



**HAL**  
open science

## A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species

Wim Gorssen, Roel Meyermans, Steven Janssens, Nadine Buys

### ► To cite this version:

Wim Gorssen, Roel Meyermans, Steven Janssens, Nadine Buys. A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. *Genetics Selection Evolution*, 2021, 53 (1), pp.2. 10.1186/s12711-020-00599-7 . hal-03105251

**HAL Id: hal-03105251**

**<https://hal.science/hal-03105251>**

Submitted on 11 Jan 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

Open Access



# A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species

Wim Gorssen<sup>†</sup>, Roel Meyermans<sup>†</sup>, Steven Janssens and Nadine Buys<sup>\*†</sup>

## Abstract

**Background:** Runs of homozygosity (ROH) have become the state-of-the-art method for analysis of inbreeding in animal populations. Moreover, ROH are suited to detect signatures of selection via ROH islands and are used in other applications, such as genomic prediction and genome-wide association studies (GWAS). Currently, a vast amount of single nucleotide polymorphism (SNP) data is available online, but most of these data have never been used for ROH analysis. Therefore, we performed a ROH analysis on large medium-density SNP datasets in eight animal species (cat, cattle, dog, goat, horse, pig, sheep and water buffalo; 442 different populations) and make these results publicly available.

**Results:** The results include an overview of ROH islands per population and a comparison of the incidence of these ROH islands among populations from the same species, which can assist researchers when studying other (livestock) populations or when looking for similar signatures of selection. We were able to confirm many known ROH islands, for example signatures of selection for the *myostatin* (*MSTN*) gene in sheep and horses. However, our results also included multiple other ROH islands, which are common to many populations and not identified to date (e.g. on chromosomes D4 and E2 in cats and on chromosome 6 in sheep).

**Conclusions:** We are confident that our repository of ROH islands is a valuable reference for future studies. The discovered ROH island regions represent a unique starting point for new studies or can be used as a reference for future studies. Furthermore, we encourage authors to add their population-specific ROH findings to our repository.

## Background

Runs of homozygosity (ROH) are long continuous homozygous stretches in the genome and are formed by the combination of two identical haplotypes in an individual [1]. Broman and Weber [2] first identified these long homozygous segments in the human genome and Gibson et al. [3] described their potential for inbreeding assessment. A genomic inbreeding coefficient based on ROH ( $F_{ROH}$ ) was first defined by McQuillan et al.

[4]. Since 2010, analysis of ROH has become a standard approach to study inbreeding and detect signatures of selection in animal populations with the first reported studies in 2010 for cattle [5], in 2010 for dogs [6], in 2012 for pigs [7], in 2013 for horses [8], in 2014 for goats [9], in 2015 for sheep [10], in 2016 for cats [11] and in 2020 for water buffaloes [12]. Moreover, ROH analyses are complementary to genome-wide association studies (GWAS), inbreeding depression studies, genomic prediction and detection of deleterious variants and population-specific major genes [1, 13].

ROH analyses allow the accurate estimation of  $F_{ROH}$  on both the population and the individual level. It is commonly accepted that short ROH are indicators of distant

\*Correspondence: nadine.buys@kuleuven.be

<sup>†</sup>Wim Gorssen and Roel Meyermans share first authorship  
Livestock Genetics, Department of Biosystems, KU Leuven, Kasteelpark Arenberg 30, Box 2472, 3001 Leuven, Belgium



consanguinity, whereas long ROH are more likely the result of recent inbreeding [14]. However, ROH may also result from small inversions that suppress recombination [1] or demographic events [15], such as a population bottleneck, genetic drift, or (artificial) selection. Moreover, it has been shown that ROH are more likely to arise in genomic regions with a low recombination rate and high linkage disequilibrium (LD), such as for the X-chromosome and near the centromere of chromosomes [1, 16]. In addition, copy number variants (CNV) and/or coverage gaps may lead to artefacts in ROH analyses [17]. The effect of gaps in SNP coverage can be reduced by adjusting the parameters set for the ROH analysis, as discussed by Meyermans et al. [18].

ROH facilitate the investigation of highly inbred genomic regions within a population, first referred to as ROH islands by Nothnagel et al. [19]. These ROH islands can provide important insights into the studied population and are likely to be signatures of positive selection due to LD [15, 19–21]. However, currently many populations have not been studied for the identification of ROH islands, although several large online genotype datasets are available for various species. Furthermore, even when such populations have been investigated for ROH islands, it is often difficult to compare the results between studies, for example because of differences in analysis methods or detection criteria.

In this paper, we provide an overview of ROH islands in 442 populations (18,633 individuals) from eight animal species (cat, cattle, dog, goat, horse, pig, sheep and water buffalo) using medium-density SNP data, which were all analyzed using a standardized protocol. The outcome and R-script of our analyses are made online available and can be used as a reference for future studies (<https://doi.org/10.17605/OSF.IO/XJTKV>). Since ROH islands

are potential signatures of selection, overlapping ROH islands across populations and species are a valuable tool in comparative genomic studies and may reveal important genetic regions.

## Methods

Medium-density SNP data from eight species (cat, cattle, dog, goat, horse, pig, sheep, and water buffalo) and 797 populations were collected from online available datasets (Table 1). Detailed information and background on these data are described in the corresponding studies.

Quality control and ROH analyses were performed using the PLINK v1.9 software [22]. Each population was subjected to the following quality control (PLINK commands in brackets). For SNPs, only autosomal SNPs were retained, and neither minor allele frequency pruning (--maf), no Hardy–Weinberg equilibrium test (--hwe), nor LD pruning were performed [18]. Individual call rate was set to 0.90 (--mind 0.10) and possibly duplicated individuals were removed (--genome; PI\_HAT > 0.95). Minimal SNP call rate was set to 0.95 (--geno 0.05) and only populations with more than 15 individuals were retained after quality control.

For the ROH analysis (--homozyg), no heterozygous SNP was allowed (--homozyg-window-het and --homozyg-het) and one SNP could be missing (--homozyg-window-missing) [35, 36]. The minimal number of SNPs per window (--homozyg-window-snp) and in the final ROH segment (--homozyg-snp) were breed specifically calculated by the L-parameter [18, 35, 37] and a window had a minimal size of 1000 kb (--homozyg-kb). The density was set to 150 kb/SNP, thus 1 SNP every 150 kb (--homozyg-density), the maximal gap was 1000 kb (--homozyg-gap), and the window threshold set to two outer SNPs [18]. Average SNP

**Table 1** Data overview per species

Animal species	Populations before QC	Animals before QC	Populations after QC	Animals after QC	SNPs before QC	References
Cat	47	2078	26	1657	58,888	[23]
Cattle	144	4103	50	2263	45,926	[28, 49–58]
Dog	146	5406	49	4414	160,432	[24]
Goat	143	4653	96	4327	49,943	[25, 47]
Horse	37	795	35	774	50,042	[26, 48]
Pig	146	2113	78	1438	52,783	[27]
Sheep	118	3609	100	3490	52,413	[28–33]
Water buffalo	16	346	8	270	53,830	[34]
Total	797	23,103	442	18,633		

The number of populations and individuals are shown for the raw datasets and after applying quality control. The number of autosomal SNPs available per species before quality control is also shown. The main cause for excluding a population was a lack of a sufficient number of individuals ( $n < 15$ )

QC quality control, SNPs single nucleotide polymorphisms

density was at least one SNP per 55 kb for all populations. Consequently, genome coverage was higher than 97% for all breeds, which means that the given settings allowed ROH detection for more than 97% of the autosomal genome.

ROH incidence was calculated as the percentage of animals with a SNP within an ROH segment for a given population and were visualised via Manhattan plots using the qqman package [38]. ROH islands were defined as regions where SNPs had a P-value for ROH incidence higher than a population specific threshold. This threshold was calculated based on standard normal z-scores derived from the distribution of ROH incidences. The top 0.1% of SNPs with a P-value higher than 0.999 using a z-score table for ROH incidence were considered to form ROH islands, as specified by Purfield et al. [15] and Gorssen et al. [39]. As an additional restriction, the minimal threshold for detection of ROH islands was set to 30%, which means that a ROH has to be present in at least 30% of the population to be included in a ROH island. For populations with high levels of inbreeding (e.g. Boxer dogs with a mean  $F_{ROH}$  of 45%), no SNP reached a P-value > 0.999 for ROH incidence. Therefore, a maximal threshold for ROH island detection was set at 80%, meaning that all ROH with an incidence higher than 80% were marked as ROH islands.

## Results

An R-script was developed for standardized breed-by-breed quality control and ROH analysis. This script uses a PLINK-format genotype file (.bim, .bed and .fam), with a unique family ID (FID) for each population. First, parameter settings are specified for quality control and ROH analysis. Second, the R-script performs quality control and a ROH analysis per population (FID). Third, the script creates Manhattan plots based on ROH incidence per SNP for every investigated population, and a summary table. For the 442 populations studied here, all figures and the R-script are deposited at Open Science Framework (OSF) (<https://doi.org/10.17605/OSF.IO/XJTKV>).

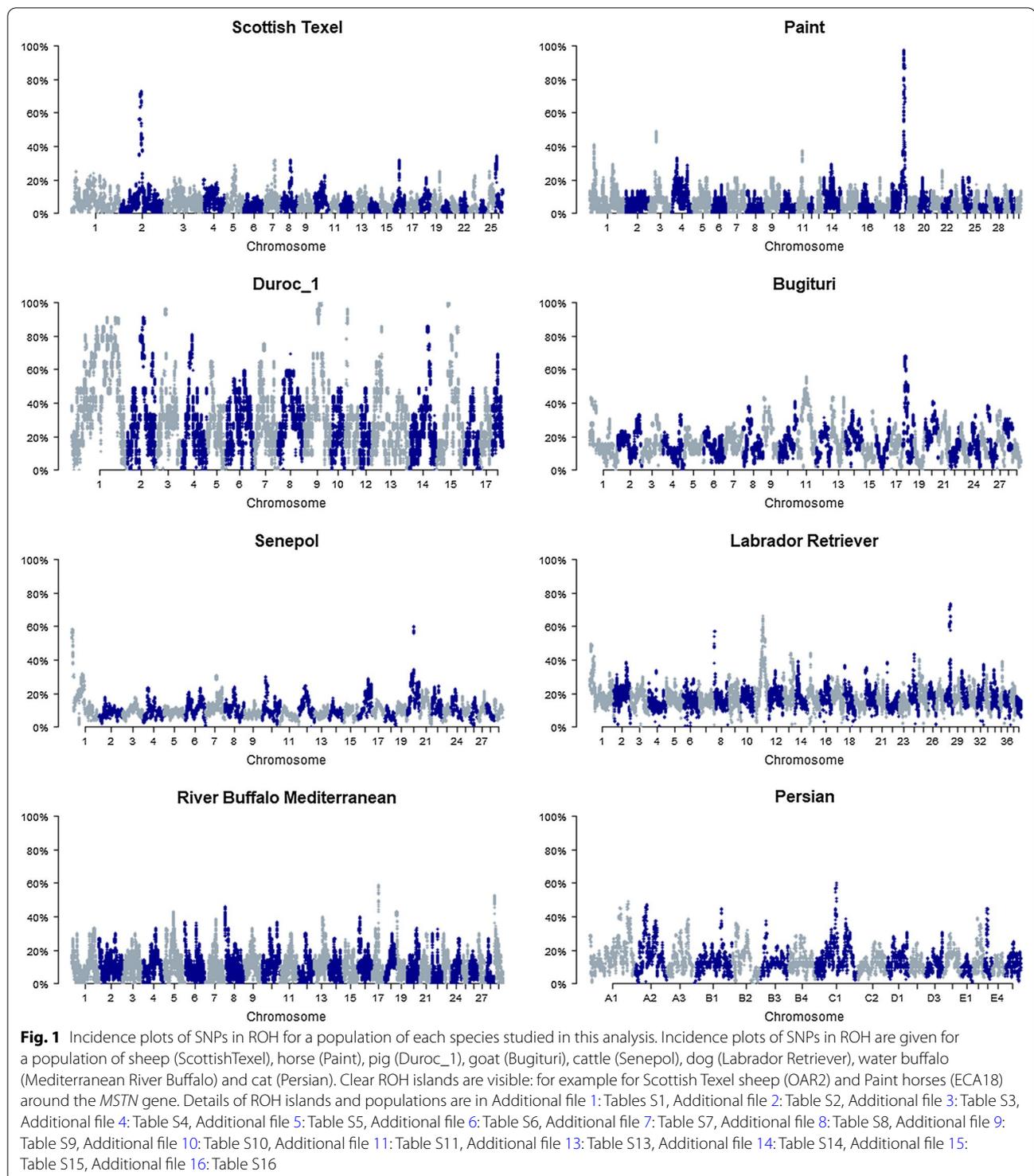
Figures 1 and 2 show an example of the ROH incidence plots. These Manhattan plots provide a quick overview of population-specific baseline ROH levels and ROH islands. For example, in Paint horses, ROH incidence levels are generally low (0 to 15%) but one remarkable ROH island is observed on *Equus caballus* chromosome (ECA)18 at 68 Mb. This ROH island was found in 23 out of the 24 studied horses in the Paint population (Fig. 1). Duroc pigs have higher baseline ROH levels (0 to 60%), with several ROH islands having an incidence higher than 80% (Fig. 1). To investigate ROH islands in more detail, we created tables with ROH island locations (bins

of one Mb) per population, which are in Additional file 1: Tables S1, Additional file 2: Table S2, Additional file 3: Table S3, Additional file 4: Table S4, Additional file 5: Table S5, Additional file 6: Table S6, Additional file 7: Table S7, Additional file 8: Table S8, Additional file 9: Table S9, Additional file 10: Table S10, Additional file 11: Table S11, Additional file 12: Table S12, Additional file 13: Table S13, Additional file 14: Table S14, Additional file 15: Table S15, Additional file 16: Table S16 and in the repository. As an example, detailed information on the ROH islands per breed and chromosomal region for cats is provided in Table 2. These figures and tables can be used to detect overlapping ROH island regions in multiple populations. In cats, for example, 19 out of the 26 studied populations (73%) have a ROH island on chromosome B3 around 27–28 Mb (Table 2). In sheep, 15 of the 100 studied populations show a ROH island on *Ovis aries* chromosome (OAR)6 at ~38 Mb (Fig. 2, Additional file 13: Table S13 and Additional file 14: Table S14).

## Discussion

We created a repository of ROH-islands for 442 populations from eight animal species (cat, cattle, dog, goat, horse, pig, sheep and water buffalo). These results are available online via OSF (<https://doi.org/10.17605/OSF.IO/XJTKV>) and examples are provided in Figs. 1 and 2, Table 2 and Additional file 1: Tables S1, Additional file 2: Table S2, Additional file 3: Table S3, Additional file 4: Table S4, Additional file 5: Table S5, Additional file 6: Table S6, Additional file 7: Table S7, Additional file 8: Table S8, Additional file 9: Table S9, Additional file 10: Table S10, Additional file 11: Table S11, Additional file 12: Table S12, Additional file 13: Table S13, Additional file 14: Table S14, Additional file 15: Table S15, and Additional file 16: Table S16.

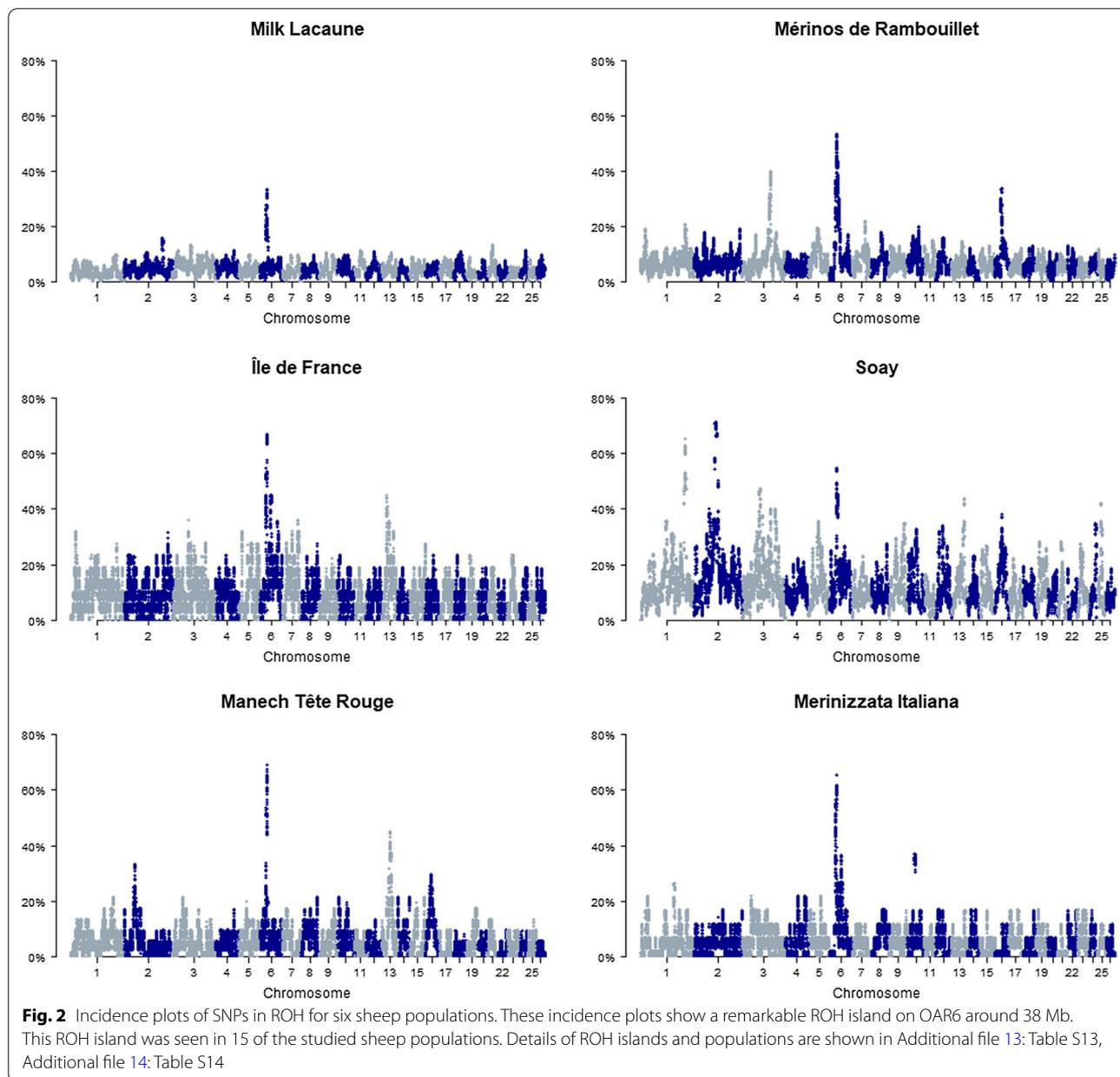
For many populations, the figures of ROH incidence show interesting ROH islands. For example, previously known ROH islands are observed around the *myostatin* (*MSTN*) gene in Texel sheep (OAR2, 129 Mb) and Paint and Quarter horses (ECA18, 66 Mb) (Fig. 1) and Additional file 9: Table S9, Additional file 10: Table S10, Additional file 13: Table S13, and Additional file 14: Table S14. Both Purfield et al. [15] and Fariello et al. [40] found a signature of selection in the region spanning the *MSTN* gene in Texel sheep using different methods [15, 40]. Petersen et al. [26] showed the existence of a clear signature of selection in the genomic region around *MSTN* in Paint and Quarter Horses using  $F_{ST}$ -based statistics [26]. *MSTN* is a major gene involved in muscle development. In sheep, selection on conformation has apparently led to a common ROH island, spanning the *MSTN* gene. In cats, we found that 19 of the 26 populations showed a ROH island on chromosome B3 between 20 and 30 Mb



(Table 2). Montague et al. [41] found strong signatures of selection in cats in these regions on chromosome B3 using  $F_{ST}$ -based analysis [41]. They suggested that the *ARID3B* gene—which is known to impact neural crest cell survival—might be the driving factor for the selection

signature on chromosome B3 and they linked it to the domestication syndrome hypothesis [42].

Although the above-mentioned ROH islands had an assumed/known underlying biological factor driving positive selection, we detected multiple ROH islands



in different species without a clear link to the underlying (biological) mechanisms, so far. These ROH islands might indicate genetic regions under positive selection, although they could also be the result of population bottlenecks, regions with repressed recombination, or artefacts caused by CNV or SNP gaps [1, 13, 17]. We observed that ROH and ROH islands occurred more frequently near the centromere of chromosomes, possibly due to a difference in recombination levels [7, 17]. We minimized ROH artefacts that are caused by large SNP gaps, by optimizing ROH detection in PLINK as described in Meyermans et al. [18], in particular by adjusting the minimal SNP

density for ROH detection. However, hemizygous deletions that are especially large might still resemble a ROH signal in SNP data [1, 4, 17]. To address this and also the possible interference between ROH and CNV via SNP data, raw genotyping data (e.g. Illumina final report files) are required. These are often not available for such large datasets and could not be examined in this study. Moreover, ascertainment bias might also impact ROH analysis as genotyping arrays can be less suitable for populations that are only distantly related to the populations used for array development. In this study, ascertainment bias was minimized by calculating the L-parameter, which takes

**Table 2 ROH island regions (bins of one Mb) for 26 cat populations listed per chromosome**

Population	Chromosome																	
	A1	A2	A3	B1	B2	B3	B4	C1	C2	D1	D2	D3	D4	E1	E2	E3	F1	F2
Abyssinian				34-42		27-30			65-91									
American Curl		17-18				143-146										15-41		
Bengal		87-90					82-89	39-40			60-77				37-38			
British Shorthair	83-87					24-30	32-37						33-33					
Burmese				35-49		22-30				26-55		28-32						
Colony						25-25									37-38			
Devon Rex							35-83											
Domestic						26-29												
LaPerm	83-84					27-29							31-34		25-38			
Lykoi						26-29									23-38			
Maine Coon	0-4					27-28							35-38			15-17		
Munchkin						25-29							30-33					
Norwegian Forest Cat	127-129	98-99				23-30	33-35								37-38			
Oriental	83-87					23-30	32-34	166-174	91-93									
Oriental Toygers	157-159		130-131			27-30	114-119					27-30			49-49			
Persian	160-203	58-61		149-150				104-112								15-17		
Peterbald	181-193					23-58		166-168	91-122									
Ragdoll	54-88			40-61		26-29				46-46								
Scottish Fold		101-103	49-50			25-30												
Selkirk Rex						24-29		103-110							37-39			
Siamese	152-186		49-52			25-30	32-34	166-174	91-93	43-45; 92-92					37-40			
Siberian	186-186				73-73	24-30							33-33					
Sphynx								136-140								53-57	15-17	
Tennessee Rex						25-31				22-41								
Turkish Van	34-34																	31-33
Wildcat																		

Within each cell of the table, the size of genomic region(s) with a ROH island is indicated in Mb. ROH islands were detected for multiple populations on e.g. chromosome B3 (25-29 Mb), E2 (37-38 Mb) and D4 (31-33 Mb)

potential heterozygosity differences into account [18, 35, 37] and by setting the minimum ROH length to 1000 kb. Furthermore, we would like to note that the presence of a ROH island does not implicate an identical underlying haplotype. For example, in the context of selection, ROH are likely to arise in genomic regions that contain important genes, but the underlying genotypes might differ among populations and individuals [39].

Regardless of their origin, ROH islands can provide a valuable clue for future research. For example, 15 sheep populations in this study show a ROH island on OAR6 at 38 Mb (Fig. 2). OAR6 appears to harbour multiple genes that are linked with milk production in sheep, but we also found several QTL for fat-tail, growth and bone-related traits near the region around 38 Mb [43, 44]. As another example, seven cat populations showed a ROH island on chromosome E2 (37–38 Mb) and six cat populations on chromosome D4 (33 Mb) (Table 2). However, to our knowledge, these regions have never been reported in the literature as regions of interest. Therefore, these overlapping ROH islands can be a starting point for researchers investigating signatures of selection but also for studies on islands of speciation, recombination hot- and coldspots and population history. Furthermore, researchers can compare the outcome of their (ROH) studies with our repository. Comparing (ROH) results among studies reported in the literature is often difficult, due to differences in quality control, software and parameter settings. The advantage of our repository is that it combines information on multiple populations and species which were analysed using a standardized method.

To define ROH islands, we implemented a population-specific threshold based on the ROH incidence distribution, which allows a comparison of ROH islands across different populations. In this study, we used PLINK for ROH detection as it is still the most frequently used software for these analyses [13]. Especially when using medium-density SNP data, it is essential to optimize (PLINK) ROH detection settings [18]. For example, the default value for density setting in PLINK is one SNP per 50 kb, which is higher than the mean density for our sheep (one SNP per 50.5 kb) and cattle (one SNP per 54.5 kb) SNP data even before quality control. Thus, using the default PLINK values can dramatically underestimate the number of ROH detected. Our recommendation for future studies using PLINK is to carefully consider these parameters, to make sure results are as correct and comparable as possible. To facilitate this, we share our R-script with other researchers upon proper citation (<https://doi.org/10.17605/OSF.IO/XJTKV>). We would like to draw attention to the fact that the dog data had a higher average SNP density (one SNP per 14 kb).

However, differences in SNP density were accounted for by following Meyermans et al. [18]. Besides PLINK, other algorithms are also available for ROH detection, for example RzooRoH [45, 46]. Results obtained by using these software can also be included in our collection, since differences between rule-based (e.g., PLINK) and model-based (e.g., RzooRoH) approaches should be small when performed correctly on medium-density SNP data [39, 45, 46].

## Conclusions

We have shown that important ROH islands can be detected by scanning multiple populations simultaneously for ROH islands using a standardized detection method. We provide our script for standardized ROH island analyses and make all the results publicly available via OSF (<https://doi.org/10.17605/OSF.IO/XJTKV>). By sharing our results, our aim is to give researchers a useful reference to compare with their own analyses or to provide a unique starting point to investigate specific signatures of selection. Moreover, we encourage authors of future ROH studies to add their Manhattan plots of ROH incidence to our collection. We strongly believe that this ROH island repository will be very valuable for comparisons with future (ROH) studies or as a starting point for new studies.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12711-020-00599-7>.

**Additional file 1: Table S1.** ROH island regions (bins of one Mb) for the studied cat populations: overview for all chromosomes. Within each cell, the genomic region(s) with an ROH island are given (in Mb).

**Additional file 2: Table S2.** ROH island regions (bins of one Mb) for the studied cat populations: details per chromosome. Within the Excel file, there is one tab for each chromosome with ROH island details per population. Cells are marked with an 'x' if an ROH island appeared in the specific region.

**Additional file 3: Table S3.** ROH island regions (bins of one Mb) for the studied cattle populations: overview for all chromosomes. Within each cell, the genomic region(s) with an ROH island are given (in Mb).

**Additional file 4: Table S4.** ROH island regions (bins of one Mb) for the studied cattle populations: details per chromosome. Within the Excel file, there is one tab for each chromosome with ROH island details per population. Cells are marked with an 'x' if an ROH island appeared in the specific region.

**Additional file 5: Table S5.** ROH island regions (bins of one Mb) for the studied dog populations: overview for all chromosomes. Within each cell, the genomic region(s) with an ROH island are given (in Mb).

**Additional file 6: Table S6.** ROH island regions (bins of one Mb) for the studied dog populations: details per chromosome. Within the Excel file, there is one tab for each chromosome with ROH island details per population. Cells are marked with an 'x' if an ROH island appeared in the specific region.

**Additional file 7: Table S7.** ROH island regions (bins of one Mb) for the studied goat populations: overview for all chromosomes. Within each cell, the genomic region(s) with an ROH island are given (in Mb).

**Additional file 8: Table S8.** ROH island regions (bins of one Mb) for the studied goat populations: details per chromosome. Within the Excel file, there is one tab for each chromosome with ROH island details per population. Cells are marked with an 'x' if an ROH island appeared in the specific region.

**Additional file 9: Table S9.** ROH island regions (bins of one Mb) for the studied horse populations: overview for all chromosomes. Within each cell, the genomic region(s) with an ROH island are given (in Mb).

**Additional file 10: Table S10.** ROH island regions (bins of one Mb) for the studied horse populations: details per chromosome. Within the Excel file, there is one tab for each chromosome with ROH island details per population. Cells are marked with an 'x' if an ROH island appeared in the specific region.

**Additional file 11: Table S11.** ROH island regions (bins of one Mb) for the studied pig populations: overview for all chromosomes. Within each cell, the genomic region(s) with an ROH island are given (in Mb).

**Additional file 12: Table S12.** ROH island regions (bins of one Mb) for the studied pig populations: details per chromosome. Within the Excel file, there is one tab for each chromosome with ROH island details per population. Cells are marked with an 'x' if an ROH island appeared in the specific region.

**Additional file 13: Table S13.** ROH island regions (bins of one Mb) for the studied sheep populations: overview for all chromosomes. Within each cell, the genomic region(s) with an ROH island are given (in Mb).

**Additional file 14: Table S14.** ROH island regions (bins of one Mb) for the studied sheep populations: details per chromosome. Within the Excel file, there is one tab for each chromosome with ROH island details per population. Cells are marked with an 'x' if an ROH island appeared in the specific region.

**Additional file 15: Table S15.** ROH island regions (bins of one Mb) for the studied water buffalo populations: overview for all chromosomes. Within each cell, the genomic region(s) with an ROH island are given (in Mb).

**Additional file 16: Table S16.** ROH island regions (bins of one Mb) for the studied water buffalo populations: details per chromosome. Within the Excel file, there is one tab for each chromosome with ROH island details per population. Cells are marked with an 'x' if an ROH island appeared in the specific region.

#### Acknowledgements

Not applicable.

#### Authors' contributions

WG and RM analyzed the data and wrote the manuscript. WG, RM, SJ and NB designed and conceived this study. SJ and NB critically reviewed the analyses and the manuscript. All authors read and approved the final manuscript.

#### Funding

This study was partially funded by an SB PhD fellowship (1S37119N) and an FR PhD fellowship (1104320N) of the Research Foundation Flanders (FWO). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

#### Availability of data and materials

The datasets analysed during the current study are available in the following repositories.

##### Cat

Gandolfi et al. [23]. Data available on: <https://www.nature.com/articles/s41598-018-25438-0#Sec26>.

##### Cattle

Sempéré et al. [28]. Data available on: <http://widde.toulouse.inra.fr/widde/>.

Batch selection of all populations with SNP data from Illumina Bovine SNP50v1, Illumina Bovine SNP50v2 and Illumina BovineHD. Only common markers (46,387) were selected from all chromosomes.

##### Dog

Shannon et al. [24]. Data available on: <https://datadryad.org/stash/dataset/doi:10.5061/dryad.v9t5h>.

##### Goat

Colli et al. [25] and data from Bertolini et al. [47]. Dryad Digital Repository. <https://doi.org/10.5061/dryad.v8g21pt>.

##### Horse

Petersen et al. [26] and data from Petersen et al. [48]. NAGPR Community Data Repository. <https://www.animalgenome.org/repository/pub/UMN2013.0125/>.

##### Pig

Yang et al. [27]. Data at the Dryad Digital Repository. <https://doi.org/10.5061/dryad.30tk6>.

##### Sheep

Sempéré et al. [28]. Data available on: <http://widde.toulouse.inra.fr/widde/>. Batch selection of all populations with SNP data from Illumina OvineSNP50v1 and AgResearch OvineHD. Only common markers (42,439) were selected from all chromosomes.

##### Water buffalo

Colli et al. [34]. Data available at: <https://datadryad.org/stash/dataset/doi:10.5061/dryad.h0cc7>.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

Received: 30 April 2020 Accepted: 11 December 2020

Published online: 04 January 2021

#### References

- Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: Windows into population history and trait architecture. *Nat Rev Genet.* 2018;19:220–34.
- Broman KW, Weber JL. Long homozygous chromosomal segments in reference families from the Centre d'Étude du Polymorphisme Humain. *Am J Hum Genet.* 1999;65:1493–500.
- Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet.* 2006;15:789–95.
- McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. *Am J Hum Genet.* 2008;83:359–72.
- Sölkner J, Ferenčaković M, Fürst C, Čurik I. Genomic metrics of individual autozygosity, applied to a cattle population. In Proceedings of the 61st Annual Meeting of the European Association for Animal Production: 23–27 August 2010; Heraklion. 2010.
- Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, et al. A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol.* 2010;8:e1000451.
- Bosse M, Megens HJ, Madsen O, Paudel Y, Frantz LAF, Schook LB, et al. Regions of homozygosity in the porcine genome: Consequence of demography and the recombination landscape. *PLoS Genet.* 2012;8:e1003100.
- Khanshour AM. Genetic diversity and population structure of the Arabian horse populations from Syria and other countries. PhD thesis, Texas A&M University. 2013.
- Guangul SA. Design of community based breeding programs for two indigenous goat breeds of Ethiopia. PhD thesis, University of Natural Resources and Life Sciences Vienna. 2014.
- Beynon SE, Slavov GT, Farré M, Sunduimijid B, Waddams K, Davies B, et al. Population structure and history of the Welsh sheep breeds determined by whole genome genotyping. *BMC Genet.* 2015;16:65.

11. Bertolini F, Gandolfi B, Kim ES, Haase B, Lyons LA, Rothschild MF. Evidence of selection signatures that shape the Persian cat breed. *Mamm Genome*. 2016;27:144–55.
12. Ghoreishifar SM, Moradi-Shahrbabak H, Fallahi MH, Jalil Sarghale A, Moradi-Shahrbabak M, Abdollahi-Arpanahi R, et al. Genomic measures of inbreeding coefficients and genome-wide scan for runs of homozygosity islands in Iranian river buffalo *Bubalus bubalis*. *BMC Genet*. 2020;21:16.
13. Peripolli E, Munari DP, Silva MVGB, Lima ALF, Irgang R, Baldi F. Runs of homozygosity: current knowledge and applications in livestock. *Anim Genet*. 2017;48:255–71.
14. Curik I, Ferenčaković M, Sölkner J. Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest Sci*. 2014;166:26–34.
15. Purfield DC, McParland S, Wall E, Berry DP. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS One*. 2017;12:e0176780.
16. Ma L, O'Connell JR, VanRaden PM, Shen B, Padhi A, Sun C, et al. Cattle sex-specific recombination and genetic control from a large pedigree analysis. *PLoS Genet*. 2015;11:e1005387.
17. Nandolo W, Utsunomiya YT, Mészáros G, Wurzinger M, Khayadzadeh N, Torrecilha RBP, et al. Misidentification of runs of homozygosity islands in cattle caused by interference with copy number variation or large intermarker distances. *Genet Sel Evol*. 2018;50:43.
18. Meyermans R, Gorssen W, Buys N, Janssens S. How to study runs of homozygosity using plink? a guide for analyzing medium density snp data in livestock and pet species. *BMC Genomics*. 2020;21:94.
19. Nothnagel M, Lu TT, Kayser M, Krawczak M. Genomic and geographic distribution of snpdefined runs of homozygosity in Europeans. *Hum Mol Genet*. 2010;19:2927–35.
20. Peripolli E, Stafuzza NB, Munari DP, Lima ALF, Irgang R, Machado MA, et al. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle. *BMC Genomics*. 2018;19:34.
21. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*. 2002;419:832–7.
22. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
23. Gandolfi B, Alhaddad H, Abdi M, Bach LH, Creighton EK, Davis BW, et al. Applications and efficiencies of the first cat 63K DNA array. *Sci Rep*. 2018;8:7024.
24. Shannon LM, Boyko RH, Castelhamo M, Corey E, Hayward JJ, McLean C, et al. Genetic structure in village dogs reveals a Central Asian domestication origin. *Proc Natl Acad Sci USA*. 2015;112:13639–44.
25. Colli L, Milanese M, Talenti A, Bertolini F, Chen M, Crisà A, et al. Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. *Genet Sel Evol*. 2018;50:58.
26. Petersen JL, Mickelson JR, Rendahl AK, Valberg SJ, Andersson LS, Axelsson J, et al. Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet*. 2013;9:e1003211.
27. Yang B, Cui L, Perez-Enciso M, Traspov A, Crooijmans RPMA, Zinovieva N, et al. Genome-wide SNP data unveils the globalization of domesticated pigs. *Genet Sel Evol*. 2017;49:71.
28. Sempéré G, Moazami-Goudarzi K, Eggen A, Laloë D, Gautier M, Flori L. WIDDE: A Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. *BMC Genomics*. 2015;16:940.
29. Ciani E, Lasagna E, D'Andrea M, Alloggio I, Marroni F, Ceccobelli S, et al. Merino and Merino-derived sheep breeds: a genome-wide intercontinental study. *Genet Sel Evol*. 2015;47:64.
30. Kijas JW, Lenstra JA, Hayes B, Boitard S, Neto LR, Cristobal MS, et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol*. 2012;10:e1001258.
31. Rochus CM, Tortereau F, Plisson-Petit F, Restoux G, Moreno-Romieux C, Tossier-Klopp G, et al. Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep. *BMC Genomics*. 2018;19:71.
32. Pickering NK, Young EA, Kijas JW, Scobie DR, Mcewan JC. Genetic origin of Arapawa sheep and adaptation to a feral lifestyle. *Proc Assoc Advmt Anim Breed Genet*. 2011;20:451–4.
33. Young EA, Kijas JW, McCulloch R, Scobie DR, McRae KM, Pickering NK, et al. Arapawa: a novel New Zealand sheep breed of distinct origin. *Proc New Zeal Soc Anim Prod*. 2011;71:248–50.
34. Colli L, Milanese M, Vajana E, Iamartino D, Bombà L, Puglisi F, et al. New insights on water buffalo genomic diversity and post-domestication migration routes from medium density SNP chip data. *Front Genet*. 2018;9:53.
35. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. *BMC Genet*. 2012;13:70.
36. Ferenčaković M, Sölkner J, Curik I. Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. *Genet Sel Evol*. 2013;45:42.
37. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci USA*. 2007;104:19942–7.
38. Turner DS. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J Open Source Softw*. 2018;3:731.
39. Gorssen W, Meyermans R, Buys N, Janssens S. SNP genotypes reveal breed substructure, selection signatures and highly inbred regions in Piétrain pigs. *Anim Genet*. 2020;51:32–42.
40. Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*. 2013;193:929–41.
41. Montague MJ, Li G, Gandolfi B, Khan R, Aken BL, Searle SMJ, et al. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc Natl Acad Sci USA*. 2014;111:17230–5.
42. Wilkins AS, Wrangham RW, Tecumseh FW. The “domestication syndrome” in mammals: A unified explanation based on neural crest cell behavior and genetics. *Genetics*. 2014;197:795–808.
43. Ruiz-Larrañaga O, Langa J, Rendo F, Manzano C, Iriando M, Estonba A. Genomic selection signatures in sheep from the Western Pyrenees. *Genet Sel Evol*. 2018;50:9.
44. Matika O, Riggio V, Anselme-Moizan M, Law AS, Pong-Wong R, Archibald AL, et al. Genome-wide association reveals QTL for growth, bone and in vivo carcass traits as assessed by computed tomography in Scottish Blackface lambs. *Genet Sel Evol*. 2016;48:11.
45. Bertrand AR, Kadri NK, Flori L, Gautier M, Druet T. RZooRoH: An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. *Methods Ecol Evol*. 2019;10:860–6.
46. Druet T, Gautier M. A model-based approach to characterize individual inbreeding at both global and local genomic scales. *Mol Ecol*. 2017;26:5820–41.
47. Bertolini F, Servin B, Talenti A, Rochat E, Kim ES, Oget C, et al. Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genet Sel Evol*. 2018;50:57.
48. Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, et al. Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One*. 2013;8:e54997.
49. Gautier M, Laloë D, Moazami-Goudarzi K. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS ONE*. 2010;5(9):e13038.
50. Decker JE, McKay SD, Rolf MM, Kim JW, Molina Alcalá A, Sonstegard TS, et al. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet*. 2014;10(3):e1004254.
51. Flori L, Thevenon S, Dayo GK, Senou M, Sylla S, Berthier D, et al. Adaptive admixture in the West African bovine hybrid zone: Insight from the Borgou population. *Mol Ecol*. 2014;23(13):3241–57.
52. Flori L, Gonzatti MI, Thevenon S, Chantal I, Pinto J, Berthier D, et al. A quasi-exclusive European ancestry in the Senepol tropical cattle breed highlights the importance of the slick locus in tropical adaptation. *PLoS One*. 2012;7(5):e36133.
53. Gao Y, Gautier M, Ding X, Zhang H, Wang Y, Wang X, et al. Species composition and environmental adaptation of indigenous Chinese cattle. *Sci Rep*. 2017;7(1):16196.
54. Gautier M, Flori L, Riebler A, Jaffrézic F, Laloë D, Gut I, et al. A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics*. 2009;10:550.

55. Iso-Touru T, Tapio M, Vilkki J, Kiseleva T, Ammosov I, Ivanova Z, et al. Genetic diversity and genomic signatures of selection among cattle breeds from Siberia, eastern and northern Europe. *Anim Genet*. 2016;47(6):647–57.
56. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, et al. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One*. 2009;4(4):e5350.
57. Flori L, Moazami-Goudarzi K, Alary V, Araba A, Boujenane I, Boushaba N, et al. A genomic map of climate adaptation in Mediterranean cattle breeds. *Mol Ecol*. 2019;28(5):1009–29.
58. Illumina Agrigenomics. Available from: <https://emea.illumina.com/areas-of-interest/agrigenomics.html?langsel=/fr/>. Accessed 15 Jan 2020.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

