (posterior probability ≤ 0.5) per sampling location and analysis. **Bold print** indicates the cluster solutions discussed in the main text

	MSW										GSW									
	K = 3					**K = 4**					K = 3					**K = 3**				
	30 PCs					**50 PCs**					30 PCs					**45 PCs**				
Sampling location	n	C1	C2	C3	UA	**C1**	**C2**	**C3**	**C4**	**UA**	n	C1	C2	C3	UA	**C1**	**C2**	**C3**	**UA**
Aargau AG	14	0	0	0	14	**13**	**0**	**0**	**1**	**0**	24	13	1	0	10	**19**	**1**	**0**	**4**
Basel BL	19	0	2	3	14	**16**	**0**	**0**	**2**	**1**	21	0	2	0	19	**1**	**1**	**0**	**19**
Hessen HE	16	13	0	1	2	**0**	**14**	**0**	**2**	**0**	21	1	1	1	18	**0**	**0**	**0**	**21**
Neuenburg NE	18	0	17	1	0	**0**	**0**	**18**	**0**	**0**	21	1	13	3	4	**0**	**16**	**1**	**4**
Schaffhausen SH	18	2	0	1	15	**1**	**1**	**0**	**16**	**0**	20	1	0	0	19	**2**	**0**	**0**	**20**
Thurgau TG	15	0	0	11	4	**0**	**0**	**0**	**14**	**1**	21	0	0	1	20	**0**	**0**	**1**	**20**
Zürich ZH	16	0	0	11	5	**1**	**0**	**0**	**14**	**1**	19	1	0	12	6	**0**	**0**	**15**	**3**

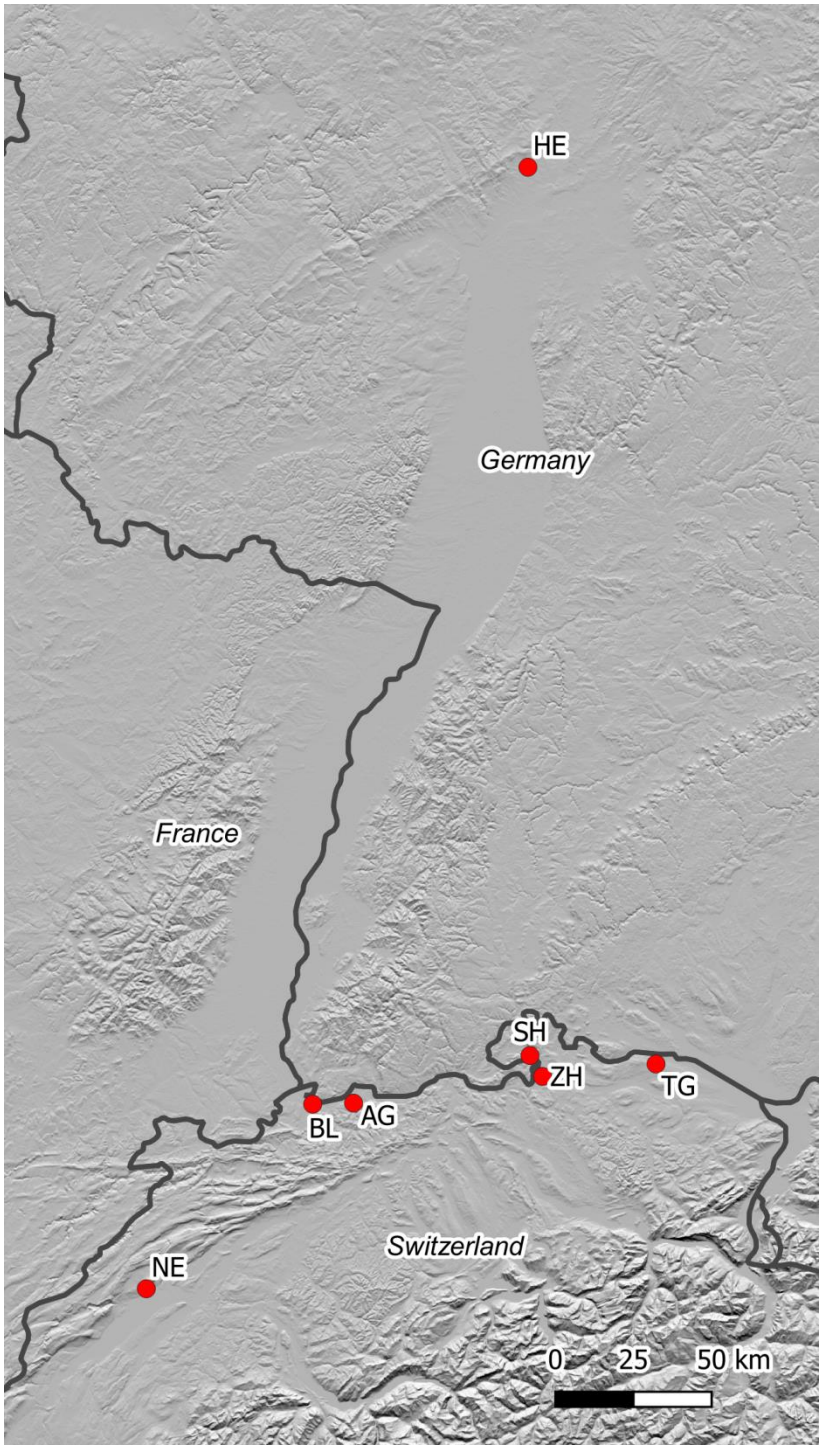

Figure S1 Locations of sampled local populations of middle and great spotted woodpeckers. AG = Aargau, BL = Basel-Landschaft, HE = Hessen, NE = Neuenburg, SH = Schaffhausen, TG = Thurgau, ZH = Zürich

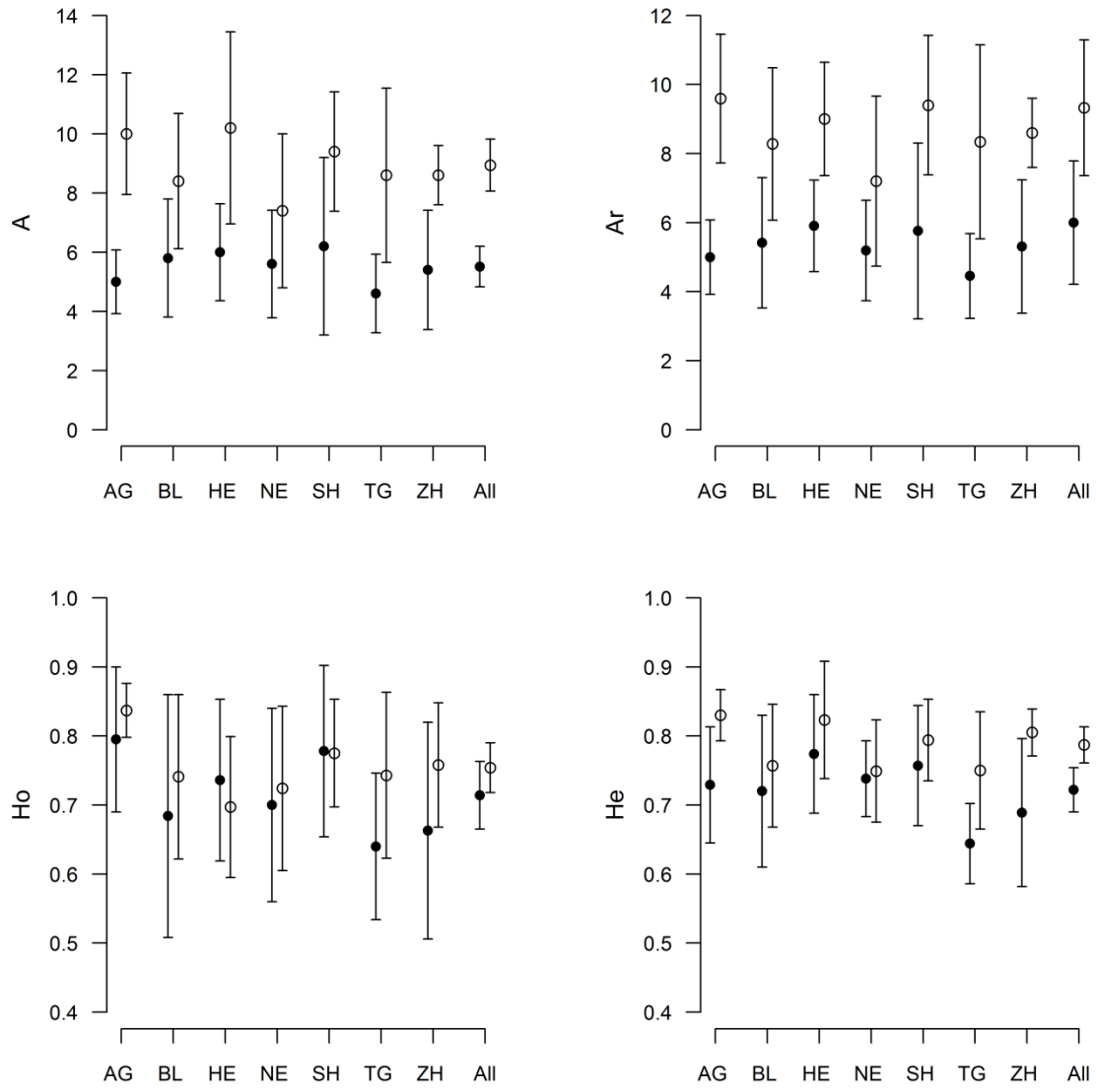

Figure S2 Measures of genetic diversity in middle spotted woodpeckers (closed circles) and great spotted woodpeckers (open circles) based on the five markers developed in the middle spotted woodpecker and typed in both species. Bars give 95% confidence intervals. A = number of alleles, A_r = allelic richness, H_o = observed heterozygosity, H_e = expected heterozygosity

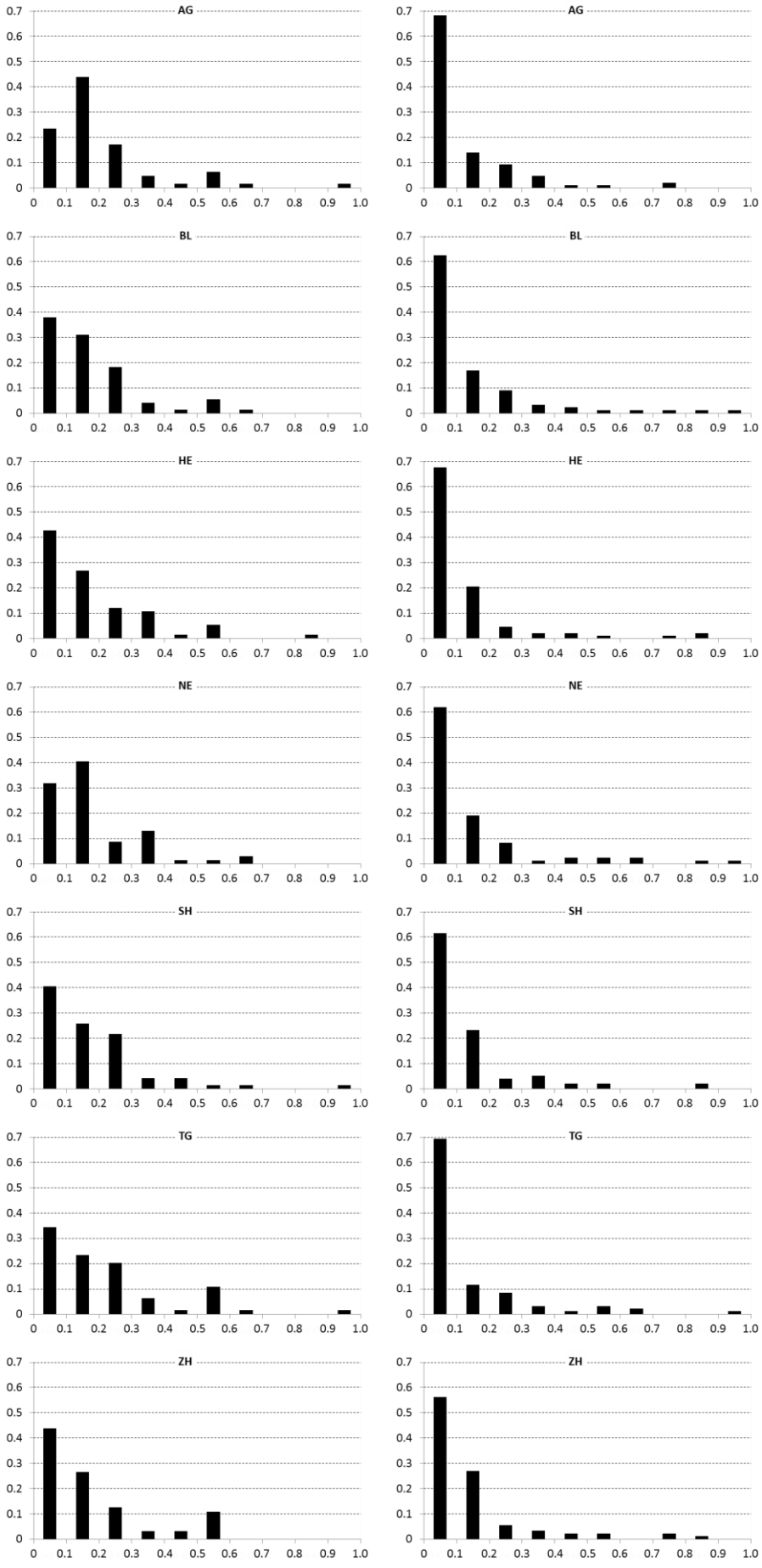
AG
AG
BL
BL
HE
HE
NE
NE
SH
SH
TG
TG
ZH
ZH

Figure S3 Allele frequency distributions from the seven populations of the middle spotted woodpecker (left panels) and the great spotted woodpecker (right panels). In each panel, the x-axis gives the allele frequency class, the y-axis shows the proportion of alleles. A population is considered to have undergone a recent bottleneck if fewer alleles are found in the lowest allele frequency class (0-0.1) than in one or more intermediate allele frequency classes (Luikart et al. 1998). Accordingly, two populations of the middle spotted woodpecker (AG, NE) experienced a recent bottleneck, while all the other populations of both species had the largest proportion of alleles in the lowest allele frequency class, suggesting absence of recent bottlenecks

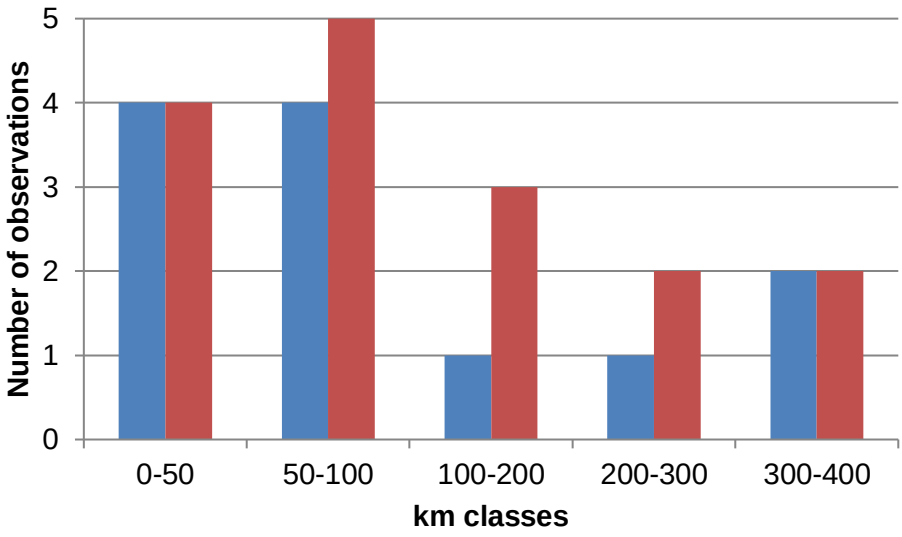

Figure S4 Number of first-generation migrant events in relation to distance in the middle spotted woodpecker (blue bars) and in the great spotted woodpecker (red bars) according to results from GENECLASS2

Text Appendix 1 DNA extraction, microsatellite markers and genotyping

Genomic DNA was extracted from the blood samples using an automated magnetic-particle protocol (BioSprint 96 DNA Blood kit; QIAGEN, Hilden, Germany) while the feather samples were processed with a spin column protocol (DNeasy Blood & Tissue Kit, QIAGEN).

Two sets of primer pairs previously developed in Picine species (Vila et al. 2008, Ellegren et al. 1999) were tested for cross amplification and polymorphism in the sampled populations. Additionally, new markers were developed for the great spotted woodpecker (Ecogenics, Zurich, Switzerland): size selected fragments from genomic DNA were enriched for SSR content by using magnetic streptavidin beads and biotin-labeled CT and GT repeat oligonucleotides. The SSR enriched library was analyzed on a Roche 454 platform using the GS FLX titanium reagents. The sequence reads were screened for microsatellite inserts with a tetra- or a trinucleotide of at least 6 repeat units or a dinucleotide of at least 10 repeat units. Primer sets were designed and a set of markers screened for polymorphisms, of which seven markers were selected for genotyping (see Table S1 for details).

PCR amplifications were carried out in five multiplex and one singleplex reactions. The total reaction volume of 6µl comprised approximately 30ng of genomic DNA, 3µl of 2x Multiplex Master Mix (Qiagen) and equal amounts of forward and reverse primers varying in concentration between 0.1 and 0.2 µM (see Table S1 for details). Thermal profiles consisted of an initial denaturation step of 94 °C for 15 min followed by 25 cycles of 94 °C for 30s, 55 to 62°C for 90s (see Table S1 for details on each reaction) and 72°C for 30s, followed by a final extension at 60°C for 30 min. Fragment analyses were performed on a 3730 DNA Analyzer using Genescan-500 LIZ (Thermo Fisher Scientific, Waltham, USA) size standard. Electropherograms were analyzed with Genemapper 4.0 (Thermo Fisher Scientific) and GeneMarker 2.2.0 (Softgenetics, State College, USA).

All of the twelve microsatellite markers developed by Vila et al. (2008) specifically for the middle spotted woodpecker worked in this species. Out of these twelve markers, seven were

tested in the great spotted woodpecker and six out of seven worked (Table S2). Three out of six microsatellite markers developed for the white-backed woodpecker (Ellegren *et al.*, 1999), worked for both the middle and the great spotted woodpecker. The seven newly developed microsatellite markers for the great spotted woodpecker (GenBank accession numbers: JX196852-JX196858) faultlessly worked in that species and were not tested in the middle spotted woodpecker.

Text appendix 2 Detailed marker-specific results on null alleles, HWE and linkage disequilibrium tests

Marker-specific results across sampling sites are given in Table S3. With the exception of markers DMC1 and DMD7 in BL in the middle spotted woodpecker and marker DMA5890 in HE in the great spotted woodpecker, all markers in all sampling sites did not significantly deviate from Hardy-Weinberg equilibrium (HWE) in both species (after sequential Bonferroni correction). Over all markers, six of the seven populations of the middle spotted woodpecker and all seven populations of the great spotted woodpecker were in HWE (Table 1 in the main text).

In the middle spotted woodpecker, null alleles were suggested to be present at marker Dlu4 in four (AG, BL, HE, TG) of the seven local populations (Table S3), at marker DMC1 in two local populations (AG, BL) and at three further markers each in one local population (DMD7 in BL, DMD9 in NE, DMA2 in HE). Marker Dlu4 was excluded from all following analyses, but all other markers were retained due to inconsistent patterns of null allele occurrence across markers and local populations.

In the great spotted woodpecker, null alleles appeared to be present at marker Dlu1 in all seven local populations (Table S3), at Dlu4 in five local populations (AG, HE, SH, TG, ZH), at DMA7242 in four local populations (AG, BL, NE, TG) and at six further markers in different local populations (DMC111 in ZH, Dlu6 in AG, DMC115 and DMA5890 in HE, DMA2693 in SH, DMA2164 in TG). Markers Dlu1, Dlu4 and DMA7242 were excluded as was the monomorphic marker DMA4, the others retained.

Of 510 and 451 pairwise locus comparisons, significant linkage disequilibrium was detected for 17 and 15 comparisons in the middle spotted woodpecker and the great spotted woodpecker, respectively. However, following Bonferroni corrections these tests were no longer significant (all $p > 0.05$).

Text Appendix 3 Identifying the numbers of PCs for DAPC

In contrast to STRUCTURE, DAPC is a multivariate method that does not make any assumptions about the underlying population genetic model.

Prior to the DAPC, the number of principal components (PCs) was determined for each species with two approaches. In the middle spotted woodpecker, 30 PCs was suggested to be optimal in 10 analyses regarding the trade-off between power of discrimination and over-fitting in relation to the number of PCs considered (a-score approach, see chapters 4.1. and 4.2 in the tutorial of the adegenet package, Jombart and Collins, 2015), and the proportion of conserved variance was 0.82. In the second approach, 50 PCs was the number most often reaching the lowest RMSE value in 20 cross-validations (training set = 90% of total data, validation set = 10% of total data, 100 replicates; see tutorial in the adegenet package for details). With 50 PCs, the proportion of conserved variance was 0.96. Both DAPCs were run with 30 and 50 PCs, respectively. Because the results were quite similar, only those with 50 PCs are presented.

In the great spotted woodpecker, the ten a-score analyses suggested retaining 45 PCs (proportion of conserved variance 0.87). In 20 cross-validations (second approach), the lowest RMSE value was achieved eight times with 20 PCs (0.63) and nine times with 30 PCs (0.76). Because the proportion of conserved variance was low for 20 PCs, DAPCs were run with the first 30 and 45 PCs, respectively. Results were again very similar and only those based on DAPCs with 45 PCs are presented below.

Text Appendix 4 Results of STRUCTURE analyses without *a priori* information on sampling locations

Without *a priori* information on sampling locations and irrespective of the settings for the prior and alpha, STRUCTURE most often suggested the presence of one genetic cluster, i.e. K = 1, as the most likely solution in both woodpecker species (Fig. 3 in the main text). This was supported by plots of the proportional membership *Q* for the different values of *K* > 1: each individual was assigned to each genetic cluster to nearly the same extent ($Q \approx 1/K$), so no meaningful spatial clustering was evident.

Using the multivariate approach to evaluate genetic structure implemented in ADEGENET, K = 2 was suggested as the best solution in the middle spotted woodpecker (based on both lowest BIC and strongest decrease in BIC) when retaining 50 PCs, and either K = 2 (strongest decrease in BIC) or K = 3 (lowest BIC) when retaining 30 PCs. In each case, however, clusters consisted of a mixture of individuals from different populations, with no clear pattern evident. In the great spotted woodpecker, the best solution was suggested to be K = 2 based on the strongest decrease in BIC and K = 3 based on the lowest BIC, respectively. In both cases and irrespective of retaining 30 PCs or 45 PCs, clusters consisted of a mixture of individuals from different population, and again no clear pattern was evident.