



# Evaluation of genotyping by sequencing for population genetics of sibling and hybridizing birds: an example using Syrian and Great Spotted Woodpeckers

Artur Gurgul<sup>1</sup> · Anna Miksza-Cybulska<sup>2</sup> · Tomasz Szmatola<sup>1</sup> · Ewelina Semik-Gurgul<sup>1</sup> · Igor Jasielczuk<sup>1</sup> · Monika Bugno-Poniewierska<sup>1</sup> · Tomasz Figarski<sup>3</sup> · Łukasz Kajtoch<sup>3</sup>

Received: 30 January 2018 / Revised: 1 October 2018 / Accepted: 24 October 2018 / Published online: 10 November 2018  
© The Author(s) 2018

## Abstract

Recent high-throughput genotyping technologies allow for comprehensive genomic analyses on an unprecedented scale. However the advantages of the most commonly used tools are strongly limited in non-model organisms, including wild birds. In this study we attempt to test the utility of genotyping by sequencing (GBS) without relying on the reference genome sequence in selected pairs of sibling bird species that are known to hybridize (Syrian and Great Spotted Woodpeckers). We found that GBS is able to produce a satisfying number of polymorphisms and can be successfully applied to the population genetics of these species. The results also suggest that urban populations of these woodpeckers, especially phenotypic Syrian Woodpeckers, consist of individuals that harbor genotypes assigned to Great Spotted Woodpeckers, which suggests intensive hybridization and introgression.

**Keywords** *Dendrocopos major* · *Dendrocopos syriacus* · Next-generation sequencing · Hybrids · Polymorphism

## Zusammenfassung

### Die Eignung von Genotypisierung mittels Sequenzierung für die Populationsgenetik von Zwillings- und Hybridarten am Beispiel von Blut- und Buntspecht

Moderne Hochdurchsatz-Genotypisierungstechnologien ermöglichen umfassende Genomanalysen in nie dagewesenem Ausmaß. Allerdings sind die Vorzüge der gängigsten Methoden stark limitiert, sobald diese außerhalb von Modellorganismen (z. B. bei Wildvögeln) angewendet werden. In der vorliegenden Studie versuchen wir, die Anwendbarkeit von Genotypisierung mittels Sequenzierung (GBS) ohne Bezug auf ein Referenzgenom zu testen, bei einem Zwillingsartenpaar, Blut- und Buntspechte, die bekanntermaßen hybridisieren. Wir fanden, dass GBS eine zufriedenstellende Anzahl von Polymorphismen produziert und erfolgreich in der Populationsgenetik dieser Arten eingesetzt werden kann. Die erzielten Ergebnisse deuten auch darauf hin, dass städtische Populationen dieser Spechte, insbesondere phänotypischer Blutspechte, aus Individuen

---

Communicated by M. Wink.

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10336-018-1601-9>) contains supplementary material, which is available to authorized users.

---

✉ Artur Gurgul  
artur.gurgul@izoo.krakow.pl

<sup>3</sup> Institute of Systematics and Evolution of Animals Polish Academy of Sciences, Sławkowska 17, 31-016 Krakow, Poland

<sup>1</sup> Department of Animal Molecular Biology, National Research Institute of Animal Production, Krakowska 1, 32-083 Balice, Poland

<sup>2</sup> Department of Sheep and Goat Breeding, National Research Institute of Animal Production, Krakowska 1, 32-083 Balice, Poland

bestehen, die Buntspecht-Genotypen aufweisen, was möglicherweise auf intensive Hybridisierung und Introgression hindeutet.

## Introduction

One of the most commonly utilized and powerful genotyping tools is the single nucleotide polymorphism (SNP) genotyping microarray, which allows one to establish genotypes of millions of SNPs in parallel. However, the microarray design process requires prior knowledge of SNPs surrounding a DNA sequence and needs to be preceded by high-throughput SNP detection techniques (Trevino et al. 2007). Furthermore, microarray design for minor experiments or initial studies, e.g., in wild species, is often economically unjustified. A solution for this may be the application of recently developed genotyping methods that are based on the enrichment of genome sequences associated with restriction enzyme (REs) digestion sites and subsequent utilization of a next-generation sequencing (NGS) technique. An RE-based enrichment approach allows for sequencing only a fraction of the genome and achieves higher multiplexing, significantly reducing experimental costs compared to whole genome sequencing (Elshire et al. 2011). One of the available RE-based enrichment and genotyping techniques is genotyping by sequencing (GBS). An important advantage of GBS is that the derived sequences can be used for polymorphism detection and genotyping without locating them in the reference sequence, and thus permits whole genome genotyping in practically any organism (Elshire et al. 2011). GBS has been shown to be highly useful in animal genomics (He et al. 2014), especially in species for which there are no commercially available microarray tools.

In this study, we attempt to test the performance of a standard, single-digest GBS protocol in selected bird taxa [Syrian Woodpecker *Dendrocopos syriacus* (SW), and Great Spotted Woodpecker *Dendrocopos major* (GW)], which are known to be closely related and able to hybridize (Michalczyk et al. 2014; Figarski and Kajtoch 2018), without relying on the reference genome sequence. We evaluated the detected SNPs in terms of their variability, quality and application in population genetics of these species, and tested the application of GBS to the detection of their hybrids.

## Materials and methods

### Sampled birds

Two sibling *Dendrocopos* woodpeckers were recently examined with the use of various genetic markers for species delimitation and identification of their hybrids (Michalczyk

et al. 2014). Sampling in this study also followed previous research, with slight modifications, and comprised nine individuals of SW, two hybrids (possible backcrosses), eight individuals of GW collected from their sympatric range in Poland, and additionally three individuals of GW from an allopatric population in Norway (Table 1). The results of GBS for three of the SW and one of the GW individuals from Poland were found to be inadequate, so the final dataset used for analyses comprised a total of 18 specimens.

### GBS library preparation and sequencing

DNA was purified from muscle or down feathers of woodpeckers with the use of a Nucleospin kit (Macherey–Nagel, Düren, Germany).

GBS libraries were prepared after a few optimization steps, following previously described methodology (Elshire et al. 2011). The protocol included *Pst*I RE digestion and ligation of 48 indexed adapters designed with GBS Barcode Generator software (<http://www.deenabio.com/services/gbs-adapters>). In brief, during the library preparation process, commercially synthesized oligonucleotides (forward and reverse complement) containing unique barcodes (4–9 nucleotides long) and a common adapter were annealed to obtain ligation-ready double-stranded oligonucleotides. Genomic DNA (200 ng) was digested overnight with a ten-fold *Pst*I-*HF* enzyme excess (100,000 U/ml; New England Biolabs) at 37 °C. Digestion products were then ligated with 4.8 ng of each adapter using T4 DNA ligase (400 U/μl; New England Biolabs) at 22 °C for 60 min. After clean-up, fragments with ligated adapters were amplified in 18 cycles of PCR, using universal primers (12.5 pmol/μl each) and 2×Taq Master Mix (New England Biolabs). Purified and diluted libraries were controlled for quality using Agilent's TapeStation2200 system (Agilent Technologies) and quantified using DNA binding dye (Qubit DS DNA assay; Thermo Fisher Scientific). Oligonucleotide sequences used here are the same as those used in our previous study (see Gurgul et al. 2018).

Library pools were ultimately sequenced in a single-end 50-bp run on a HiScanSQ system (Illumina, San Diego, CA) using a single lane of TruSeq v3 HiSeq flowcell and v3 reagents (Illumina). Raw, demultiplexed, unfiltered reads were deposited in the NCBI Sequence Read Archive (SRA) database under BioProject number PRJNA397002.

### SNP detection, filtering and sample differentiation analysis

The obtained raw reads were assessed for quality using FastQC software and analyzed with the use of recently developed GBS SNP Calling Reference Optional Pipeline (GBS-SNP-CROP) (Melo et al. 2016) dedicated to species

**Table 1** General characterization of samples used for genotyping by sequencing (GBS) of selected taxa (Syrian Woodpecker *Dendrocopos syriacus*, Great Spotted Woodpecker *Dendrocopos major* and hybrid *D. syriacus* × *D. major*)

<i>Dendrocopos</i> species	Location	Phenotype	Genotype		GBS symbol
			mtDNA	nucDNA	
<i>D. syriacus</i>	Poland-Krakow	Syriacus-like	SW	SW	SW1
<i>D. syriacus</i>	Poland-Krakow	Syriacus-like	SW	SW/GW?	SW2
<i>D. syriacus</i>	Poland-Krakow	Syriacus-like	SW	SW	SW3
<i>D. syriacus</i>	Poland-Krakow	Syriacus-like	SW	GW?	SW4
<i>D. syriacus</i>	Poland-Krakow	Syriacus-like	SW	SW	SW6
<i>D. syriacus</i>	Poland-Krakow	Syriacus-like	SW	SW	SW9
<i>D. syriacus</i> × <i>D. major</i>	Poland-Krakow	Intermediate	GW	SW/GW	SW12
<i>D. syriacus</i> × <i>D. major</i>	Poland-Krakow	Intermediate	GW	SW/GW	GW1
<i>D. major</i>	Norway	Major-like	GW	GW	GW3
<i>D. major</i>	Norway	Major-like	GW	GW	GW4
<i>D. major</i>	Norway	Major-like	GW	GW	GW5
<i>D. major</i>	Poland-unknown	Major-like	GW	GW	GW6
<i>D. major</i>	Poland-unknown	Major-like	GW	GW	GW7
<i>D. major</i>	Poland-Krakow	Major-like	GW	GW	GW9
<i>D. major</i>	Poland-Bieszczady	Major-like	GW	GW	GW10
<i>D. major</i>	Poland-Krakow	Major-like	GW	GW	GW11
<i>D. major</i>	Poland-Krakow	Major-like	GW	GW	GW12
<i>D. major</i>	Poland-Krakow	Major-like	GW	GW	GW2

Individuals were assigned to species based on their phenotype (plumage) and genotype [following methods presented by Michalczyk et al. (2014)]

mtDNA Mitochondrial DNA control region, nucDNA nuclear DNA transforming growth factor intron 5, ? uncertain genotypes

with no reference genome sequence available. GBS-SNP-CROP handles the lack of reference genome sequences by building a mock reference sequence based on collapsed reads and clustering strategy. The software was set to a minimum read length of 40 bp and other settings recommended for diploid genomes (Melo et al. 2016). Moreover, to reduce the number of false positives, the SNP matrix was further reduced by applying filters for read depth ( $n=4$ ), a minimum number of reads supporting each allele for heterozygote calls ( $n=2$ ) and other parameters allowing us to achieve SNP confidence at the 95% level (Melo et al. 2016).

Detected SNPs were additionally filtered to remove those with minor allele frequency (MAF) lower than 0.01, missing genotypes in more than 20% of individuals and deviating from Hardy–Weinberg equilibrium with  $P < 0.001$ . Moreover, individuals with more than 20% missing genotypes (four) were also removed. Supporting SNPs analyses were performed using PLINK (Purcell et al. 2007) or TASSEL (Bradbury et al. 2007) software.

To evaluate the performance SNP in a population differentiation analysis, the inter-individual distance matrix was calculated as: 1—identity by state (IBS) similarity, with IBS defined as the probability that alleles drawn at random from two individuals at the same locus are the same. The obtained pairwise IBS distances were then used to create cladograms based on a neighbor-joining (NJ) method (Saitou and Nei

1987) and visualized with the use of a modified version of Archaeopteryx Tree software (Han and Zmasek 2009). After SNP genotypes conversion to sequences, following Lee et al. (2014), a phylogenetic tree was constructed using a maximum likelihood (ML) method along with a bootstrap test implemented in MEGAX software (Kumar et al. 2018). The tree was inferred based on the Jukes-Cantor model (Jukes and Cantor 1969). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein 1985) is presented. Initial tree(s) for the heuristic search were obtained by applying the NJ method to a matrix of pairwise distances estimated using the maximum composite likelihood approach. The additional visualization of population differentiation was made by a multidimensional scaling (MDS) method applied to the IBS distance matrix. To further validate the population structure and the patterns of admixture among populations, Structure software (Falush et al. 2003) was run ten times per each  $K$ , initially from  $K=2$  and up to  $K=4$ . An admixture model was used and the parameters of the runs were as follows: 200,000 iterations and 100,000 burn-in period. Then, the Clumpak software (Kopelman et al. 2015) was used to solve the problems of label switching and genuine multimodality (Jakobsson and Rosenberg 2007) and to visualize the results.

## Results

### PstI-generated library fragment size distribution

The presence of DNA fragments that originated from enzyme digestion within the repetitive genomic DNA sequences was evaluated based on Agilent TapeStation2200 system traces. *PstI* endonuclease generated a fairly uniform fragment distribution for woodpeckers (Supplementary File 1).

### Sequencing reads and SNP statistics

The applied sequencing depth allowed the generation of, on average, 862,900 good barcoded reads per sample (Table 2).

With the use of the GBS-SNP-CROP pipeline, 3329 of biallelic SNPs were detected with average read coverage across a population of 19.7 (Supplementary File 2). The final set of SNPs was obtained by applying additional population-level filtering criteria in terms of polymorphisms and missing genotypes. This resulted in a set of 12,479 filtered SNPs. The mean MAF across all variants was 0.242 ( $\pm 0.129$ ). The MAF distribution analysis showed that 46.9% of SNPs showed MAF up to 0.2 and 81.3% SNPs had MAF up to 0.4. The observed heterozygosity averaged for all SNPs was high (0.475). The average call rate per SNP was 90.1%. When both species were analyzed separately, the average MAF for the GW population (0.251) was slightly higher than that for the SW one (0.231) and the averaged observed heterozygosity was slightly higher in GW samples (0.498 vs. 0.461). The percentage of missing genotypes was lower for SW than for GW (0.09 vs. 0.08).

Raw, demultiplexed, unfiltered sequencing reads were deposited in the NCBI SRA database under BioProject no. PRJNA397002.

### SNP performance in population differentiation analysis

The assignment of woodpeckers to clusters with the use of GBS resulted in different outputs from NJ and MDS methods. What is common to both approaches is that they showed that GW had much greater genetic diversity than SW. According to the NJ cladogram there are two clusters that generally correspond to GW and SW; however, in the first one, some individuals seem to have intermediate genotypes (Fig. 1a). GW individuals originating in Norway and Poland formed clearly separated clusters, with one of the analyzed hybrids clustering near the Norwegian samples. In the MDS plot only some analyzed individuals of SW form a clearly separated cluster, whereas two others could be assigned as genetically closer to GW and had genotypes highly similar to those found in the analyzed hybrids. Moreover, GW is divided into two weakly separated groups, of which one encompasses some individuals from Poland and the second birds from Poland and Norway (Fig. 1b).

The results of population structure and degrees of admixture of the populations identified with Structure software are presented in Fig. 2a. The highest likelihood was obtained for  $K=2$ . However, the likelihood for  $K=3$  and  $K=4$  was very similar and presented exactly the same patterns as  $K=2$  suggesting it as the most optimal result of population structure of the studied individuals. The results obtained with Structure software assigned the analyzed hybrids as well as two other SW individuals (which were clustered together with hybrids using MDS) to the GW population (Fig. 2a). Very similar results were obtained when the maximum likelihood method was applied to the construction of the cladogram (Fig. 2b). In this analysis, both hybrids and the two SW individuals with an uncertain nuclear DNA (Table 1) genotype

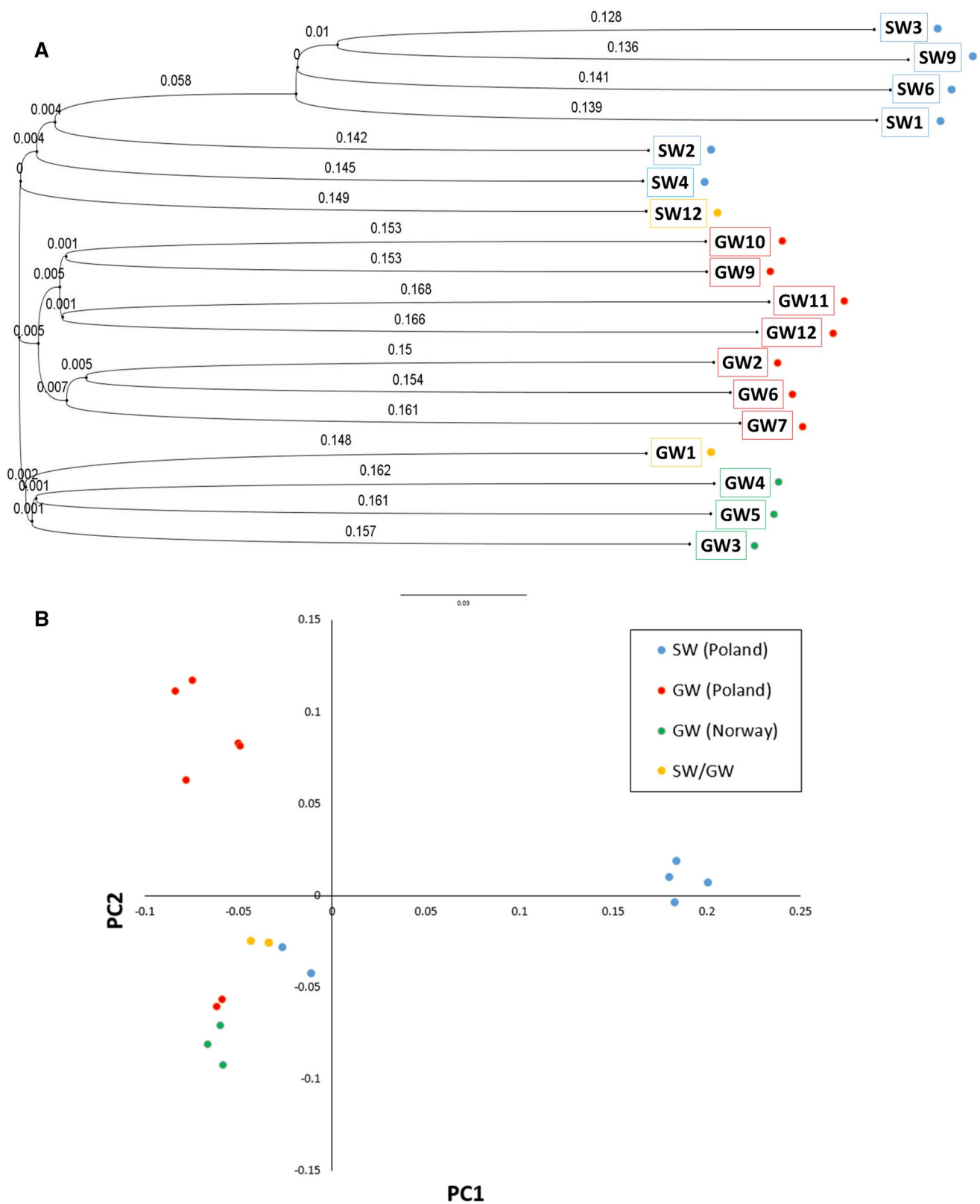
**Table 2** Sequencing and single nucleotide polymorphism (SNP) statistics

Species	Individuals	Number of species	Average number of good barcoded reads per individual ( $K$ )	No. of initial variants	No. of filtered SNPs	Average read depth at SNPs across the population	Average MAF	$H_o$	Frequency of missing genotypes per SNP (average)
All	22 <sup>a</sup>	2	862.9	3329	2479	19.66 ( $\pm 23.11$ )	0.242 ( $\pm 0.129$ )	0.475	0.099
SW	6	1					0.227 ( $\pm 0.163$ )	0.448	0.102
GW	10						0.251 ( $\pm 0.161$ )	0.498	0.090
SW <sup>b</sup>	4 <sup>b</sup>	1	–	–	–	–	0.231 ( $\pm 0.207$ )	0.461	0.079

MAF Minor allele frequency,  $H_o$  observed heterozygosity; for other abbreviations, see Table 1

<sup>a</sup>Sequencing and SNP discovery statistics were calculated for all studied animals. Population indexes were calculated for 18 individuals after removing four individuals with a high ratio of missing genotypes

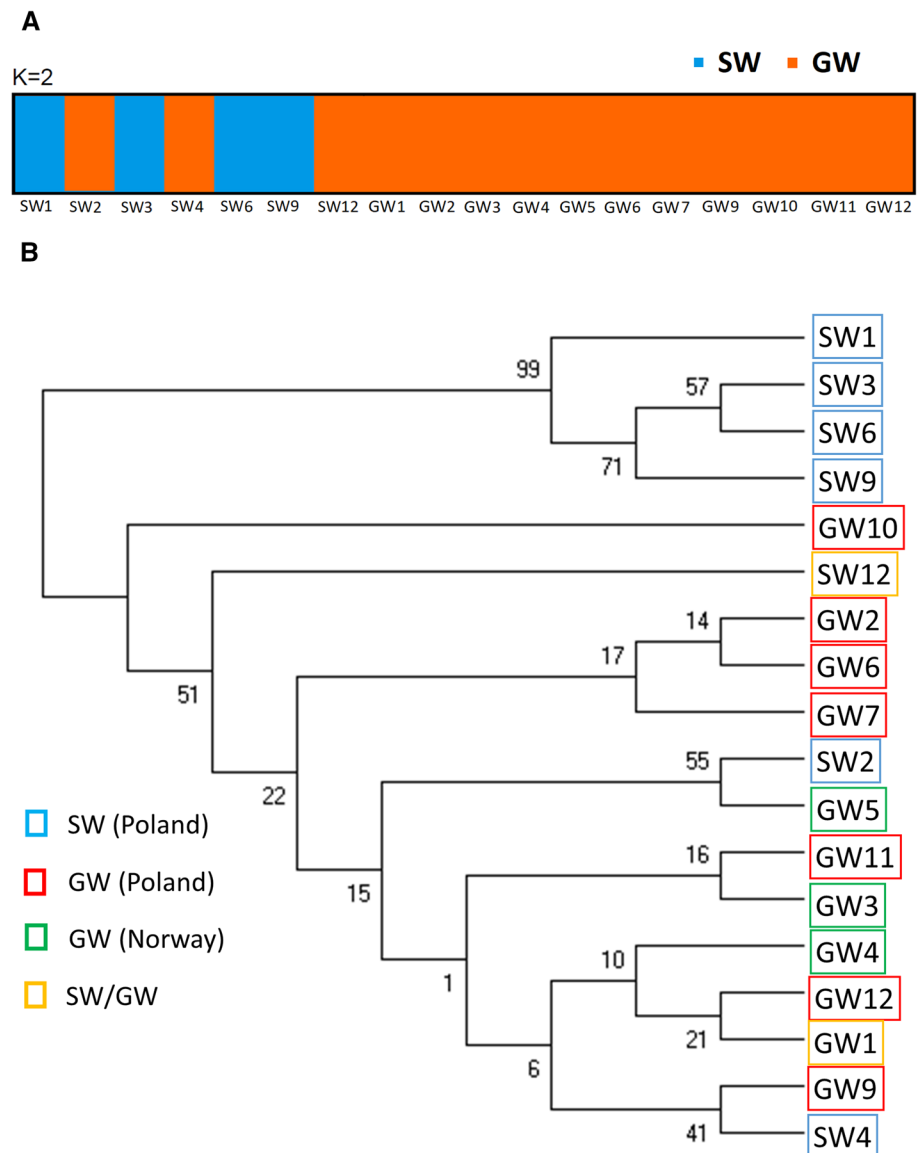
<sup>b</sup>With the exclusion of individuals with uncertain ancestry after GBS analysis (SW2, SW4)



**Fig. 1** **a** Cladogram for the studied woodpecker individuals/species based on identity by state (IBS) and the neighbor-joining method; **b** multidimensional scaling plot (MDS) based on the IBS distance matrix for the studied woodpeckers (GW *Dendrocopos major*, SW

*Dendrocopos syriacus*, SW/GW hybrids). The geographic origin of each individual is given in *parentheses*. Branches lengths are provided

**Fig. 2** **a** Graphical representation of individual contributions obtained with Structure software for  $K=2$ ; **b** Cladogram generated using maximum likelihood (ML) method. Geographic origin of each individual is given in *parentheses*. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown *next to the branches*. For abbreviations, see Fig. 1



were randomly clustered with the GW population, while the remaining SW individuals formed a clearly separated cluster. No visible subdivision amongst individuals according to their geographical origin could be seen in the ML results.

## Discussion

Previous studies have confirmed the usefulness of GBS for domestic species of birds (Pértille et al. 2016; Zhu et al. 2016). However, this technique has been scarcely applied to wild birds that lack a reference genome sequence (see review in Toews et al. 2016). Thus, in this study, we attempt to detect and genotype a panel of SNPs in selected pairs of sibling woodpeckers and their hybrids sufficient for the purposes of population genetics. With the applied sequencing

depth and filtering criteria we were able to create a panel of 2479 informative SNPs for these species.

Genotyping of SW and GW woodpeckers with the use of GBS resulted in patterns only partially concordant with previous species delimitation and hybrid identification studies (Michalczuk et al. 2014). In Michalczuk et al. (2014) the usage of a set of sequence-based markers resulted in reliable delimitation of these species. The most variable marker was found to be the control region of mtDNA, which includes several discriminating SNPs. Discriminating SNPs were also found in each of three autosomal and one sex-lined intron, but the majority of the others were common to both species. Finally, these sequence-based markers were insufficient for the definitive identification of hybrids as the number of unlinked SNPs was too low. Moreover, the use of microsatellites was found



to be very problematic, as only a few loci could be amplified and were simultaneously polymorphic in both species. However, these microsatellites allowed for the assignment of some specimens as possible hybrids (Michalczuk et al. 2014). GBS resulted in a quite unexpected pattern as it showed that some phenotypic SW and both phenotypic hybrids possessed genotypes that are clustered within GW genotypes.

GBS is expected to be much more powerful than mtDNA sequences or microsatellites for the identification of species and their hybrids. Assignment of some phenotypic SW and all hybrids to a genetic pool characteristic of GW suggests that hybridization between GW and SW is a much more frequent phenomenon than expected based on previous genetic studies (Michalczuk et al. 2014) and field observations (Figarski and Kajtoch 2018). Apparently, ongoing hybridization, including backcrossing with GW, has resulted in a high reduction of characteristic SW alleles in the genomes of birds from sympatric populations. It is important to highlight that the described pattern is characteristic of urban populations, where contacts between SW and GW are quite common as GW occupies parks and larger wooded areas in the vicinity of orchards and avenues of trees occupied by SW (Kajtoch and Figarski 2017). It would be interesting to verify the genomics of woodpeckers from rural populations where SW only rarely encounters and mates with GW [SW occupies mostly orchards, whereas GW mainly occupies forests (Kajtoch and Figarski 2017)]. If this pattern is confirmed from larger samples from different populations, it would mean that a large number of phenotypic SW have GW-like genome characteristics. This could have serious consequences for any ecological, ethological or biological studies on these woodpeckers, as plumage may not be a reliable indicator of species affinity of birds observed in the field.

Our findings indicate that GBS could have very wide applications in ornithology, probably covering not only species delimitation and population genetics, but also hybridization and parental studies.

**Acknowledgments** We are extremely grateful to J. Ranke, “Dzika Klinika” (J. D. Wójcik) and “Vetika” (K. Ptak) for providing woodpecker samples. This paper has greatly benefited from the constructive comments of the anonymous reviewers. The study was financed by funds of the National Research Institute of Animal Production (research project number 04-010.1), and statutory research of the Institute of Systematics and Evolution of Animals of the Polish Academy of Sciences.

## Compliance with ethical standards

**Ethical standards** This study complies with current Polish and international laws.

**Conflict of interest** The authors declare no conflict of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6(5):e19379
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791
- Figarski T, Kajtoch Ł (2018) Hybrids and mixed pairs of Syrian and Great Spotted Woodpeckers in urban populations. *J Ornithol* 159:311–314
- Gurgul A, Miksza-Cybulska A, Szmatoła T, Jasielczuk I, Piestrzyńska-Kajtoch A, Fornal A, Semik-Gurgul E, Bugno-Poniewierska M (2018) Genotyping-by-sequencing performance in selected livestock species. *Genomics*. <https://doi.org/10.1016/j.ygeno.2018.02.002>
- Han MV, Zmasek CM (2009) phyloXML: xML for evolutionary biology and comparative genomics. *BMC Bioinform* 10:356
- He J, Zhao X, Laroche A, Lu ZX, Liu H, Li Z (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front Plant Sci* 5:484
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) *Mammalian protein metabolism*. Academic Press, New York, pp 21–132
- Kajtoch Ł, Figarski T (2017) Comparative distribution of Syrian and Great Spotted Woodpeckers in different landscapes of Poland. *Folia Zool* 66:29–36
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Res* 15:1179–1191
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549
- Lee TH, Guo H, Wang X, Kim C, Paterson AH (2014) SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genom* 15:162
- Melo AT, Bartaula R, Hale I (2016) GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC Bioinform* 17:29
- Michalczuk J, McDevitt A, Mazgajski TD, Figarski T, Ilieva M, Bujoczek M, Malczyk P, Kajtoch Ł (2014) Tests of multiple molecular markers for the identification of Great Spotted and Syrian Woodpeckers and their hybrids. *J Ornithol* 3:591–600
- Pértille F, Guerrero-Bosagna C, Henrique da Silva V, Boschiero C, de Ribamar da Silva Nunes J, Corrêa Ledur M, Jensen P, Lehmann

- Coutinho L (2016) High-throughput and cost-effective chicken genotyping using next-generation sequencing. *Sci Rep* 6:26929
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 81:559–575
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Toews DPL, Campagna L, Taylor SA, Balakrishnan CN, Baldassarre DT, Deane-Coe PE, Harvey MG, Hooper DM, Irwin DE, Judy CD, Mason NA et al (2016) Genomic approaches to understanding population divergence and speciation in birds. *Auk Ornithol Adv* 133:13–30
- Trevino V, Falciani F, Barrera-Saldaña HA (2007) DNA microarrays: a powerful genomic tool for biomedical and clinical research. *Mol Med* 13:527–541
- Zhu F, Cui Q-Q, Hou Z-C (2016) SNP discovery and genotyping using genotyping-by-sequencing in Pekin Ducks. *Sci Rep* 6:36223