An Annotated Draft Genome of the Mountain Hare (*Lepus timidus*)

João P. Marques [1,2,3,†], Fernando A. Seixas[1,2,3,†], Liliana Farelo[1], Colin M. Callahan[4], Jeffrey M. Good[4,5], W. Ian Montgomery[6], Neil Reid[6], Paulo C. Alves [1,2,5], Pierre Boursot [3,*], and José Melo-Ferreira [1,2,*]

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, Vairão, Portugal

[2]Departamento de Biologia, Faculdade de Ciências do Porto, Portugal

[3]Institut des Sciences de l'Evolution Montpellier (ISEM), Univ Montpellier, CNRS, IRD, EPHE, France

[4]Division of Biological Sciences, University of Montana

[5]Wildlife Biology Program, University of Montana

[6]Institute of Global Food Security, School of Biological Sciences, Queen's University Belfast, United Kingdom

†These authors contributed equally to this work.

*Corresponding authors: E-mails: pierre.boursot@umontpellier.fr; jmeloferreira@cibio.up.pt.

Accepted: December 7, 2019

Data deposition: This project has been deposited in the NCBI Sequence Read Achive Database under BioProject PRJNA561582. The hare pseudo-reference genome is available in Dryad: https://doi.org/10.5061/dryad.x95x69pd8.

Abstract

Hares (genus *Lepus*) provide clear examples of repeated and often massive introgressive hybridization and striking local adaptations. Genomic studies on this group have so far relied on comparisons to the European rabbit (*Oryctolagus cuniculus*) reference genome. Here, we report the first de novo draft reference genome for a hare species, the mountain hare (*Lepus timidus*), and evaluate the efficacy of whole-genome re-sequencing analyses using the new reference versus using the rabbit reference genome. The genome was assembled using the ALLPATHS-LG protocol with a combination of overlapping pair and mate-pair Illumina sequencing (77x coverage). The assembly contained 32,294 scaffolds with a total length of 2.7 Gb and a scaffold N50 of 3.4 Mb. Re-scaffolding based on the rabbit reference reduced the total number of scaffolds to 4,205 with a scaffold N50 of 194 Mb. A correspondence was found between 22 of these hare scaffolds and the rabbit chromosomes, based on gene content and direct alignment. We annotated 24,578 protein coding genes by combining ab-initio predictions, homology search, and transcriptome data, of which 683 were solely derived from hare-specific transcriptome data. The hare reference genome is therefore a new resource to discover and investigate hare-specific variation. Similar estimates of heterozygosity and inferred demographic history profiles were obtained when mapping hare whole-genome re-sequencing data to the new hare draft genome or to alternative references based on the rabbit genome. Our results validate previous reference-based strategies and suggest that the chromosome-scale hare draft genome should enable chromosome-wide analyses and genome scans on hares.

Key words: whole-genome sequencing, de novo assembly, annotation, hares, Leporids, Lagomorpha.

Introduction

The ability to sequence whole genomes has revolutionized our power to study the evolution of non-model organisms. Hares (genus *Lepus*) have recently emerged as useful evolutionary models to understand introgressive hybridization and local adaptation (Alves et al. 2008; Jones et al. 2018; Seixas et al. 2018; Giska et al. 2019). Genomic analyses on this group have primarily relied on comparisons to the high-quality reference genome of another leporid, the more extensively studied European rabbit (*Oryctolagus cuniculus*) (Carneiro et al. 2014), estimated to share a most recent common ancestor with hares ~12 million years ago (Matthee et al. 2004). Although these studies have generally used iterative mapping approaches to reduce divergence and increase mapping

efficiency (e.g., Jones et al. 2018; Seixas et al. 2018), it remains unclear to what extent reliance on an outgroup reference may have limited genomic inferences.

We extend the genomic resources of Leporids by assembling the first draft genome of a hare species, the mountain hare (*Lepus timidus*). The mountain hare is an arcto-alpine species widely distributed in the northern Palearctic, from western Europe to eastern Asia, with some isolated populations, as in the Alps, Poland, Great Britain, and Ireland. The current distribution of the species reflects the colonization of previously glaciated areas in the north, and the retreat from southernmost regions in the south (Waltari and Cook 2005; Hamill *et al.* 2006; Melo-Ferreira et al. 2007; Smith et al. 2018). The species has been implicated in recurrent events of introgressive hybridization with other hare species from Europe (Thulin et al. 1997; Alves et al. 2003; Melo-Ferreira et al. 2009; Seixas et al. 2018; Giska et al. 2019), and displays important locally adapted traits, such as varying ecologies (Caravaggi et al. 2017), size differences among regions, or distinctive coat color (Smith et al. 2018; Giska et al. 2019). Furthermore, genus *Lepus* is distributed worldwide with more than 30 classified species, which show adaptions to contrasting environments, from arctic to arid regions. Detailed investigation of relevant evolutionary processes in the genus can benefit from the availability of hare-specific genomic resources (Fontanesi et al. 2016).

Materials and Methods

DNA Sampling, Extraction, and Sequencing

One female mountain hare (*Lepus timidus hibernicus*) specimen (NCBI BioSample ID SAMN12621015) was captured from the wild for scientific research purposes by the Irish Coursing Club (ICC) at Borris-in-Ossory, County Laois under National Parks & Wildlife (NPWS) license no. C 337/2012 issued by the Department of Arts, Heritage and the Gaeltacht (dated October 31, 2012). Genomic DNA was extracted from kidney, muscle, and ear tissue using the JETquick Tissue DNA Spin Kit (GENOMED), with RNAse and proteinase K to remove RNA and protein contamination. Genomic libraries of different insert lengths were generated following the standard ALLPATHS-LG protocol (Gnerre et al. 2011): one Illumina TruSeq DNA library of 180 bp fragments was sequenced with overlapping paired-end (OPE) reads, and three Illumina TruSeq DNA mate-pair (MP) libraries of 2.5, 4.5, and 8.0 kb insert sizes. Whole-genome sequencing was performed at The Genome Analysis Center (TGAC, currently Earlham Institute, Norwich, UK)—seven HiSeq2000 lanes (five OPE and two 4.5 kb MP)—and CIBIO's New-Gen sequencing platform—three HiSeq1500 lanes (2.5 and 8.0 kb MP). Raw sequencing reads were deposited in the Sequence Read Archive (details in supplementary table S1, Supplementary Material online).

Read Quality Assessment and Filtering

Exact duplicates were removed both from OPE and MP libraries using PRINSEQ v0.20.4 (Schmieder and Edwards 2011b). PhiX sequences were identified using Bowtie2-v2.2.3 (Langmead and Salzberg 2012) and removed. Adapter sequences were removed using Cutadapt v1.4.1 (Martin 2011) for OPE reads and Skewer v1.3.1 (Jiang et al. 2014) for mate-pairs. For the latter, only pairs in the correct orientation determined by the presence of the junction adapter were retained.

Genome Size Estimation

Genome size was estimated using a *k*-mer-based approach (Marçais and Kingsford 2011). First, the frequency distribution of 17 bp *k*-mers was obtained using jellyfish v2.2.6 (Marçais and Kingsford 2011) based on the OPE raw reads—supplementary figure S1, Supplementary Material online. The sequencing depth was then calculated following $M = N * (L - k + 1)/L$, where M is the peak of the *k*-mer depth frequency distribution, L is the read length, and k is the chosen *k*-mer length in bp. Finally, the genome size was estimated by dividing the total number of bases sequenced by the sequencing depth.

Genome Assembly and Annotation

De novo assembly was performed using ALLPATHS-LG (Gnerre et al. 2011) with default parameters using OPE and mate-pair reads. The resulting assembly was evaluated with REAPR v1.0.18 (Hunt et al. 2013) to break incorrect scaffolds, by mapping the paired-end and the 4.5 kb mate-pair reads on the assembled genome. Another round of scaffolding was then performed using SSPACE v3.0 (Boetzer et al. 2011), with a minimum overlap of 32 bp and supported by a minimum of 20 reads. Finally, we leveraged the existence of the high-quality assembly of the genome of the European rabbit (*Oryctolagus cuniculus—Ensembl OryCun2.0*), to improve the contiguity of the assembly using the reference-based scaffolder MeDuSa v.1.6 with five iterations (Bosi et al. 2015). This re-scaffolding orders and re-orientates scaffolds without affecting intra-scaffold sequence. Quality control of the assembly at different stages was assessed based on metrics obtained with QUAST v.3.2 (Mikheenko et al. 2016). The completeness of the *L. timidus* re-scaffolded genome was evaluated using BUSCO v.3.0.2 (Simão et al. 2015), based on the presence and absence of core single-copy genes (from mammalia_odb9 database). We then checked consistency of gene content in the larger chromosome-like scaffolds and rabbit chromosomes using blastp from NCBI BLAST v2.7.1+ (Camacho et al. 2009), considering the best hit per gene with similarity above 90% over 500 bp. The 22 rabbit chromosomes were aligned against inferred corresponding

L. timidus re-scaffolded scaffolds using D-Genies v. 1.2.0 Mashmap (Cabanettes and Klopp 2018).

Repetitive regions were identified using RepeatModeler v.1.0.11 (Smit and Hubley 2008) and masked using RepeatMasker v.4.0.7 (Smit et al. 2013). The masked genome was used as input for gene prediction in MAKER v.3.01.02 (Cantarel et al. 2008), using ab-initio predictions, *L. timidus* transcriptome data, and rabbit protein annotations (*O. cuniculus*) (supplementary text, Supplementary Material online). Functional inference for genes and transcripts was performed using the translated CDS features of each coding transcript. Each predicted protein sequence was based on blastp searches against the Uniprot-Swissprot database to retrieve gene name and function, and InterProscan v5.30-69 (Jones et al. 2014) to retrieve Interpro, Pfam v31.0 (Finn et al. 2016), GO (Mi et al. 2017), KEGG (Kanehisa et al. 2016), and Reactome (Fabregat et al. 2018) information.

Analyses of Whole-Genome Re-Sequencing Data

To compare the performance of using the *L. timidus* genome or other strategies based on the rabbit genome for whole-genome analyses, we analysed re-sequencing data from the mountain hare and another hare species, the Iberian hare, *L. granatensis*, mapping the reads to 1) the new *L. timidus* re-scaffolded genome, 2) the rabbit reference genome (available from Ensembl—OryCun2.0, release 80), and 3) a hare pseudo-reference genome built through iterative mapping of hare sequence reads on the rabbit genome, followed by reference updating (from Seixas et al. 2018). For the re-sequencing data (NCBI Sequence Read Archive Biosamples SAMN07526960 and SAMN07526963; Seixas et al. 2018), adapters were removed using cutadapt version 1.8 (Martin 2011) and low quality bases (quality $<$ 20 at the end of reads, and 4 consecutive bp with average quality $<$ 30) were trimmed using Trimmomatic v0.33 (Bolger et al. 2014). Mapping was done using BWA-MEM v0.7.15 (Li 2013). Mapped reads were sorted with samtools v1.3.1 (Li et al. 2009) and read duplicates removed using Picard Markduplicates (Picard toolkit 2019). Realignment around INDELs was performed using GATK v3.2-2 (Van der Auwera et al. 2013). Genotype calling was performed for each species independently using bcftools v.1.5 (Li 2011), with minimum site (QUAL) and RMS mapping (MQ) qualities of 20, coverage (FMT/DP) between 6X and twice the average genomic coverage, and minimum genotype quality (FMT/GQ) of 20. Indels and flanking 10 bp were coded as missing data. Only sites covered in the two analysed individuals were retained. Heterozygosity was calculated in sliding windows of 50 kb, using the popgenWindows.py script available at https://github.com/simonhmartin/genomics_general, and 500 windows were randomly sampled per references and species.

The Pairwise Sequentially Markovian Coalescent (PSMC) model (Li and Durbin 2011) was then used to compare single-genome demographic inferences of the mountain hare using alternative genome references (*L. timidus* assembled genome, prior to re-scaffolding, was also included, to control for potential biases arising from the reference-based re-scaffolding process), and to infer the demographic profiles of *L. timidus* and *L. granatensis* using the *L. timidus* re-scaffolded reference (as in Seixas et al. 2018, who used the hare pseudo-reference). Changes in the density of called variants among references should cause important differences in the inferred profiles. Diploid consensus sequences were built using samtools v1.3.1 *mpileup* and *call* modules, and only sites with minimum base and mapping qualities of 20, and coverage between 8X and twice the average depth were considered (atomic time intervals were set to $4 + 50*2 + 2 + 4$ as in Seixas et al. 2018). Results were scaled using a generation time of 2 years (Marboutin and Peroux 1995) and a mutation rate (μ) of 2.8×10^{-9} substitutions/site/generation (Seixas et al. 2018). The variance of effective population size (N_e) estimates was assessed by 50 bootstraps in randomly sampled segments with replacement.

Results and Discussion

De Novo Reference Genome Assembly and Annotation

Genome assembly and sequencing metrics are in table 1 and supplementary table S1, Supplementary Material online. The assembly length, 2.70 Gb, was consistent with the *k*-mer estimate (~2.75 Gb) and the assembled length of the rabbit genome (~2.74 Gb; Carneiro et al. 2014). The *L. timidus* re-scaffolded genome contained 4,205 scaffolds, being 99.9% of the total assembly size comprised in 605 scaffolds with a minimum length of 35 kb. Of the 4,104 mammalian core genes, 3,793 (92.4%) were present in our assembly, 3,445 of which (90.8%) were found as complete single copies. The number of predicted and annotated genes (29,238 and 24,578, respectively—table 1 and supplementary fig. S2, Supplementary Material online) are in line with several published mammalian genomes that used similar sequencing approaches (Keane et al. 2015; Li et al. 2017; Koepfli et al. 2019; Ming et al. 2019), and with the extrapolation from the BUSCO completeness assessment, suggesting that the majority of genes present in our draft genome was covered by the annotation process. A total of 683 predicted genes were uniquely annotated based on the hare transcriptome and possibly represent hare-specific genes (supplementary fig. S2, Supplementary Material online).

Through the characterization of gene content (supplementary fig. S3, Supplementary Material online) and chromosome-scaffold alignment (supplementary Figs. S4 and S5, Supplementary Material online), we were able to establish correspondence between the rabbit chromosomes

Table 1
Summary Statistics of the *L. timidus* Genome Assembly, Curation and Annotation

	Value
Raw reads	3,396,221,990
Clean reads	2,084,055,186 (61%)
Raw de novo assembly (ALL-PATHS)	
Number of scaffolds	30,212
Largest (bp)	4,384,173
Total length (bp)	2,741,083,031
Contig N50 (bp)	11,800
Scaffold N50 (bp)	347,000
GC content (%)	43.54
Post assembly curation (REAPR)	
Number of scaffolds	70,707
Largest (bp)	1,921,307
Total length (bp)	2,662,248,649
N50 (bp)	156,456
GC content (%)	43.54
Scaffolding (SSPACE >1 kb)—*L. timidus* assembled genome	
Number of scaffolds	32,294
Largest (bp)	3,358,433
Total length (bp)	2,703,715,767
N50 (bp)	3,358,433
GC content (%)	43.54
Reference-based scaffolder—*L. timidus* re-scaffolded genome	
Number of scaffolds	4,205 (83% on the 22 scaffolds corresponding to the 22 rabbit chromosomes)
Largest (bp)	194,083,885
Total length (bp)	2,703,257,129
N50 (bp)	117,222,600
GC content (%)	43.40
L. timidus re-scaffolded genome characteristics and annotation statistics	
Haploid chromosome number	22 (constrained by the rabbit reference)
Estimated genome length	2.75 Gb
Sequencing coverage	77x
Total assembly length	2.70 Gb
Gaps	18.8% (supplementary fig. S6, Supplementary Material online)
Repeats	28.0%
BUSCO genes present (estimation of genome completeness)	3,793 (92%)
Ab-initio gene prediction	29,238 genes
Gene space (exons, introns, etc.)	196.6 Mb (7.3% of assembly)
Gene length (median)	5.8 kb
Gene fragmentation	137,913 exons
Exon space	26.6 Mb (1.0% of the assembly)
Exon length (median)	193 bp
Homology-based gene annotation	24,578 proteins
(details in supplementary fig. S2, Supplementary Material online)	
Uniprot-Swissprot	23,375
O. cuniculus transcriptome	18,678
L. timidus transcriptome	15,418 (683 hare transcriptome specific)
Functional annotation	
Pfam domains	28,803 (5,068 unique)
Gene ontology (GO)	14,530 (1,382 unique)
Interpro	25,395 (4,740 unique)
KEGG	748 (391 unique)
Reactome	5,066 (1,050 unique)

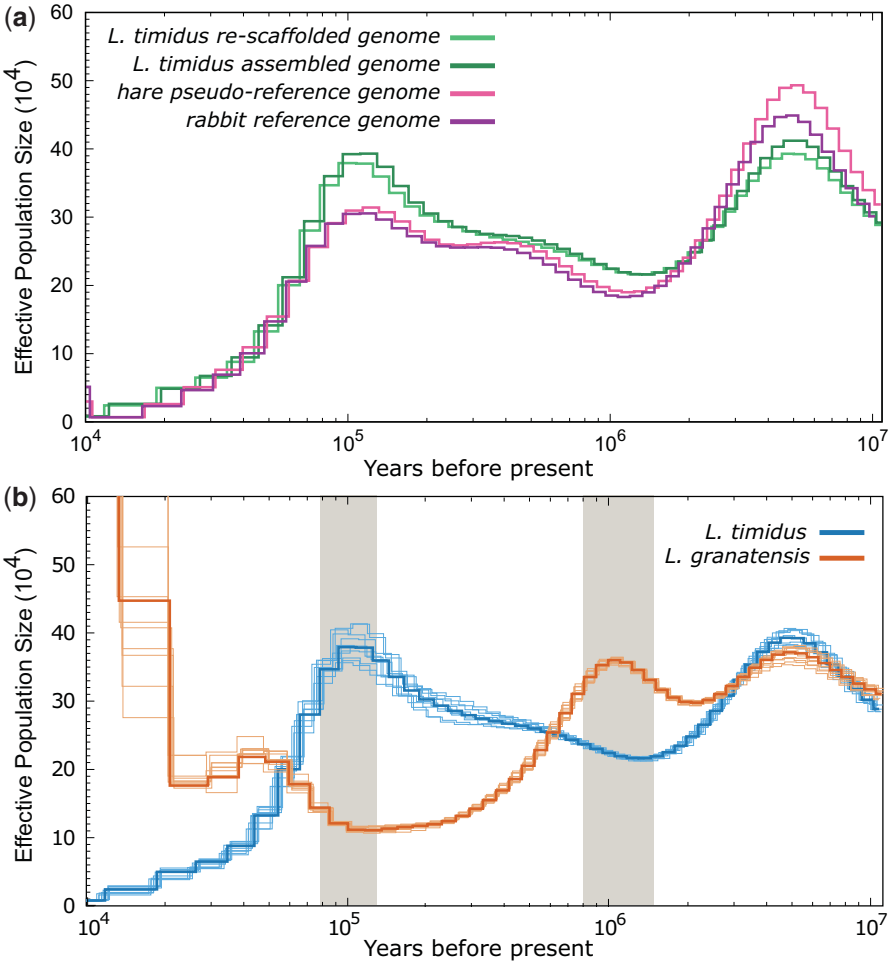

Fig 1.—PSMC inference of demographic profiles. (*a*) *Lepus timidus* demographic profile inferred using different reference genomes: *L. timidus* re-scaffolded genome, *L. timidus* assembled genome (prior to re-scaffolding), hare pseudo-reference genome, and the (European) rabbit reference genome. (*b*) PSMC inference of the demographic profiles of two European hare species—*L. timidus* and *L. granatensis*, using the *L. timidus* re-scaffolded genome (replicating the analysis by Seixas et al. 2018, which used a hare pseudo-reference). Each bold line represents a full run for each species and thin lines represent 50 randomly sampled fragments bootstrap. A rate of 2.8×10^{-9} substitutions per site per generation and a generation time of two years were assumed. Inflection points are denoted by the gray vertical bars, as in Seixas et al. (2018).

($2N = 44$) and 22 scaffolds of the re-scaffolded version of the *L. timidus* assembly. The 22 scaffolds correspond to 83% of the total length of the assembly (2.24 Gb). It should be noted that hares have 24 chromosomes, since rabbit's chromosomes 1 and 2 are each split into two in hares (presumably resulting from two fusions in the rabbit lineage; Robinson et al. 2002). This karyotype difference was naturally not recovered in our re-scaffolded assembly, which highlights the inherent shortcomings of reference-guided scaffolding. While the new genome should be accurate in resolving small-scale structural variation (small insertions–deletions, repeats, and/or inversions as recovered by the original assembled contigs/scaffolds; Salzberg et al. 2012), larger genomic rearrangements, should be missed due to the assumption of synteny with the reference.

The Impact of Alternative Reference Mapping Strategies on Genomic Analyses

The proportion of mapped reads from whole-genome re-sequencing was higher using the hare pseudo-reference, but the number of uniquely mapped reads was larger using the *L. timidus* reference genome (supplementary table S2, Supplementary Material online). These statistics suggest that the hare pseudo-reference increases both mapping proportion and efficiency, but the new hare reference allows increased confidence in mapping, measured as the proportion of uniquely mapped reads. The distributions of heterozygosity estimated in 50 kb windows along the genome did not differ significantly across analyses with different references ($P > 0.05$, Wilcoxon Ranked-sum test; supplementary fig. S7, Supplementary Material online). In agreement, the

PSMC demographic profiles also displayed similar shapes across references, with only slight differences of inferred effective population sizes (fig. 1*a*). Also, the demographic profiles of both *L. timidus* and *L. granatensis* inferred using the new *L. timidus* re-scaffolded genome are similar to those inferred by Seixas et al. (2018) using the hare pseudo-reference (fig. 1*b*). These results suggest that the use of the alternative tested references does not impact heterozygosity tract patterns, and thus that approaches based on hare pseudo-references has not limited evolutionary inference and genome scans on hares. It also shows that re-scaffolding our de novo assembly using the rabbit genome enables the use of the new hare genome reference for genomic scale scans, where the ordering along the chromosome is important. Finally, the new hare draft genome can be useful to reveal hare-specific variation, reflected for example in the putative hare-specific genes annotated here, which needs to be evaluated and investigated.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

In Ireland, a license (no. C 337/2012 dated October 31, 2012) to take animals for scientific purposes under the Wildlife Acts 1976—2012 Sections 9, 23, and 34 was issued by Laura Claffey, Wildlife Licensing Unit, National Parks & Wildlife Service (NPWS), Department of Arts, Heritage & the Gaeltacht (DAHG). Thanks to Jimi Conroy, the local NPWS Conservation Officer, who consented to sampling within his jurisdiction. The Irish Coursing Club (ICC) captured animals from the wild and we thank D.J. Histon, Chief Executive and Secretary, who supported the research. William Fitzgerald, Department of Agriculture, Food and the Marine (DAFM) dispatched animals in a humane manner and provided dissection skills. This work was funded by the project HybridAdapt, supported by Portuguese National Funds through the Fundação para a Ciência e a Tecnologia (FCT) (reference FCT-ANR/BIA-EVF/0250/2012), and by French Agence Nationale de la Recherche (ANR) (reference ANR-12-ISV7–0002-01). Additional support was obtained from Programa Operacional Potencial Humano – Quadro de Referência Estratégica Nacional (POPH-QREN) funds from the European Social Fund (ESF) and FCT (IF/00033/2014/CP1256/CT0005 FCT Investigator grant to J.M.-F., and SFRH/BD/115089/2016 and SFRH/BD/87126/2012 PhD grants to J.P.M. and F.A.S., respectively), and from a National Science Foundation (NSF) EPSCoR grant (OIA-1736249 to J.M.G.). Instrumentation, laboratory, and/or computational support were provided by CIBIO NEW-GEN sequencing platform, supported by the European Union under grant agreement no. 286431, by the Montpellier Bioinformatics Biodiversity platform supported by LabEx CeMEB, an ANR "Investissements d'avenir" program (ANR-10-LABX-04-01), and by the University of Montana Genomics Core. Additional support was obtained from the projects NORTE-01-0145-FEDER-000007 (NORTE2020 Programme, PORTUGAL 2020, through the European Regional Development Fund – ERDF), and POCI-01-0145-FEDER-022184 (COMPETE2020, PORTUGAL2020, and ERDF). It was also supported by Portugal–France bilateral cooperation programs PESSOA (FCT in Portugal; Ministère de l'Europe et des Affaires étrangères (MEAE) and Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche (MESRI) in France) and Cotutelles/Cotutelas PESSOA (FCT through Conselho de Reitores das Universidades Portuguesas (CRUP) in Portugal; MEAE and MESRI through Conférence des Présidents d'Université (CPU) in France). This research was conducted in the frame of the Laboratoire International Associé (LIA) "Biodiversity and Evolution" supported by Institut écologie et environnement – InEE (Centre national de la recherche scientifique – CNRS, France) and FCT (Portugal).

Author Contributions

J.M.-F., P.B., F.S., and P.C.A. conceived the study. N.R., W.I.M., P.C.A., and J.M.-F. organized and performed sampling. L.F. and C.M.C. performed laboratory work under supervision of J.M.-F. and J.M.G., respectively. J.P.M. and F.A.S. analysed the data. J.P.M., F.A.S., and J.M.-F. wrote the manuscript, with contributions from P.B. and J.M.G. and comments from the other authors. All authors read, revised, and approved the manuscript.

Literature Cited

Alves PC, Ferrand N, Suchentrunk F, Harris DJ. 2003. Ancient introgression of *Lepus timidus* mtDNA into *L. granatensis* and *L. europaeus* in the Iberian Peninsula. Mol Phylogenet Evol. 27(1):70–80.

Alves PC, Melo-Ferreira J, Freitas H, Boursot P. 2008. The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus *Lepus*. Philos Trans R Soc B Biol Sci. 363(1505):2831–2839.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27(4):578–579.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Bosi E, et al. 2015. MeDuSa: a multi-draft based scaffolder. Bioinformatics 31(15):2443–2451.

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ. 6:e4958.

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10(1):421.

Cantarel BL, et al. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18(1):188–196.

Caravaggi A, et al. 2017. Niche overlap of mountain hare subspecies and the vulnerability of their ranges to invasion by the European hare; the (bad) luck of the Irish. Biol Invasions 19(2):655–674.

Carneiro M, et al. 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science 345(6200):1074–1079.
Fabregat A, et al. 2018. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 46(D1):D649–D655.
Finn RD, et al. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44(D1):D279–D285.
Fontanesi L, et al. 2016. LaGomiCs—lagomorph genomics consortium: an international collaborative effort for sequencing the genomes of an entire mammalian order. J Hered. 107(4):295–308.
Giska I, et al. 2019. Introgression drives repeated evolution of winter coat color polymorphism in hares. Proc Natl Acad Sci U S A. 116(14):24150–24156.
Gnerre S, et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 108(4):1513–1518.
Hamill RM, Doyle D, Duke EJ. 2006. Spatial patterns of genetic diversity across European subspecies of the mountain hare, *Lepus timidus* L. Heredity (Edinb). 97(5):355–365.
Hunt M, et al. 2013. REAPR: a universal tool for genome assembly evaluation. Genome Biol. 14(5):R47.
Jiang H, Lei R, Ding S-W, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15(1):182.
Jones MR, et al. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. Science 360(6395): 1355–1358.
Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240.
Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44(D1):D457–D462.
Keane M, et al. 2015. Insights into the evolution of longevity from the bowhead whale genome. Cell Rep. 10(1):112–122.
Koepfli K-P, et al. 2019. Whole genome sequencing and re-sequencing of the sable antelope (*Hippotragus niger*): a resource for monitoring diversity in ex situ and in situ populations. G3 9(6):1785–1793.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357–359.
Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27(21):2987–2993.
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv 00: 3.
Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. Nature 475(7357):493–496.
Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.
Li Z, et al. 2017. Draft genome of the reindeer (*Rangifer tarandus*). Gigascience 6(12):1–5.
Marboutin E, Peroux R. 1995. Survival pattern of European hare in a decreasing population. J Appl Ecol. 32(4):809–816.
Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. Bioinformatics 27(6):764–770.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. Embnet J. 17(1):10.
Matthee CA, Van Vuuren BJ, Bell D, Robinson TJ. 2004. A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the miocene. System Biol. 53(3):433–447.
Melo-Ferreira J, Alves PC, Freitas H, Ferrand N, Boursot P. 2009. The genomic legacy from the extinct *Lepus timidus* to the three hare species of Iberia: contrast between mtDNA, sex chromosomes and autosomes. Mol Ecol. 18(12):2643–2658.
Melo-Ferreira J, et al. 2007. The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. Mol Ecol. 16(3):605–618.
Mi H, et al. 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45(D1):D183–D189.
Mikheenko A, Saveliev V, Gurevich A. 2016. MetaQUAST: evaluation of metagenome assemblies. Bioinformatics 32(7):1088–1090.
Ming Y, Jian J, Yu X, Wang J, Liu W. 2019. The genome resources for conservation of Indo-Pacific humpback dolphin, *Sousa chinensis*. Sci Data 6(1):68.
Picard toolkit. 2019. Broad Institute, available at http://broadinstitute.github.io/picard/.
Robinson TJ, Yang F, Harrison WR. 2002. Chromosome painting refines the history of genome evolution in hares and rabbits (order Lagomorpha). Cytogenet. Genome. Res. 96(1–4):223–227.
Salzberg SL, et al. 2012. GAGE: a critical evaluation of genome assemblies and assembly algorithms. Genome Res. 22(3):557–567.
Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27(6):863–864.
Seixas FA, Boursot P, Melo-Ferreira J. 2018. The genomic impact of historical hybridization with massive mitochondrial DNA introgression. Genome Biol. 19(1):91.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.
Smit A, Hubley R. 2008. RepeatModeler Open-1.0, available at http://www.repeatmasker.org.
Smit A, Hubley R, Green P. 2013. RepeatMasker Open-4.0, available at http://www.repeatmasker.org.
Smith, A.T., Johnston, C.H., Alves, P.C., Hackländer, K., editors. 2018. Lagomorphs: pikas, rabbits, and hares of the world. Baltimore (MD): Johns Hopkins University Press.
Thulin CG, Jaarola M, Tegelström H. 1997. The occurrence of mountain hare mitochondrial DNA in wild brown hares. Mol Ecol. 6(5): 463–467.
Van der Auwera GA, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11.10.1–33.
Waltari E, Cook JA. 2005. Hares on ice: phylogeography and historical demographics of *Lepus arcticus*, *L. othus*, and *L. timidus* (Mammalia: Lagomorpha). Mol Ecol. 14(10):3005–3016.

Associate Editor: Judith Mank