

Population structure of *Apodemus flavicollis* and comparison to *Apodemus sylvaticus* in northern Poland based on whole-genome genotyping with RAD-seq

Maria Luisa Martin Cerezo^{1,2}, Marek Kucka³, Karol Zub⁴, Yingguang Frank Chan³ and Jarosław Bryk^{1*}

*Correspondence:

j.bryk@hud.ac.uk

¹School of Applied Sciences,
University of Huddersfield,
Quennsgate, Huddersfield, UK
Full list of author information is
available at the end of the article
†

Abstract

Background: Mice of the genus *Apodemus* are one the most common mammals in the Palaearctic region. Despite their broad range and long history of ecological observations, there are no whole-genome data available for *Apodemus*, hindering our ability to further exploit the genus in evolutionary and ecological genomics context.

Results: Here we present results from the whole-genome, high-density genotyping using double-digest restriction site-associated DNA sequencing (ddRAD-seq) on 72 individuals of *A. flavicollis* and 10 *A. sylvaticus* from four populations, sampled across 500 km distance in northern Poland. Our data present clear genetic divergence of the two species, with average sequence divergence, based on 21377 common loci, of 1.51% and an evolutionary rate of 0.0019 substitutions per site per million years. We provide a catalogue of 117 highly divergent loci that enable genetic differentiation of the two species in Poland and to a large degree of 20 unrelated samples from several European countries and Tunisia. We also show evidence of admixture between the three *A. flavicollis* populations but demonstrate that they have negligible average genome-wide population structure, with largest pairwise $F_{ST} < 0.086$.

Conclusion: Our study demonstrates the feasibility of genome-wide, high-density genotyping in *Apodemus* and provides the first insights into the population genomics of one of the species.

Keywords: RAD-seq; genotyping; population structure; rodents; *Apodemus flavicollis*; *Apodemus sylvaticus*

1
2

Background

Mice of the genus *Apodemus* (Kaup, 1829) (*Rodentia*: *Muridae*) are one the most common mammals in the Palaearctic region [43]. The genus comprises of three subgenera (*Sylvaemus*, *Apodemus* and *Karstomys*) [43], however the systematic classification of the 20 species belonging to the genus [16] is not fully settled [36]. In the Western Palearctic, the yellow-necked mice *A. flavicollis* (Melchior, 1934) and the woodmice *A. sylvaticus* (Linnaeus, 1758) are widespread, sympatric and occasionally syntopic species. They are often difficult to distinguish morphologically in their southern range [30], but in the Central and Northern Europe both are easily recognisable by the yellow collar around the neck of *A. flavicollis*, absent in *A. sylvaticus*.

Their prevalence in Western Palearctic and common status in Western and Central Europe made them one of the model organisms to study post-glacial movement of

15

mammals [24, 45]. Both species have traditionally been studied in a parasitological context, as one of the vectors of *Borellia*-carrying ticks *Ixodes ricinus*, who often feed on *Apodemus* [13, 61], tick-borne encephalitis virus [47] and hantaviruses [33, 50] and have been used as markers for environmental quality [40, 66]. Lastly, they have extra-autosomal chromosomes, called B chromosomes, with varied distribution among the populations [59] but unknown role, although it has been suggested they are involved in cellular metabolism [34, 39].

Previous studies on *Apodemus* typically employed a small number of microsatellite [62] and mtDNA markers [24, 42, 44, 45], which are insufficient to learn about the species' population structure and admixture patterns in detail, or to identify loci under selection. In the absence of high-quality reference genome, which remains cost-prohibitive for complex genomes, whole-genome marker discovery enabled by restriction site-associated DNA sequencing presents a cost-effective method to study species on a population scale even with no previous genetic and genomic resources available [4].

Here we employ the whole-genome high-density genotyping, using double-digest restriction site-associated DNA sequencing (ddRAD-seq) to elucidate the genetic structure and connectivity of three populations of *A. flavicollis* and compare it to a population of *A. sylvaticus* in Poland. We demonstrate clear divergence between the two species and very low differentiation within populations of *A. flavicollis*. Our results provide the first whole-genome-based estimates of population parameters in *A. flavicollis*, genome-wide calculation of divergence between the two *Apodemus* species, as well as a selection of loci enabling their accurate identification.

Results

Sequencing and variant calling

The sequencing produced a total of 92741120 reads. The number of reads per individual varied from 346810 to 4157586, with an average of 1078385 reads per individual and median of 905786,5 (Supplementary Table S2). The best parameters for calling the stacks and variants for the entire dataset were: minimum number of identical, raw reads required to create a stack $m = 2$, number of mismatches allowed between loci for each individual $M = 4$ and number of mismatches allowed between loci when building the catalogue $n = 5$ (Supplementary Figure S1). The best parameters calculated for *A. flavicollis* samples only were: $m = 2$, $M = 4$ and $n = 3$ (Supplementary Figure S3). The coverage per sample ranged from 4.95x to 26.20x with an average of 10.13x and median of 9.32x for the entire dataset (Supplementary Figures S2 and S4).

SNPs and loci co-identification rates

Analysis of the duplicated samples showed that loci and allele misassignment rates were of similar magnitude, on average, between all pairs of duplicates. The duplicate pair F06-B02 showed the highest discrepancy between loci, of 10 %, and also between alleles, of 8 %. When only shared loci were included in the comparisons, all four sets of duplicates showed on average $0.5\% \pm 0.2\%$ SNPs called differently (Table 1).

	F06-B02	A12-F12	H11-G06	G02-D01	MEAN	SD
Reads (D1/D2)	0.19	3.54	1.29	1.380		
Coverage	8.93/11.20	15.95/10.22	8.05/10.51	7.62/8.54		
Locus error rate	0.10	0.08	0.04	0.04	0.07	0.031
Allele error rate	0.08	0.06	0.07	0.05	0.07	0.01
SNP error rate 1	0.15	0.12	0.05	0.07	0.10	0.04
SNP error rate 2	0.006	0.004	0.007	0.004	0.005	0.002

Table 1: Error rates calculated by comparing four sets of duplicated samples. For explanation of different errors please see Methods. D1/D2: ratio of reads from Duplicate 1 to Duplicate 2.

58 Comparison of *A. flavicollis* and *A. sylvaticus*

59 The number of assembled loci per individual ranged from 46286 to 117366 (mean:
60 73711, median: 71395, standard deviation: 29917). 52494 loci passed the population
61 filters established for species differentiation, representing 8,3% of the total 632063
62 loci included in the catalogue. Out of 158144 SNPs called, 60366 (38.1%) were
63 removed after filtering for MAF and 52298 (33%) were removed after failing the
64 HWE test at $p < 0.05$; further 35302 (22.3%) were removed due to a minimum mean
65 depth lower than 20, leaving 10178 SNPs (6.6%) to be used in the downstream
66 analyses (Figure 1). PCA plot of the first two components (Figure 2), accounting
67 for 13.13% of the total variance, shows differentiation of the two species but also
68 distinguish different populations of *A. flavicollis*.

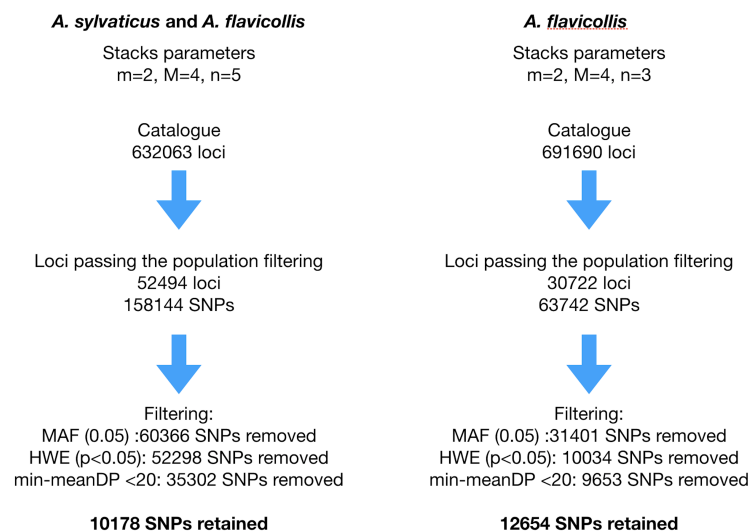


Figure 1: Summary of catalogue construction and SNP filtering steps for the complete dataset (left) and *Apodemus flavicollis* dataset. The graphic includes: Stacks parameters values (m, M, n), number of loci in the catalogue, number of SNPs filtered by minor allele frequency (MAF), which failed the Hardy-Weinberg equilibrium test at $p < 0.05$ (HWE), SNPs removed due to an average depth, across individuals, lower than 20 (min-meanDP) and the total number of SNPs retained for further analysis

69 Similarly, the phylogenetic tree shows *A. sylvaticus* as a separate clade to the
70 three populations of *A. flavicollis*, with *A. flavicollis* from geographically closer

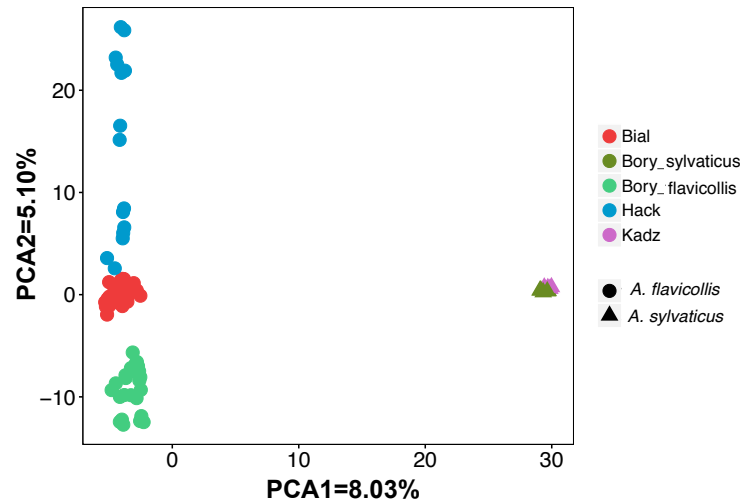


Figure 2: Principal Component Analysis of all samples analysed in the study. Each point represents one sample; the shape of the point represents the species (circles: *Apodemus flavicollis* (n = 72), triangles: *Apodemus sylvaticus* (n = 10), whereas the colour represents the location where the samples were collected: Bial - Białowieża, Kadz - Kadzidło, Hack - Haćki, Bory - Bory Tucholskie.

71 regions (Białowieża and Haćki, 50 km) grouped closer than a population from Bory
72 Tucholskie, 450 km away from Białowieża (Figure 3). The *A. sylvaticus* and *A.*
73 *flavicollis* clusters have high bootstrap value support (100% and 99% respectively).

74 We then investigated the suitability of the loci we identified on Polish populations
75 to distinguish *A. sylvaticus* and *A. flavicollis* from other European populations. The
76 genotyping of the extra 10 samples from each species (see Methods) produced 179763
77 SNPs. 62158 (34.58%) were removed after filtering for MAF and 69125 (38.45%)
78 were removed after failing the HWE test at $p < 0.05$; further 42054 (23.39%) were
79 removed due to a minimum mean depth lower than 20 and 5203 (2.89%) were
80 removed due to more than 5% missing data, leaving 1223 SNPs (0.68%) to be used
81 in the downstream analyses.

82 The first axis of the PCA plot (Figure 4) constructed from this data accounts for
83 the 65.73 % of the total variance and shows clear differentiation between the two
84 species. All the *A. flavicollis* samples cluster with the Polish *A. flavicollis* samples,
85 while all but Tunisian samples of *A. sylvaticus* cluster with the Polish samples of
86 the same species. Tunisian *A. sylvaticus* appear as a separate cluster but still closer
87 to the *A. sylvaticus* group. The catalogue of loci used for species identification is
88 included in the Supplementary Materials, Section 6.

89 Genetic diversity and population structure of *A. flavicollis*

90 The number of assembled loci per individual in the Polish populations ranged from
91 46286 to 117366 (mean: 72738, median: 70592, stdev: 12575). 30722 loci passed the

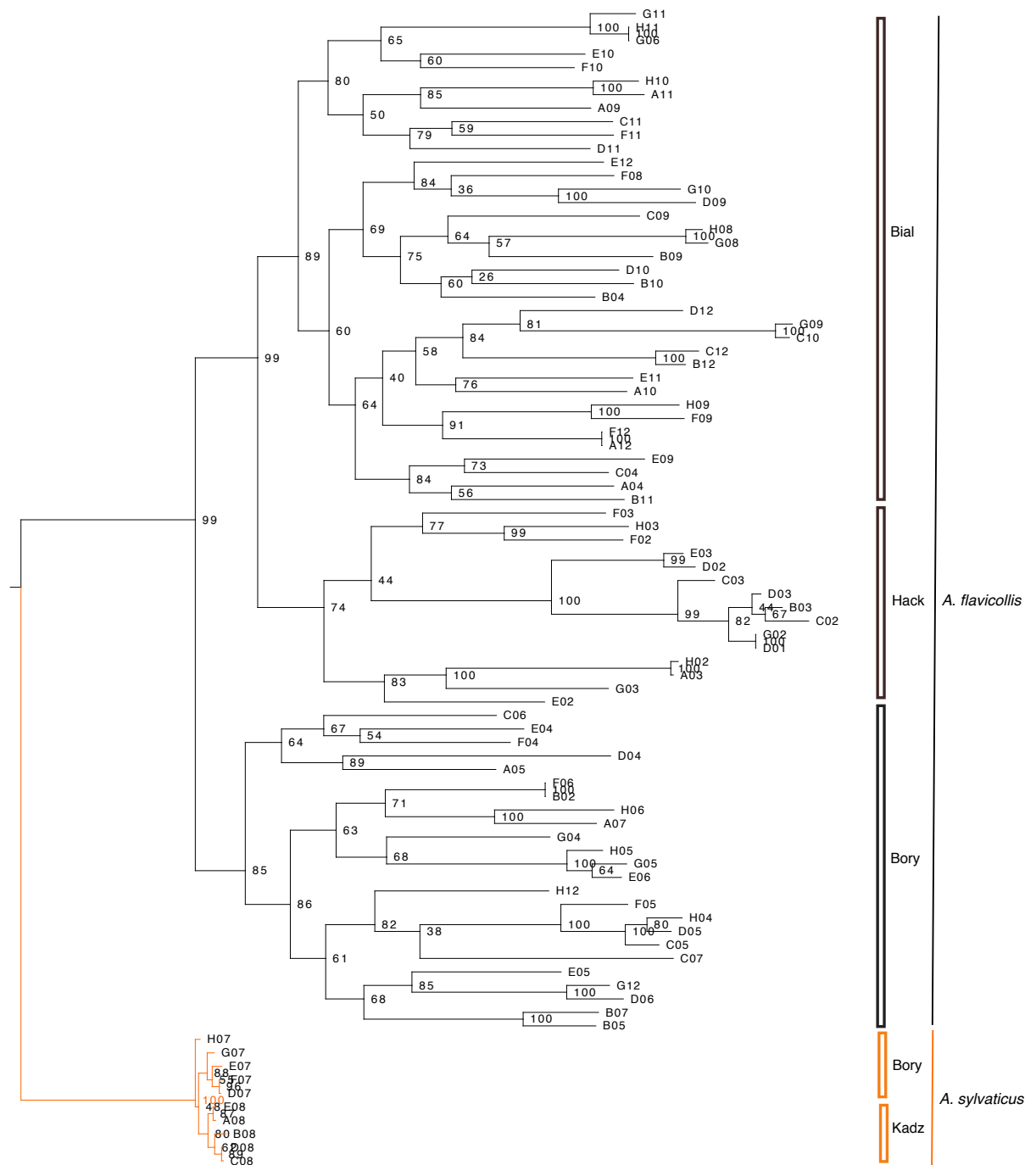


Figure 3: Maximum likelihood phylogenetic tree of all the samples analysed in the study. Colour represents the species: *A. sylvaticus* (n=10) in orange and *A. flavicollis* (n=72) in black. Duplicates samples are included: F06-B02 from Bory Tucholskie, F12-A12 and H11-G06 from Białowieża and G02-D01 from Hacki. Bootstrap support values from 100 replicates are indicated at the nodes of the tree. Bial - Białowieża, Kadz - Kadzidło, Hack - Haćki, Bory - Bory Tucholskie.

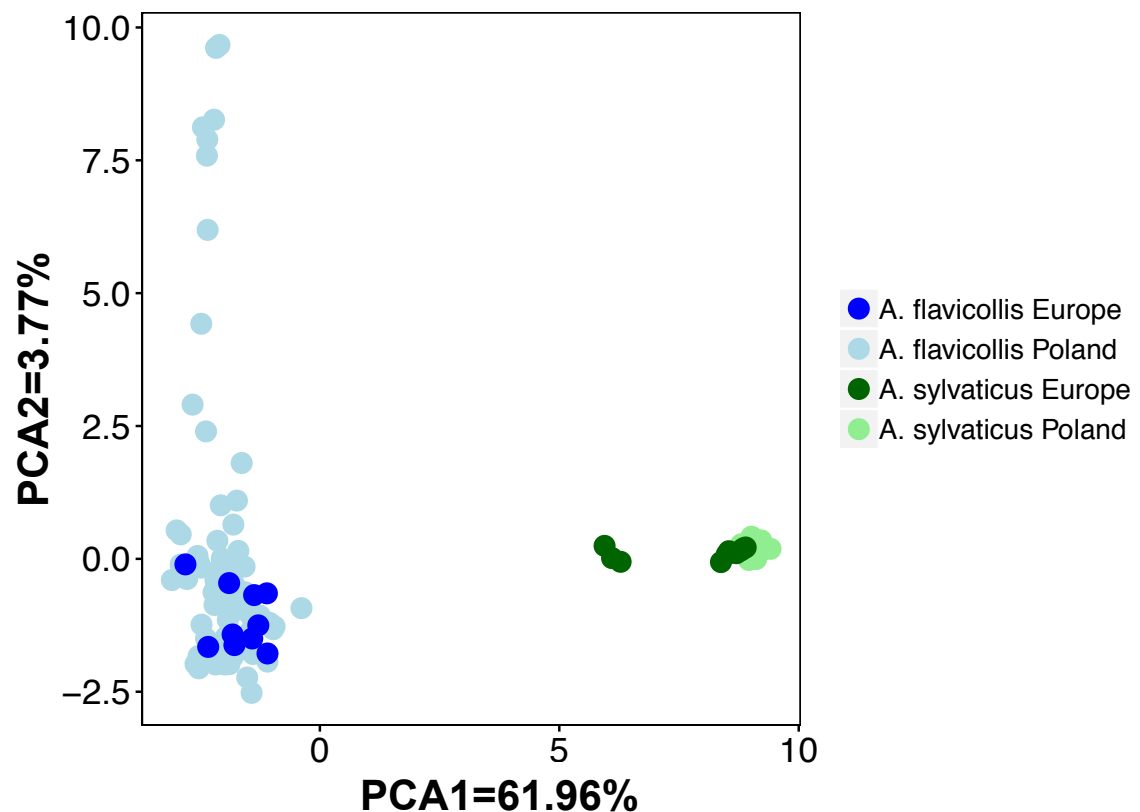


Figure 4: Species identification through Principal Component Analysis using a catalogue of 632060 loci and 1223 final SNPs. Light colours represent samples from Poland while dark colours represent samples from other European regions and Tunisia (collectively named "Europe"). Green: *A. sylvaticus*, blue: *A. flavicollis*.

92 population filters established for population differentiation, representing and 4,43%
 93 of the total 691960 loci included in the catalog. Out of 63742 SNPs called, 31401
 94 (49.26%) were removed after filtering for MAF and 10034 (15.74%) were removed
 95 after failing the HWE test at $p < 0.05$. Further 9653 (15.14%) were removed due to a
 96 minimum mean depth lower than 20, leaving 12654 (19.85%) SNPs to be used in
 97 the downstream analyses (Figure 1).

98 PCA plot (Figure 5) shows differentiation between the three Polish *A. flavicollis*
 99 populations, with PC1 and PC2 cumulatively explaining 10.47% of the total variance.
 100 Haćki population shows larger diversity than the other populations, with some
 101 Haćki individuals closer to Białowieża individuals than to others from this location.
 102 Phylogenetic tree (Figure 6) supports this pattern of differentiation. Bory Tucholskie
 103 and Haćki populations each form a cluster with a 100% of bootstrap support
 104 value, whereas Białowieża forms a third cluster with an 95% of bootstrap support.
 105 Białowieża and Bory Tucholskie population together form a large cluster with a
 106 100% bootstrap support.

107 In the ADMIXTURE analysis, the lowest cross-validation errors [1] were always
 108 found for $K = 3$, indicating contribution of three ancestral populations (Figure 7).

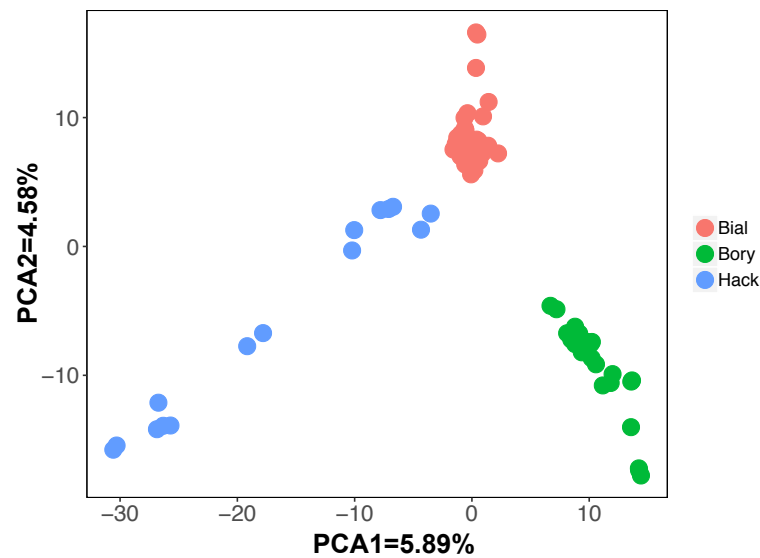


Figure 5: PCA plot showing Polish samples of *A. flavicollis* from Białowieża (red) (n=35), Hački (blue) (n=14) and Bory Tucholskie (green) (n=23). Biał - Białowieża, Kadz - Kadzidło, Hack - Hački.

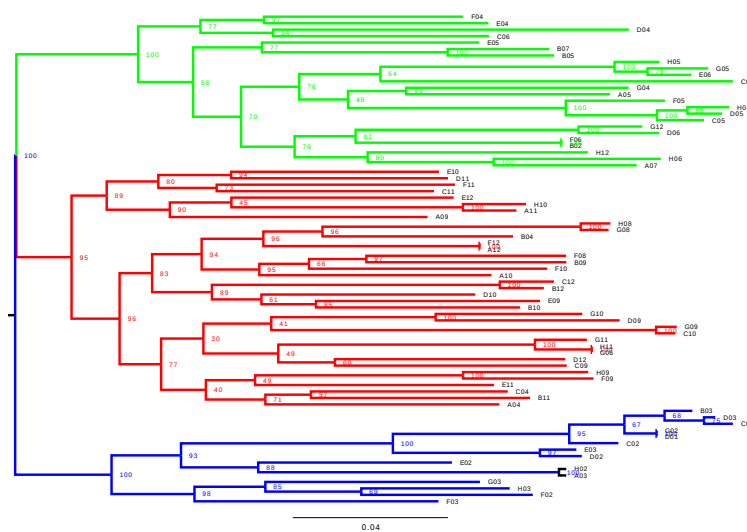


Figure 6: Maximum likelihood phylogenetic tree of n = 72 *A. flavicollis* samples from Białowieża (red, n = 35), Hački (blue, n = 14) and Bory Tucholskie (green, n = 23). Bootstrap support values from 100 replicates are indicated at the nodes of the tree.

Majority of samples from each of the populations show a single dominant component of ancestry with little contribution from other populations, with the exception of four individuals from Hački, which show clear admixture of the Białowieża population.

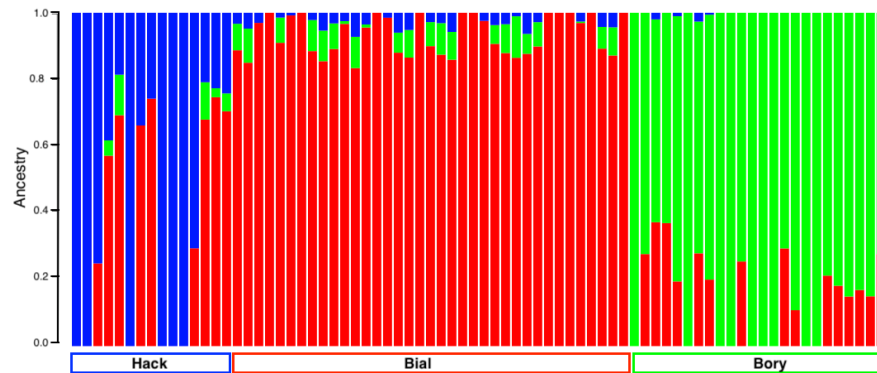


Figure 7: Maximum likelihood Admixture analysis of all *A. flavicollis* samples for the optimal $K = 3$. Each bar represents an individual and each colour represents its ancestry component (red: Białowieża, blue: Haćki, green: Bory Tucholskie).

112 Recognising that STRUCTURE-type analyses (on which ADMIXTURE is based)
 113 may be sensitive to the effects of uneven number of samples in compared groups [57],
 114 we repeated the ADMIXTURE analysis 10 times, each time randomly drawing the
 115 same number of individuals ($n = 15$) from each population. In all cases, the lowest
 116 cross-validation errors were found for $K = 2$, followed by $K = 3$ (Supplementary
 117 Figure S5). At even sampling, ADMIXTURE pattern found for $K = 3$ was the
 118 closest to the observed ecological and geographical distribution of the samples and
 119 closely matched our results when all samples were included (Supplementary Figure
 120 S6).

121 The patterns of heterozygosity highlight Haćki as the only population where the
 122 values of H_o is higher than H_e , where the F_{IS} is negative (Table 2). As parameters
 123 such as number of private alleles, nucleotide diversity and heterozygosity can vary
 124 with sample size, we performed 100 calculations of the above parameters using
 125 random sampling of the same number of individuals ($n = 15$) from each population.
 126 The parameters showed similar relationships except for the number of private alleles
 127 (data not shown).

128 F_{st} values are consistently very low between all the populations, even though
 129 populations from Haćki and Bory Tucholskie show three-fold higher F_{st} values that
 130 for the other two pairs of populations (Table 3).

131 Species divergence

132 Finally, we calculated that the average divergence between *A. flavicollis* and *A.*
 133 *sylvaticus*, based on 21377 shared loci, is 1.51 % (min=0%, max= 6.38%, median=
 134 1.42%, stdev= 1.11%).

Pop ID	N	Npa	Ind per loci	Obs Het	Exp Het	Pi	Fis
Hački	15	32	14.42	0.30	0.27	0.28	-0.04
Bory Tucholskie	24	74	22.93	0.28	0.28	0.29	0.02
Białowieża	37	148	35.13	0.29	0.30	0.30	0.01

Table 2: Population genetic parameters calculated based on 12654 SNPs from all 72 individuals of *A. flavicollis*. N, number of individuals; Npa, number of private alleles; Ind per loci, Mean number of individuals per locus in this population; H_o observed and H_e expected heterozygosity; π , average nucleotide diversity; F_{IS} inbreeding coefficient

	Bory Tucholskie	Białowieża
Hački	0.085	0.055
Bory Tucholskie		0.045

Table 3: Pairwise F_{ST} values for the three populations of *A. flavicollis*.

We then identified the top 117 most divergent loci between the species, which all had the divergence larger than 4.9% (The loci ID are provided in the Supplementary Table S3), and checked whether these loci alone allow for accurate assignment of samples to the two species. We constructed PCA plots from the Polish samples only and from the Polish, other European and Tunisian samples together. They demonstrate that while the 117 loci are sufficient to clearly assign Polish samples to the two species (Supplementary Figure S7), some uncertainty remains when we use these loci for the broader set of samples. Whereas all *A. flavicollis* samples do cluster together, *A. sylvaticus* samples do not form a clearly differentiated group (Supplementary Figure S7).

Discussion

RAD-sequencing approaches, including double-digest RAD-seq and its variants [5, 20, 46, 53, 54], have allowed a cost-effective discovery of thousands of genetic markers in both model and non-model organisms [22, 63], proving to be a transformative research tool in population genetics ([7, 12, 26]), phylogeography and phylogenetics [3, 25, 29, 60], marker development [52], linkage mapping studies [6], species differentiation [49] and detecting selection [65]. However, despite the widespread use of this approach to genome-wide marker discovery, only few studies have used RAD-seq in mammals [18, 32, 35, 48, 64]. Here, we have identified over 10000 markers in two closely related and common species of *Apodemus* in Western Palearctic, characterised the population structure of *A. flavicollis* and compared it to *A. sylvaticus*, for the first time providing genome-wide estimates of the species divergence and population genetic parameters.

Technical considerations

We have used four pairs of technical duplicates to check the accuracy of the RAD-seq genotyping based on the Poland protocol [55]. The largest source of discrepancy in SNP calls between the duplicates is caused by unequal identification of loci: the difference in our case averaged approximately 10% (Table 1) and was similar to allele misidentification rates. However, when considering only shared loci between the duplicates, the discrepancy in SNP calls fell by over an order of magnitude to an average of 0.5%, indicating high accuracy and reliability of calls in once-defined

shared loci. Our finding of loci calls being the major source of genotyping variability agrees with Mastretta *et al.* (2015), although our discrepancies are almost an order of magnitude smaller. Moreover, despite the differences in number of loci included in the analysis, each duplicated pair of samples clustered together with a 100% bootstrap values support and branch length equal to 0 on the phylogenetic tree (Figure 6), indicating that the samples were identical. Overall, our finding reiterates the importance of the influence of stochastic events and imprecise size selection in the library preparations on genotyping calls [41]. We note that some of these variables could be better controlled with automated size-selection approaches [53]. Our findings also illustrate the usefulness of including technical replicates during library preparation.

Effect of group size

Permutations performed for the calculations of genetic diversity parameters (Table 4) have shown that with the exception of the number of private alleles, the results are comparable, regardless of the number of samples included per each population.

Pop ID	N	Npa	Ind per loci	Obs Het	Exp Het	Pi	Fis
Hačky	15	115.53	14.42	0.31	0.28	0.29	-0.05
Bory Tucholskie	15	183.95	14.33	0.29	0.28	0.29	0.02
Białowieża	15	204.84	14.23	0.30	0.29	0.31	0.02

Table 4: Average genetic diversity parameters for *Apodemus flavicollis* calculated from 100 permutations of 45 individuals (15 samples per population, 12654 SNPs). N, number of individuals; Npa, number of private alleles; Ind per loci, Mean number of individuals per locus in this population; H_o observed and H_e expected heterozygosity; π , average nucleotide diversity; F_{IS} inbreeding coefficient

In the ADMIXTURE, we observed different optimal K depending on whether all samples were included in the analysis ($K = 3$) or a set of 15 randomly-chosen set of samples from each population ($K = 2$, although closely followed by $K = 3$). While the previously reported tendency of STRUCTURE-like analyses to produce $\Delta K = 2$ does not apply in our case due to different method to select optimal number of clusters [28], we chose to use $K = 3$ for our analyses due to close match to the spatial and ecological locations from which our populations were sampled. The results obtained for $K = 3$ in the evenly-sampled dataset were similar to the clusters obtained for $K = 3$ with the complete dataset.

Population structure

The F_{ST} values calculated in this study between all three pairs of populations of *A. flavicollis*, based on 12654 SNPs, are consistently low and are not affected when we randomly draw the same number of individuals from each population to compute pairwise F_{ST} (Table 5). Previous studies of *A. flavicollis* populations in north-eastern Poland based on a small number of microsatellites showed similarly and consistently low values [14, 21], even though Gortat *et al.* [21] suggested some population structure based on statistically significant differences between very low pairwise F_{ST} values. Czarnomska *et al.* [14] also suggest large, broadly geographically defined clusters of *A. flavicollis* in north-eastern Poland that are separated by highly

	Bory Tucholskie	Białowieża
Haćki	0.045 ± 0.002	0.057 ± 0.001
Bory Tucholskie		0.045 ± 0.002

Table 5: Average pairwise F_{ST} values $\pm$ standard deviation for the three populations of *A. flavicollis* calculated from 100 permutations of 45 individuals (15 samples per population, 12654 SNPs)

admixed individuals, but, again, F_{ST} between those clusters are as low as those reported by Gortat et al. [21] and this study.

We would argue, based on a much larger set of genome-wide markers reported here, that *A. flavicollis* has a negligible population structure across the entire area studied. Large number of markers nevertheless allows us to discover evidence for admixture of Białowieża population and Haćki (Figure 7), further indicated by relatively high heterozygosity and negative F_{IS} in this population. It is therefore intriguing that such a low differentiation occurs across hundreds of kilometres of varying landscape in a species that typically has a limited range of about 4 km² and that suffers up to 86% winter mortality rate [56], which would lead to multiple bottlenecks and drift-driven population differentiation. With this in mind, our data suggests a much larger dispersal ability of the species, a much better connectivity between populations, or both.

The heterozygosity values reported in this study are smaller than in previous work by Czarnomska et al. [14], Gortat et al. [21]. They range from 0.28 to 0.31, in comparison to ranges between 0.84 to 0.88 in [14] and 0.56 and 0.7 in [21] for most but not all of their markers. However, as their work was based on few microsatellites, these differences likely reflect the higher variability of microsatellites compared to SNPs [17, 19, 23].

Both low overall F_{ST} and moderate heterozygosity suggest it would be worthwhile to conduct a genome-wide scan for selection using F_{ST} as a metrics of local genomic differentiation to identify geographically local regions under selection. This, however, is not yet possible given the lack of high-quality reference genome for *Apodemus* and unknown synteny to the available genome of *Mus musculus*.

Divergence and differentiation of *A. flavicollis* and *A. sylvaticus*

Given that accurate identification of the two species using morphological characters is problematic, especially in their southern range [9], a large collection of markers identified in this study allowed us to create a catalogue of 632060 loci that allow clear differentiation between species. This identification is somewhat biased, as the catalogue was built using many more samples of *A. flavicollis* than *A. sylvaticus* (72 vs 10) and both from a relatively limited geographical range. Nevertheless, it allowed for accurate assignment of *A. flavicollis* samples and to a large degree of *A. sylvaticus*, as we demonstrated on a set of 20 independent samples from other European countries and Tunisia (Figure 4). Given the wide distribution of both species in Western Palearctic, a more representative sample from both species from a broader geographic range would likely provide more accurate set of markers for their identification.

Finally, we calculated the nucleotide divergence between the two species, based on 21377 shared loci, which is 1.51%. Considering a divergence time between *A. flavicollis*

and *A. sylvaticus* estimated from archeological data of 4 Mya [42], the evolution rate is 0.0019 substitutions per site per million of years. This estimate of sequence divergence level is in broad agreement with calculations based on mitochondrial 12S rRNA, IRBP and Cytochrome b genes [43]. However, our calculation is likely an underestimate, as we only used shared loci to calculate divergence and did not include the potential impact of insertion/deletion events, which can significantly affect the total genomic divergence between species Britten [8], Li et al. [38]. Highly divergent sequences would have been identified as different loci, and would not be compared to their true homologous sequences.

Conclusions

We have successfully applied the ddRAD-seq approach to discover tens of thousands of SNPs in wide-spread and common mammalian species of *A. flavicollis* and *A. sylvaticus*. The high resolution data obtained here allowed us to delineate geographically close populations, including identifying admixture between them, but suggest that *A. flavicollis* effectively forms a single population in an entire sampling area that spans 500 km in the W-E direction. Comparing *A. flavicollis* and *A. sylvaticus*, we have calculated their genome-wide divergence and identified a set of genomic loci that enable effective molecular identification of the species. We anticipate that with the development of further whole-genome resources, *Apodemus*, thanks to its common status, broad geographic range and long history of ecological observations, will become an excellent model species for evolutionary and ecological research in the genomic era.

Methods

Sample collection and DNA extraction

Eighty two individuals (10 *Apodemus sylvaticus* and 72 *Apodemus flavicollis*) from four locations in northern Poland spanning 500 km were trapped in 2015 (Figure 8). *A. flavicollis* were collected in Białowieża (E23.8345814, N52.7231935), an oak-lime-hornbeam forest (n = 35), Bory Tucholskie (E17.5160265, N53.7797608), in an oak-lime-hornbeam and pine forest (n = 23) and Haćki (E23.1793284, N52.834369), in a xerothermic meadow (n = 14). *A. sylvaticus* were trapped in Kadzidło (E21.3778496, N53.2089113) in a dry pine forest (n = 5) and in Bory Tucholskie, mainly in a pine forest (n = 5) (Supplementary Table S1). While *A. flavicollis* are present in all sampled locations, there have been no trappings of *A. sylvaticus* in Białowieża for the last 20 years, despite Białowieża being within the European range of this species (Dr Karol Zub, personal communication). The sampling procedures were approved by the Local Ethical Commission on Experimentation on Animals in Białystok, Poland, under permission number 2015/99.

Tail clippings were collected, preserved in $\geq 95\%$ ethanol and stored at -20°C until DNA extraction. The tissues were digested by incubating at 55°C overnight with lysis buffer (10mM Tris, 100mM NaCl, 10mM EDTA, 0.5% SDS) and proteinase K (20mg/ml). Subsequently, potassium acetate and RNase A were used to remove protein and RNA contamination. Three ethanol washes were performed using Sera-Mag SpeedBeads solution (GElifesciences, Marlborough, MA, USA). The quality and integrity of the DNA was tested in a 2% agarose gel. Twenty-fold dilutions of

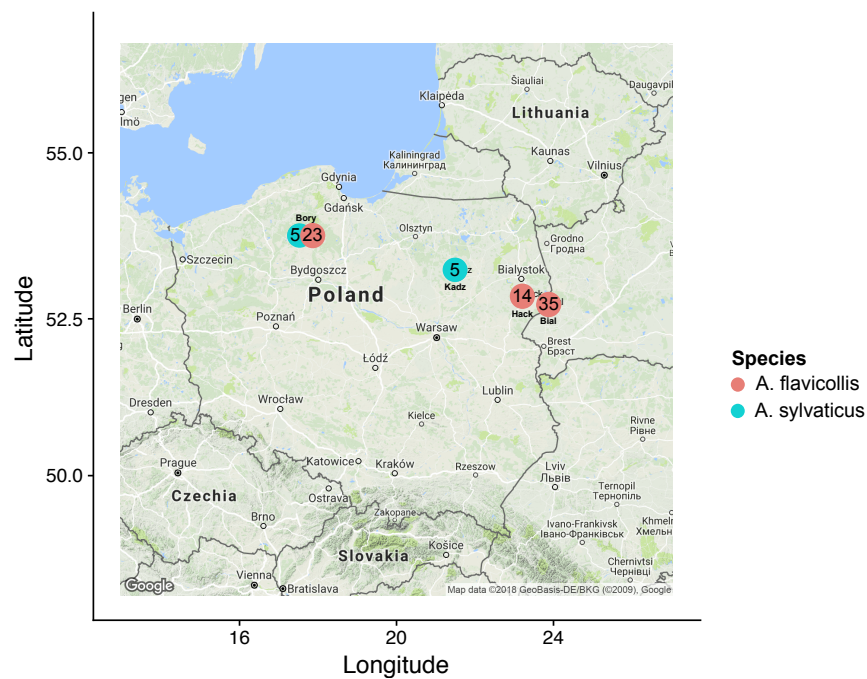


Figure 8: Locations of the Polish samples used in this study. Red circles represent samples from *Apodemus flavicollis* while blue dots represent samples from *Apodemus sylvaticus*. The number inside the circles are the number of samples from each locality. Bial - Białowieża, Kadz - Kadzidło, Hack - Haćki, Bory - Bory Tucholskie.

the samples were used to measure the DNA concentration using Quant-iT PicoGreen dsDNA assay kit (Invitrogen, Carlsbad, CA, USA) and concentration of each sample was then normalised to 10 ng/ μ l in 20 μ l volume. Four samples were used as technical duplicates (F06-B02, G02-D01, H11-G06, F12-A12). Technical duplicates had the same DNA but were digested and ligated to barcodes independently.

ddRAD-seq library preparation

ddRAD-seq library was prepared following the protocol from Poland and Rife [54], adapted to a different combination of enzymes. Briefly, genomic DNA was digested in a 20 μ l reaction with CutSmart[®] buffer, 8 units of *SbfI* and 8 units of HF-*MseI* (New England Biolabs, Frankfurt am Main, Germany). Digestion was performed at 37°C for 2 hours. Enzymes were inactivated at 65°C for 20 minutes and the reactions were kept at 8°C. Adapter ligation was performed at 22°C for 2 hours and the ligase was inactivated by incubating the samples at 65°C for 20 minutes. Samples were cooled down to 8°C and multiplexed by combining 5 μ l of each sample. P1 adapters contained barcodes with a length between 5 and 10 bp.

PCR amplification was conducted in 25 μ l with 1 μ l of each primer (IlluminaF_PE: AATGATACGCGACCAACGAGATCTACACTCTTCCCTACAC-GACGCTCTTCCGATCT and IlluminaR_PE: CAAGCAGAAGACGGCATAAC-GAGATCGGTCTCGGCATTCCTGCTGAA) at 10mM, 0.5 μ l of 10 mM dNTPs, 13.25 μ l of PCR-grade water, 5 μ l of 5x Phusion HF Buffer, 0.25 μ l of Phusion DNA

303 Polymerase (New England Biolabs, Frankfurt am Main, Germany) and 4 μ l of the
304 multiplexed DNA. After an initial denaturation step of 30s at 98°C, PCR reaction
305 was carried out for 12 cycles (10s at 98°C, 20s at 58°C and 15s at 72°C). Final
306 elongation step was performed at 72°C for 5 minutes.

307 PCR products were loaded into a single lane on a 1% agarose gel with 100 bp DNA
308 ladder (New England Biolabs, Frankfurt am Main, Germany). Fragments between
309 200 and 500 bp were cut from the gel with a scalpel and purified using the QIAquick
310 gel extraction kit (QIAGEN, Hilden, Germany), followed by the second cleanup
311 step with Sera-Mag SpeedBeads (GELifesciences, Marlborough, MA, USA). Sizing,
312 quantification and quality control of the DNA was performed using Bioanalyzer
313 2100 (Agilent, Santa Clara, CA, USA) before paired-end sequencing on an Illumina
314 HiSeq 3500 with 150 bp read length.

315 Processing of RAD-tags

316 Sequences were analysed with Stacks version 1.48 [10]. Samples were demultiplexed
317 using process_radtags allowing no mismatches in barcodes and cutting sites. Se-
318 quences with uncalled bases and low quality scores were removed and all reads were
319 trimmed to 141 bp. The four files generated per sample by process_radtags were
320 concatenated using a custom bash script. The best parameters for building and
321 calling SNPs *de novo*, using denovo_map, were calculated following Paris et al.
322 [51] approach, using either samples from both species or only from *A. flavicollis*.
323 Secondary reads were not used to call haplotypes in denovo_map (option -H).

324 SNPs and loci co-identification rates

325 We estimated the loci and SNP co-identification rates by analysing a set of four
326 samples that were prepared and sequenced in duplicates. Sequences for 52494 loci
327 from both species, were extracted using -fasta_samples option from the population
328 package in Stacks. We extracted sequences for each of the duplicated samples with a
329 custom script and calculated co-identification rates as described by [41]. Briefly, the
330 locus misassignment rate is the percentage of unidentified loci, calculated by dividing
331 the number of loci found only in one of the duplicates by the total number of loci
332 in each sample. The allele misassignment rate is the percentage of mismatches
333 between the UIPAC consensus sequences between homologous loci from each pair of
334 duplicates. Finally, the two SNP error rates: the percentage of different SNPs called
335 in each of the duplicated samples using either all 10178 SNPs or using the SNPs
336 called without missing data between duplicate samples excluded (see Table 1).

337 Variant calling and filtering

338 We combined the data from *A. sylvaticus* and *A. flavicollis* to establish species
339 differentiation and then filtered the SNPs using the population package from Stacks
340 [10] and VCFtools [15]. We kept SNPs common to the 80% of the individuals in each
341 species ($p=1$, $r=0.8$) and excluded SNPs with minor allele frequencies $MAF < 0.05$
342 and which deviated from the Hardy-Weinberg equilibrium (HWE) at $P < 0.05$. We
343 also removed sites with mean depth values lower than 20. We manually modified
344 the chromosome numbers in the vcf file to input it into SNPhylo [37], which we
345 used to build the tree. We set a missing rate (-M) of 1, minor allele frequencies

(-m) of 0, linkage disequilibrium threshold (-l) of 1 and the -r option to skip the step of removing low quality data. Confidence values were estimated using 1000 bootstrap replicates. The root was manually fixed to separate both species. Principal Component Analysis (PCA) was performed using the R package Adegenet [31] (Figure 3).

The set of divergent loci identified between the two species in Polish samples was tested for its ability to differentiate an extra set of samples from other locations in Europe and Tunisia. Ten *A. flavicollis* (2 samples from Austria, 5 from Lithuania and 3 from Romania) and 10 samples of *A. sylvaticus* (4 samples from Wales, 3 from Tunisia and 3 from Scotland) were kindly provided by Dr Jeremy Herman, National Museums Scotland, Dr Johan Michaux, University of Liege and Dr Karol Zub, Mammal Research Institute of the Polish Academy of Sciences (MRI) (Supplementary Table S4). We considered all 20 test samples as a different group from Polish *A. sylvaticus* and *A. flavicollis* for SNP calling. We kept SNPs common to the 80% of the individuals in each group ($p=1$, $r=0.8$) and excluded SNPs with minor allele frequencies $MAF < 0.05$, SNPs which deviated from the Hardy-Weinberg equilibrium (HWE) at $P < 0.05$, sites with mean depth values lower than 20 and with more than 5% of missing data.

Population divergence

To analyse genetic diversity and population connectivity within *A. flavicollis*, we analysed the three populations (Bory Tucholskie, Białowieża and Haćki) separately ($p=3$, $r=0.8$), while keeping the other parameters as described above. Due to the lack of outgroup, a mid-point root was chosen in the phylogenetic tree. Individual ancestries were estimated following a maximum likelihood approach with ADMIXTURE [2], after conversion of the VCF file to ped with plink version 1.9 [11, 58]. ADMIXTURE analysis was run for each of $K=1$ to $K=5$, each using 10 different seeds. Weighted (Weir-Cockerham) F_{st} was calculated with VCFtools v0.1.13. Heterozygosity, P_i and F_{is} were calculated with the population package from Stacks [10].

Species divergence

To calculate the divergence between the two species, a set of common loci was extracted with a custom script and the strict consensus sequences for each species were calculated with Consensus.pl script [27]. Sequence divergence was then calculated using a custom R script (Supplementary Materials, Section 9).

Competing interests

The authors declare that they have no competing interests.

Funding

This project was supported by the University of Huddersfield, the Friedrich Meiser laboratory of the Max Planck Society and Microsoft Azure Research Award CRM:0518338.

Author's contributions

MLMC designed the study, prepared the sequencing library, analysed the data and wrote the manuscript. MK contributed to the experimental design and preparation of the sequencing library. KZ performed the sampling and contributed to the manuscript. YFC contributed to the analysis of the data and the manuscript. JB designed the study, analysed the data and wrote the manuscript.

390 Acknowledgements

391 The authors wish to thank Jeremy Herman (National Museums Scotland) and Johan Michaux (Université de
392 Liège) for providing the European samples for testing our catalog of SNPs.

393 Author details

394 ¹School of Applied Sciences, University of Huddersfield, Quennsgate, Huddersfield, UK. ²AVIAN Behavioural
395 Genomics and Physiology Group, IFM Biology, Department of Zoology, Linköping University, Linköping,
396 Sweden. ³Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany. ⁴The Mammal
397 Research Institute, Polish Academy of Sciences, Białowieża, Poland.

398 References

- 399 1. David H Alexander and Kenneth Lange. Enhancements to the admixture algorithm for individual ancestry
400 estimation. *BMC bioinformatics*, 12(1):246, 2011.
- 401 2. David H Alexander, John Novembre, and Kenneth Lange. Fast model-based estimation of ancestry in
402 unrelated individuals. *Genome research*, 19(9):1655–1664, 2009.
- 403 3. S Elizabeth Alter, Jason Munshi-South, and Melanie LJ Stiassny. Genome wide snp data reveal cryptic
404 phylogeographic structure and microallopatric divergence in a rapids-adapted clade of cichlids from the
405 congo river. *Molecular ecology*, 26(5):1401–1419, 2017.
- 406 4. Kimberly R Andrews, Jeffrey M Good, Michael R Miller, Gordon Luikart, and Paul A Hohenlohe.
407 Harnessing the power of radseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2):
408 81, 2016.
- 409 5. N.A. Baird, P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver, Z.A. Lewis, E.U. Selker, W.A. Cresko, and
410 E.A. Johnson. Rapid snp discovery and genetic mapping using sequenced rad markers. *PloS one*, 3(10):
411 e3376, 2008.
- 412 6. Simon W Baxter, John W Davey, J Spencer Johnston, Anthony M Shelton, David G Heckel, Chris D
413 Jiggins, and Mark L Blaxter. Linkage mapping and comparative genomics using next-generation rad
414 sequencing of a non-model organism. *PloS one*, 6(4):e19315, 2011.
- 415 7. L Blanco-Bercial and A Bucklin. New view of population genetics of zooplankton: Rad-seq analysis reveals
416 population structure of the north atlantic planktonic copepod *centropages typicus*. *Molecular ecology*, 25
417 (7):1566–1580, 2016.
- 418 8. Roy J Britten. Divergence between samples of chimpanzee and human dna sequences is 5%, counting
419 indels. *Proceedings of the National Academy of Sciences*, 99(21):13633–13635, 2002.
- 420 9. Vanja Bugarski-Stanojević, Jelena Blagojević, Tanja Adnadjević, Vladimir Jovanović, and Mladen
421 Vujošević. Identification of the sibling species *apodemus sylvaticus* and *apodemus flavicollis* (rodentia,
422 muridae)—comparison of molecular methods. *Zoologischer Anzeiger-A Journal of Comparative Zoology*,
423 252(4):579–587, 2013.
- 424 10. Julian M Catchen, Angel Amores, Paul Hohenlohe, William Cresko, and John H Postlethwait. Stacks:
425 building and genotyping loci de novo from short-read sequences. *G3: Genes, genomes, genetics*, 1(3):
426 171–182, 2011.
- 427 11. Christopher C Chang, Carson C Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M Purcell, and
428 James J Lee. Second-generation plink: rising to the challenge of larger and richer datasets. *Gigascience*, 4
429 (1):7, 2015.
- 430 12. Gareth A Cromie, Katie E Hyma, Catherine L Ludlow, Cecilia Garmendia-Torres, Teresa L Gilbert, Patrick
431 May, Angela A Huang, Aimée M Dudley, and Justin C Fay. Genomic sequence diversity and population
432 structure of *saccharomyces cerevisiae* assessed by rad-seq. *G3: Genes, Genomes, Genetics*, 3(12):
433 2163–2171, 2013.
- 434 13. Benjamin Cull, Alexander GC Vaux, Lisa J Ottowell, Emma L Gillingham, and Jolyon M Medlock. Tick
435 infestation of small mammals in an english woodland. *Journal of Vector Ecology*, 42(1):74–83, 2017.
- 436 14. Sylwia D Czarnomska, Magdalena Niedziałkowska, Tomasz Borowik, and Bogumiła Jędrzejewska. Regional
437 and local patterns of genetic variation and structure in yellow-necked mice—the roles of geographic distance,
438 population abundance, and winter severity. *Ecology and evolution*, 8:8171–8186, 2018.
- 439 15. Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A Albers, Eric Banks, Mark A DePristo, Robert E
440 Handsaker, Gerton Lunter, Gabor T Marth, Stephen T Sherry, et al. The variant call format and vcftools.
441 *Bioinformatics*, 27(15):2156–2158, 2011.
- 442 16. Jamshid Darvish, Zeinolabedin Mohammadi, Fatemeh Ghorbani, Ahmad Mahmoudi, and Sylvain Dubey.
443 Phylogenetic relationships of *apodemus kaup*, 1829 (rodentia: Muridae) species in the eastern
444 mediterranean inferred from mitochondrial dna, with emphasis on iranian species. *Journal of mammalian
445 evolution*, 22(4):583–595, 2015.
- 446 17. María E Fernández, Daniel E Goszczynski, Juan P Lirón, Egle E Villegas-Castagnasso, Mónica H Carino,
447 María V Ripoli, Andrés Rogberg-Muñoz, Diego M Posik, Pilar Peral-García, and Guillermo Giovambattista.
448 Comparison of the effectiveness of microsatellites and snp panels for genetic identification, traceability and
449 assessment of parentage in an inbred angus herd. *Genetics and molecular biology*, 36(2):185–191, 2013.
- 450 18. R Fernández, M Schubert, AM Vargas-Velázquez, A Brownlow, GA Vikingsson, U Siebert, LF Jensen,
451 N Øien, D Wall, E Rogan, et al. A genomewide catalogue of single nucleotide polymorphisms in
452 white-beaked and atlantic white-sided dolphins. *Molecular ecology resources*, 16(1):266–276, 2016.
- 453 19. Martin C Fischer, Christian Rellstab, Marianne Leuzinger, Marie Roumet, Felix Gugerli, Kentaro K Shimizu,
454 Rolf Holderegger, and Alex Widmer. Estimating genomic diversity and population differentiation—an
455 empirical comparison of microsatellite and snp variation in *arabidopsis halleri*. *BMC genomics*, 18(1):69,
456 2017.
- 457 20. Paolo Franchini, Daniel Monné Parera, Andreas F Kautt, and Axel Meyer. quaddrad: a new
458 high-multiplexing and pcr duplicate removal ddrad protocol produces novel evolutionary insights in a
459 nonradiating cichlid lineage. *Molecular Ecology*, 26(10):2783–2795, 2017.

21. Tomasz Gortat, Alicja Gryczyńska-Siemiatkowska, Robert Rutkowski, Anna Kozakiewicz, Antoni Mikoszewski, and Michał Kozakiewicz. Landscape pattern and genetic structure of a yellow-necked mouse *apodemus flavicollis* population in north-eastern poland. *Acta Theriologica*, 55(2):109–121, 2010.
22. Nicholas M Hammerman, Ramon E Rivera-Vicens, Matthew P Galaska, Ernesto Weil, Richard S Appeldoorn, Monica Alfaro, and Nikolaos V Schizas. Population connectivity of the plating coral *agaricia lamarcki* from southwest puerto rico. *Coral Reefs*, 37(1):183–191, 2018.
23. Lorenz Hauser, Melissa Baird, RAY Hilborn, Lisa W Seeb, and James E Seeb. An empirical comparison of snps and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*oncorhynchus nerka*) population. *Molecular ecology resources*, 11:150–161, 2011.
24. Jeremy S Herman, Frída Jóhannesdóttir, Eleanor P Jones, Allan D McDevitt, Johan R Michaux, Thomas A White, Jan M Wójcik, and Jeremy B Searle. Post-glacial colonization of europe by the wood mouse, *apodemus sylvaticus*: evidence of a northern refugium and dispersal with humans. *Biological Journal of the Linnean Society*, 120(2):313–332, 2017.
25. Andrew L Hipp, Deren AR Eaton, Jeannine Cavender-Bares, Elisabeth Fitzek, Rick Nipper, and Paul S Manos. A framework phylogeny of the american oak clade based on sequenced rad data. *PLoS One*, 9(4): e93975, 2014.
26. Paul A Hohenlohe, Mitch D Day, Stephen J Amish, Michael R Miller, Nick Kamps-Hughes, Matthew C Boyer, Clint C Muhlfeld, Fred W Allendorf, Eric A Johnson, and Gordon Luikart. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end rad sequencing. *Molecular ecology*, 22(11):3002–3013, 2013.
27. Joseph Hughes. Sequence-manipulation. *GitHub repository*, 2011.
28. Jasmine K Janes, Joshua M Miller, Julian R Dupuis, René M Malenfant, Jamieson C Gorrell, Catherine I Cullingham, and Rose L Andrew. The $k=2$ conundrum. *Molecular Ecology*, 26(14):3594–3602, 2017.
29. Daniel L Jeffries, Gordon H Copp, Lori Lawson Handley, K Håkan Ölsén, Carl D Sayer, and Bernd Hänfling. Comparing radseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the crucian carp, *carassius carassius*, l. *Molecular ecology*, 25(13):2997–3018, 2016.
30. Vida Jojić, Vanja Bugarski-Stanojević, Jelena Blagojević, and Mladen Vujošević. Discrimination of the sibling species *apodemus flavicollis* and *a. sylvaticus* (rodentia, muridae). *Zoologischer Anzeiger-A Journal of Comparative Zoology*, 253(4):261–269, 2014.
31. Thibaut Jombart. adegenet: a r package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11):1403–1405, 2008.
32. L Lacey Knowles, Rob Massatti, Qixin He, Link E Olson, and Hayley C Lanier. Quantifying the similarity between genes and geography across alaska's alpine small mammals. *Journal of Biogeography*, 43(7): 1464–1476, 2016.
33. Marcin Kolodziej, Alicja Melgies, Justyna Joniec-Wiechetek, Aleksander Michalski, Anna Nowakowska, Grzegorz Pitucha, and Marcin Niemcewicz. First molecular characterization of *dobrava-belgrade* virus found in *apodemus flavicollis* in poland. *Annals of Agricultural and Environmental Medicine*, 25(2): 368–373, 2018.
34. J. Kozłowski, M. Konarzewski, and A.T. Gawelczyk. Cell size as a link between noncoding dna and metabolic rate scaling. *Proceedings of the National Academy of Sciences*, 100(24):14080–14085, 2003.
35. Hayley C Lanier, Rob Massatti, Qixin He, Link E Olson, and L Lacey Knowles. Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in collared pikas (*ochotona collaris*). *Molecular Ecology*, 24(14):3688–3705, 2015.
36. A.G. Lapinski, M.V. Pavlenko, L.L. Solovenchuk, and V.V. Gorbachev. Some limitations in the use of the mitochondrial dna *cytb* gene as a molecular marker for phylogenetic and population-genetic studies by the example of the *apodemus* genus. *Russian Journal of Genetics: Applied Research*, 6(1):84–90, 2016.
37. Tae-Ho Lee, Hui Guo, Xiyin Wang, Changsoo Kim, and Andrew H Paterson. Snphilo: a pipeline to construct a phylogenetic tree from huge snp data. *BMC genomics*, 15(1):162, 2014.
38. Wen-Hsiung Li, Masako Tanimura, and Paul M Sharp. An evaluation of the molecular clock hypothesis using mammalian dna sequences. *Journal of molecular evolution*, 25(4):330–342, 1987.
39. S. Maciak, E. Bonda-Ostaszewska, M. Czarnofęski, M. Konarzewski, and J. Kozłowski. Mice divergently selected for high and low basal metabolic rates evolved different cell size and organ mass. *Journal of evolutionary biology*, 27(3):478–487, 2014.
40. M. Martiniaková, R. Omelka, J. Jancová, R. Stawarz, and G. Formicki. Heavy metal content in the femora of yellow-necked mouse (*apodemus flavicollis*) and wood mouse (*apodemus sylvaticus*) from different types of polluted environment in slovakia. *Environmental monitoring and assessment*, 171(1-4):651–660, 2010.
41. A Mastretta-Yanes, Nils Arrigo, Nadir Alvarez, Tove H Jorgensen, D Piñero, and BC Emerson. Restriction site-associated dna sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15(1):28–41, 2015.
42. Johan René Michaux, Elodie Magnanou, Emmanuel Paradis, Caroline Nieberding, and Roland Libois. Mitochondrial phylogeography of the woodmouse (*apodemus sylvaticus*) in the western palearctic region. *Molecular Ecology*, 12(3):685–697, 2003.
43. JR Michaux, P Chevret, M-G Filippucci, and M Macholan. Phylogeny of the genus *apodemus* with a special emphasis on the subgenus *sylvaemus* using the nuclear *irbp* gene and two mitochondrial markers: cytochrome b and 12s rrna. *Molecular phylogenetics and evolution*, 23(2):123–136, 2002.
44. JR Michaux, Roland Libois, E Paradis, and M-G Filippucci. Phylogeographic history of the yellow-necked fieldmouse (*apodemus flavicollis*) in europe and in the near and middle east. *Molecular phylogenetics and evolution*, 32(3):788–798, 2004.
45. JR Michaux, Roland Libois, and MG Filippucci. So close and so different: comparative phylogeography of two small mammal species, the yellow-necked fieldmouse (*apodemus flavicollis*) and the woodmouse (*apodemus sylvaticus*) in the western palearctic region. *Heredity*, 94(1):52–63, 2005.

- 531 46. M.R. Miller, J.P. Dunham, A. Amores, W.A. Cresko, and E.A. Johnson. Rapid and cost-effective
532 polymorphism identification and genotyping using restriction site associated dna (rad) markers. *Genome*
533 *research*, 17(2):240–248, 2007.
- 534 47. Luwanika Mlera and Marshall E Bloom. The role of mammalian reservoir hosts in tick-borne flavivirus
535 biology. *Frontiers in cellular and infection microbiology*, 8:298, 2018.
- 536 48. Andre E Moura, John G Kenny, Roy Chaudhuri, Margaret A Hughes, Andreanna J. Welch, Ryan R
537 Reisinger, PJ Nico de Bruyn, Marilyn E Dahlheim, Neil Hall, and A Rus Hoelzel. Population genomics of
538 the killer whale indicates ecotype evolution in sympatry involving both selection and drift. *Molecular*
539 *Ecology*, 23(21):5179–5192, 2014.
- 540 49. Eric Pante, Jawad Abdelkrim, Amélia Viricel, Delphine Gey, SC France, Marie-Catherine Boisselier, and
541 Sarah Samadi. Use of rad sequencing for delimiting species. *Heredity*, 114(5):450, 2015.
- 542 50. Anna Papa, Elton Rogozi, Enkelejda Velo, Evangelia Papadimitriou, and Silvia Bino. Genetic detection of
543 hantaviruses in rodents, albania. *Journal of medical virology*, 88(8):1309–1313, 2016.
- 544 51. Josephine R Paris, Jamie R Stevens, and Julian M Catchen. Lost in parameter space: a road map for
545 stacks. *Methods in Ecology and Evolution*, 2017.
- 546 52. Venkatramana Pegadaraju, Rick Nipper, Brent Hulke, Lili Qi, and Quentin Schultz. De novo sequencing of
547 sunflower genome for snp discovery using rad (restriction site associated dna) approach. *BMC genomics*,
548 14(1):556, 2013.
- 549 53. Brant K Peterson, Jesse N Weber, Emily H Kay, Heidi S Fisher, and Hopi E Hoekstra. Double digest
550 radseq: an inexpensive method for de novo snp discovery and genotyping in model and non-model species.
551 *PLoS one*, 7(5):e37135, 2012.
- 552 54. Jesse A Poland and Trevor W Rife. Genotyping-by-sequencing for plant breeding and genetics. *The Plant*
553 *Genome*, 5(3):92–102, 2012.
- 554 55. Jesse A Poland, Patrick J Brown, Mark E Sorrells, and Jean-Luc Jannink. Development of high-density
555 genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS*
556 *one*, 7(2):e32253, 2012.
- 557 56. Zdzisław Pucek, Włodzimierz Jędrzejewski, Bogumiła Jędrzejewska, and Michalina Pucek. Rodent
558 population dynamics in a primeval deciduous forest (białowieża national park) in relation to weather, seed
559 crop, and predation. *Acta Theriologica*, 38(2):199–232, 1993.
- 560 57. Sebastien J Puechmaile. The program structure does not reliably recover the correct population structure
561 when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology*
562 *Resources*, 16(3):608–627, 2016.
- 563 58. S Purcell and CC Chang. Plink 1.9 package. 2018. URL <https://www.cog-genomics.org/plink/1.9/>.
- 564 59. Marija Rajičić, Svetlana A Romanenko, Tatyana V Karamysheva, Jelena Blagojević, Tanja Adnađević,
565 Ivana Budinski, Aleksey S Bogdanov, Vladimir A Trifonov, Nikolay B Rubtsov, and Mladen Vujošević. The
566 origin of b chromosomes in yellow-necked mice (apodemus flavicollis)—break rules but keep playing the
567 game. *PLoS one*, 12(3):e0172704, 2017.
- 568 60. AM Reitzel, S Herrera, MJ Layden, MQ Martindale, and TM Shank. Going where traditional markers have
569 not gone before: utility of and promise for rad sequencing in marine invertebrate phylogeography and
570 population genomics. *Molecular ecology*, 22(11):2953–2970, 2013.
- 571 61. Dania Richter, Daniela B Schlee, and Franz-Rainer Matuschka. Reservoir competence of various rodents
572 for the lyme disease spirochete borrelia spielmanii. *Applied and environmental microbiology*, 2011.
- 573 62. Adriana Rico, Pavel Kindlmann, and František Sedláček. Can the barrier effect of highways cause genetic
574 subdivision in small mammals? *Acta theriologica*, 54(4):297–310, 2009.
- 575 63. Naiara Rodriguez-Ezpeleta, Paula Álvarez, and Xabier Irigoien. Genetic diversity and connectivity in
576 maurolicus muelleri in the bay of biscay inferred from thousands of snp markers. *Frontiers in genetics*, 8:
577 195, 2017.
- 578 64. Aaron Shafer, Claire R Peart, Sergio Tusso, Inbar Maayan, Alan Brelsford, Christopher W Wheat, and
579 Jochen BW Wolf. Bioinformatic processing of rad-seq data dramatically impacts downstream population
580 genetic inference. *Methods in Ecology and Evolution*, 8(8):907–917, 2017.
- 581 65. Allison J Shultz, Allan J Baker, Geoffrey E Hill, Paul M Nolan, and Scott V Edwards. Snp s across time
582 and space: population genomic signatures of founder events and epizootics in the house finch (haemorrhous
583 mexicanus). *Ecology and evolution*, 6(20):7475–7489, 2016.
- 584 66. M. Velickovic. Measures of the developmental stability, body size and body condition in the black-striped
585 mouse (apodemus agrarius) as indicators of a disturbed environment in northern serbia. *Belgian Journal of*
586 *Zoology*, 137(2):147, 2007.

587 Additional Files

588 Additional file 1 — Supplementary Materials