The genomic impact of historical hybridization with massive mitochondrial DNA introgression

Fernando Seixas, Pierre Boursot, José Melo-Ferreira

Genome Biology

CrossMark

The genomic impact of historical hybridization with massive mitochondrial DNA introgression

Fernando A. Seixas[1,2,3], Pierre Boursot[3*] and José Melo-Ferreira[1,2*]

Abstract

Background: The extent to which selection determines interspecific patterns of genetic exchange enlightens the role of adaptation in evolution and speciation. Often reported extensive interspecific introgression could be selection-driven, but also result from demographic processes, especially in cases of invasive species replacements, which can promote introgression at their invasion front. Because invasion and selective sweeps similarly mold variation, population genetics evidence for selection can only be gathered in an explicit demographic framework. The Iberian hare, *Lepus granatensis*, displays in its northern range extensive mitochondrial DNA introgression from *L. timidus*, an arctic/boreal species that it replaced locally after the last glacial maximum. We use whole-genome sequencing to infer geographic and genomic patterns of nuclear introgression and fit a neutral model of species replacement with hybridization, allowing us to evaluate how selection influenced introgression genome-wide, including for mtDNA.

Results: Although the average nuclear and mtDNA introgression patterns contrast strongly, they fit a single demographic model of post-glacial invasive replacement of *timidus* by *granatensis*. Outliers of elevated introgression include several genes related to immunity, spermatogenesis, and mitochondrial metabolism. Introgression is reduced on the X chromosome and in low recombining regions.

Conclusions: General nuclear and mtDNA patterns of introgression can be explained by purely demographic processes. Hybrid incompatibilities and interplay between selection and recombination locally modulate levels of nuclear introgression. Selection promoted introgression of some genes involved in conflicts, either interspecific (parasites) or possibly cytonuclear. In the latter case, nuclear introgression could mitigate the potential negative effects of alien mtDNA on mitochondrial metabolism and male-specific traits.

Keywords: Introgressive hybridization, Iberian hares, Whole genome sequencing, Range invasion, Adaptation

Background

Genetic introgression between closely related species can be a major source of adaptive variation, in addition to standing variation and new mutation [1, 2]. Introgression of pre-tested genetic combinations may provide important advantages to prosper or invade some habitats [3, 4], although it could also be non-adaptive if involving selfish genetic elements or compensatory mechanisms [5–7]. An increasing number of studies report the role of adaptive introgression in species evolution and interactions [8–15]. There is also growing evidence for the role of introgression in promoting adaptive speciation and radiation, including in conditions of apparent sympatry [16, 17]. The most indisputable cases of a role for introgression in adaptation concern genes whose function can clearly be related to a known or presumed adaptation in the recipient species.

Although gathering systematic and genome-wide empirical and statistical evidence for introgression promoted by selection is now at hand with the development of genomics, demonstrating selection-driven

* Correspondence: pierre.boursot@umontpellier.fr; jmeloferreira@cibio.up.pt
[3]Institut des Sciences de l'Évolution, Université de Montpellier, CNRS, IRD, EPHE, Place Eugène Bataillon, 34095 Montpellier, France
[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal
Full list of author information is available at the end of the article

Seixas et al. Genome Biology (2018) 19:91

Page 2 of 20

introgression is challenging for at least two reasons. First, one must be able to disentangle the effects of introgression from those of incomplete lineage sorting (i.e., sharing of ancestral variation among daughter populations/species), which is expected to be pervasive between recently diverged taxa. Second, interpreting a pattern of introgression as driven by selection based on its geographic and frequency patterns needs a comparison with a null, neutral expectation that depends on the complex and generally unknown historic, geographic, and demographic conditions of genetic admixture. For example, during invasion of the range of a species by another with hybridization, drift in initially small founding populations and repeated hybridization at the invasion front may bring variants introgressed from the resident species into the invading one to high frequencies well beyond the initial contact between the two interacting taxa [18, 19]. Such high prevalence and geographic extent may thus not suffice to invoke selection as driving introgression.

The vast majority of the reported cases of introgression in animals involve the mitochondrial genome (mtDNA) [20], often occurring at high frequencies over extended regions [21–23]. Explanations for the apparent tendency of mtDNA to extensively cross species boundaries include pure demography/drift, sex-biased interspecific mating, and very often adaptation (reviewed by Toews and Brelsford [20]). However, the occurrence and persistence of introgression during range replacements are favored by high drift at the invasion front and low intraspecific migration rates, preventing the dilution of introgression at the front by subsequent migration from the non-affected source of the geographic expansion. These two parameters can vary across genomic regions with different modes of sex-linked transmission if the two sexes have different migration rates. In species where females are more philopatric than males, the female-transmitted mitochondrial genome is expected to be the most affected by massive introgression [18, 19, 24]. Since the mitochondrial genome is non-recombining, it represents a single realization of the demographic processes at play, and the patterns of sequence variation resulting from invasion-driven introgression are expected to resemble those predicted following a selective sweep. Sequence variation of mtDNA alone is therefore unable to provide unequivocal evidence of selection-driven introgression. In contrast, the recombining nuclear genome provides numerous independent realizations of the processes at play; it should be generally affected by demographic processes alone and only locally by selective processes, thus allowing adjustment of a neutral demographic model that can then be applied to test mtDNA patterns. Furthermore, such model could also be used to detect nuclear outliers, candidate for selection-driven introgression. The discovery that these outliers are potentially involved in functional interactions with the mitochondrial genome would provide strong

evidence for co-adaptation between the nuclear and mitochondrial genomes [7]. These co-introgressions could, however, also result from compensatory introgression of nuclear genes, mitigating the deleterious effects of demography-driven alien mtDNA invasion (the so-called mother's curse [25]). In both cases, this would result from independent cytonuclear coevolution in the two taxa, either adaptive or in response to genetic conflicts resulting from the different sex-linked transmission modes of the two genomes.

In this work, we explicitly test the influence of range replacements in determining patterns of introgression in a natural system with geographically confined but extensive mtDNA introgression, providing the opportunity to assess the relative contributions of demographic and selective processes to genetic admixture. The three species of hares (genus *Lepus*) thriving in the north of the Iberian Peninsula (*Lepus castroviejoi*, *Lepus europaeus*, and *Lepus granatensis*) are strongly affected by mitochondrial DNA introgression from *Lepus timidus*, an arctic-boreal species now extinct in Iberia but present in the fossil record until the last glacial maximum [26]. The Iberian species may have replaced *L. timidus* in this region after the last glacial maximum, under conditions that promote introgression during invasive replacement [27]. Several aspects of mtDNA variation in *L. granatensis* appear compatible with such a scenario. These include a south–north gradient of increasing mitochondrial introgression frequency [21, 28], from absent in the southern half of the peninsula to almost fixed in some northernmost populations, and an east–west phylogeographic structure of mtDNA of *timidus* origin [29]. However, the prevalence of this mitochondrial genome of arctic/boreal origin in three species (fixed in *L. castroviejoi* and quasi-fixed in *L. europaeus* and in some northern populations of *L. granatensis*) and its restriction to Northern Iberia could suggest that it confers some adaptive advantage corresponding to environmental conditions in this region [29, 30]. Possible signs of competitive replacement of the native mtDNA genome by the alien one (which would be compatible with adaptive introgression) were also proposed [27, 30]. Studies of a small number of nuclear markers in *L. granatensis* suggested evidence of south–north range expansion [31], low frequency introgression from *L. timidus*, but all over the distribution area, contrarily to mtDNA [28], and geographically widespread high frequency introgression of an X chromosome fragment [29]. These preliminary results draw a contrasted and incomplete picture, leaving open the question of the relative importance of demographic and selective factors in determining introgression into *L. granatensis*, including for mtDNA.

Here, we use whole genome sequences from the two species to infer the genomic and geographic patterns of

Seixas et al. Genome Biology (2018) 19:91

Page 3 of 20

nuclear introgression from *L. timidus* into *L. granatensis*. We then simulate expectations of introgression prevalence under a geographically explicit model of species replacement and assess whether this model can reconcile the contrasted nuclear and mitochondrial DNA introgression patterns. This null model was then used to identify regions of the genome with outlying high frequencies of introgression, which could therefore be driven by selection.
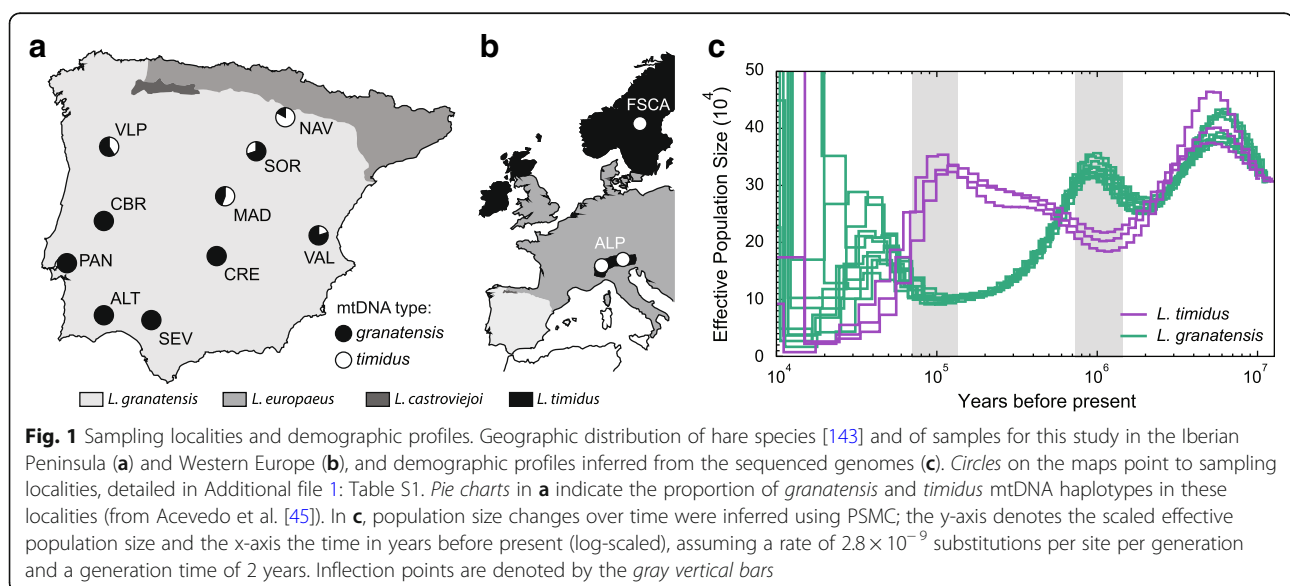
Results

Sampling and genomic datasets

We sequenced the genomes of ten *L. granatensis* specimens sampled over the species distribution range in Iberia, five in the southern region not affected by mitochondrial introgression and five along the gradient of mitochondrial introgression in the northern half of the Peninsula (Fig. 1a). Three *L. timidus* genomes, two from the Alps and one from Fennoscandia, were also sequenced (Fig. 1b), and one *L. americanus* genome was used as outgroup for some analyses. All sequenced specimens were females.

Using an iterative mapping approach [32], we built a hare pseudo-reference genome using the rabbit genome as template. This procedure increased average read mapping proportions from 92.3 to 93.6%. The median sequencing depth was 25.9X, with a range between 22.8X and 37.4X per genome (see Additional file 1: Table S1 for sequencing statistics and sampling details). Broad synteny between the rabbit and hare karyotypes is expected but some known fusions/fissions exist [33] and were taken into account in our analyses. The final dataset consisted of 46,583,958 single nucleotide polymorphisms (SNPs).

Inference and broad impact of genome-wide introgression

Estimated mean uncorrected distance between *L. granatensis* and *L. timidus* was 0.69%, and mean genome-wide F_{ST} per site was 0.35. As expected given estimates of effective population size [34], *L. timidus* was found to be more polymorphic than *L. granatensis* ($\pi = 0.0022$ and 0.0014, respectively).

We inferred regions of the ten sequenced *L. granatensis* genomes that were affected by introgression from *L. timidus*. Methods aimed at detecting local ancestry in admixed populations generally rely on the observation of presumably pure parental populations [35–37]. However, previous analyses of *L. granatensis*, although based on a limited number of markers, had suggested that nuclear introgression from *L. timidus* was present all over the range of *L. granatensis* [28], so that none of the samples sequenced here could be considered a pure *L. granatensis* reference. We therefore used the ancestry inference method implemented in ELAI (Efficient Local Ancestry Inference [38]), which can accommodate such a situation. The method partitions linkage disequilibrium into two layers corresponding to intra- and interspecific disequilibrium. It is not based on an arbitrary segmentation of the genome and is able to infer the boundaries of the introgression tracts in the genome. When one of the parental populations is unobserved, the method is expected to perform properly if the admixed population has a high proportion of ancestry from this unobserved origin, which previous results suggested for *L. granatensis* [28]. We tested the power of the method by artificially introgressing fragments of different lengths from *L. timidus* into the *L. granatensis* genome and found that the power of the method is very high for large

Fig. 1 Sampling localities and demographic profiles. Geographic distribution of hare species [143] and of samples for this study in the Iberian Peninsula (**a**) and Western Europe (**b**), and demographic profiles inferred from the sequenced genomes (**c**). *Circles* on the maps point to sampling localities, detailed in Additional file 1: Table S1. Pie charts in **a** indicate the proportion of *granatensis* and *timidus* mtDNA haplotypes in these localities (from Acevedo et al. [45]). In **c**, population size changes over time were inferred using PSMC; the y-axis denotes the scaled effective population size and the x-axis the time in years before present (log-scaled), assuming a rate of 2.8×10^{-9} substitutions per site per generation and a generation time of 2 years. Inflection points are denoted by the *gray vertical bars*

Seixas *et al. Genome Biology* (2018) 19:91

Page 4 of 20

introgression tracts (50 kb; 91.2%), high for intermediate tracts (30 kb, 71.4%), and low for small fragments (10 kb, 18.8%) (Additional file 1: Table S2). According to ELAI-based estimates, the proportion of the genome affected by introgression varied between 1.38 and 2.44% among *L. granatensis* specimens (Table 1), which may thus represent an underestimation given our power analyses and mean inferred introgression tract sizes (~ 29 kb).

Historic and geographic context of introgressive hybridization events

The PSMC [39] profiles of the *L. granatensis* and *L. timidus* individual genomes suggest at least two episodes of population size fluctuation in both species after their divergence (occurring when the two curves merge in the past; Fig. 1c). Remarkably, population sizes of the two species appear to vary in phase but in opposite directions, periods of expansion for one species corresponding to periods of retraction for the other. The method was, however, unable to reliably infer demography in the recent past, since the last glacial maximum.

The partitioning of *L. granatensis* diversity assessed with a principal component analysis (PCA) from a subset of independent SNPs and including *L. timidus* revealed differentiation on the first two axes (Fig. 2a) that, in both cases, correlated with distance to the southernmost sample (Spearman's rank correlation test *p* value < 0.05; Fig. 2b), located at the inferred origin of expansion of the species in southwest Iberia [31]. Since the first axis discriminates the two species, the spread of *L. granatensis* along this axis likely corresponds to a gradient of introgression by *L. timidus*. Accordingly, when the analysis is performed with *L. americanus* instead of *L. timidus*, the significant intraspecific geographic gradient of differentiation along the

species-discriminating axis is lost (Fig. 2c, d). The geographic differentiation along the second axis appears independent of introgression, as it remains significant whether polarizing the PCA with *L. timidus* or *L. americanus*. The similar geographic patterns along the two axes in the PCA with *L. timidus* (Fig. 2a, b) are striking and must result from the same demographic process. This is likely the range expansion of *L. granatensis* from southwest Iberia previously inferred [31], based on a much more limited number of markers (100 SNPs) but a much larger species-wide sample.
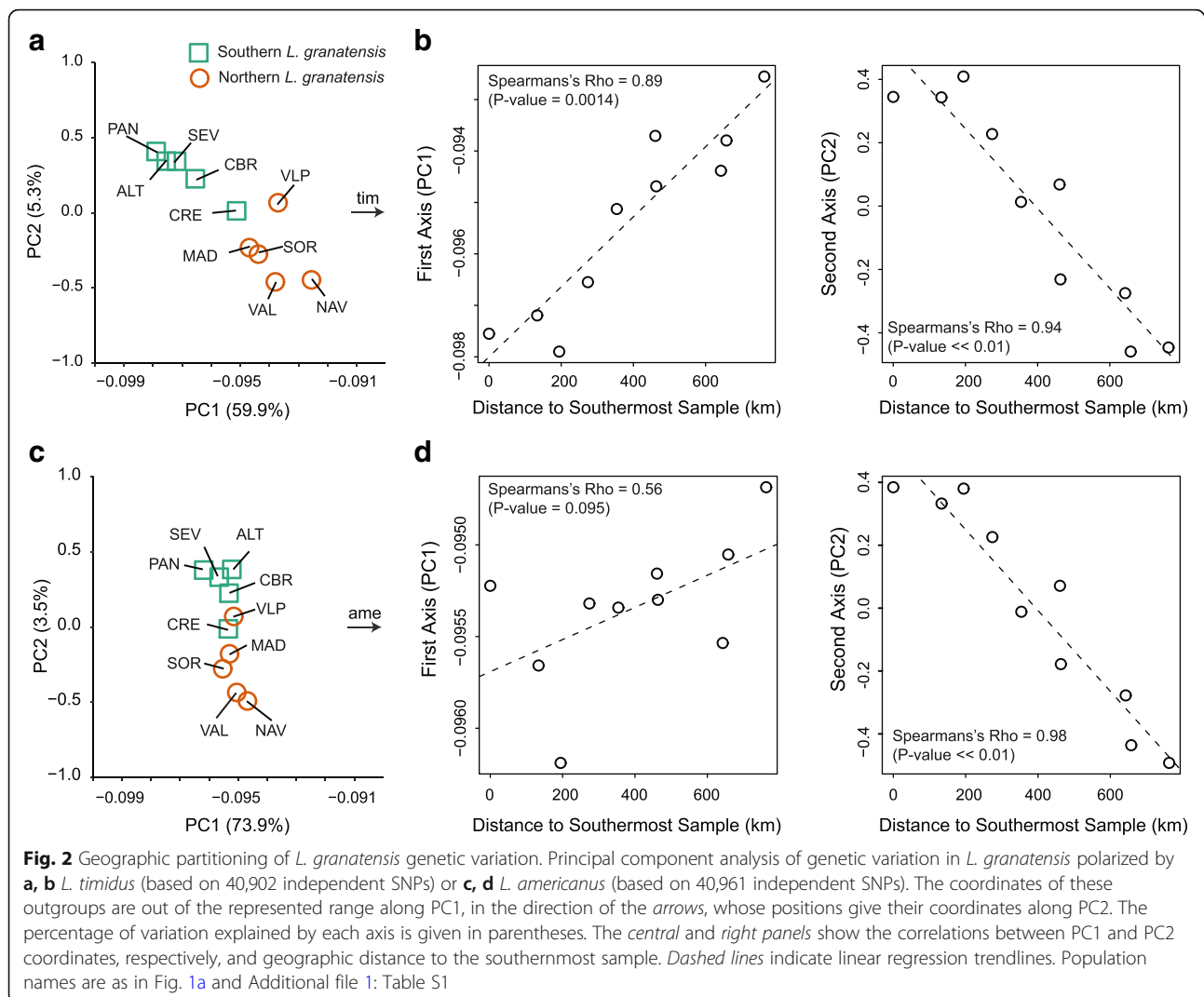
We also inferred that genomic proportions of introgression per individual significantly increase towards the north, with distance to the origin of the range expansion (Spearman's rank correlation test *p* value = 0.00086; Fig. 3a). Introgression tract lengths are expected to decay since the initial hybridization, due to recombination with native tracts, and we used their distribution (Fig. 3c) to estimate the age of hybridization [40]. However, this method has limitations, including a potential bias in the empirical size distribution due to the crypticity of small tracts and the possibly unrealistic underlying model of instantaneous admixture [41, 42]. We therefore used a second method, based on the size distribution of DNA tracts identical by state (IBS) within and between species, and tested models with multiple introgression pulses [43]. The resulting estimates suggest that introgression likely occurred between the last glacial maximum (24.3 thousand years ago (kya), based on IBS tracts; Additional file 1: Table S3), and early Holocene (7 kya, considering the distribution of introgressed tract lengths; Fig. 3c). Remarkably, mean inferred introgression tract lengths significantly increase towards the

Table 1 Mean population introgression proportions based on empirical inference and simulated datasets (using SPLATCHE2)

Parameters					Introgression proportions (%)											Max.[a]	Sign.[b]
Set	K_G	K_T	G	M	A	Mean	ALT	SEV	PAN	CBR	CRE	VLP	MAD	VAL	SOR	NAV		
Empirical																		
ELAI	–	–	–	–	–	2.0	1.3	1.6	1.5	1.9	1.9	2.4	2.2	2.2	2.2	2.4	–	–
Simulated																		
par1	1000	500	0.5	0.2	0.005	3.9	4.0	4.0	3.9	3.9	3.9	4.0	4.0	3.9	4.0	3.9	70	35
par2	1000	500	0.5	0.02	0.005	8.5	7.0	6.7	7.6	8.6	7.6	10.3	8.8	8.2	9.8	10.1	80	45
par3	1000	500	0.5	0.2	0.03	22.9	22.9	22.9	22.9	22.9	22.9	22.9	22.9	22.9	22.9	22.9	95	70
par4	1000	500	0.5	0.02	0.03	34.9	28.9	28.1	31.1	34.9	31.7	40.9	36.2	34.5	40.7	42.4	95	85
par5	10,000	5000	0.5	0.2	0.005	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	5.4	50	30
par6	10,000	5000	0.5	0.02	0.005	11.3	9.5	9.0	10.2	11.5	10.2	13.5	11.7	10.8	12.9	13.2	65	40
par7	10,000	5000	0.5	0.2	0.03	25.2	25.3	25.2	25.2	25.2	25.2	25.2	25.2	25.3	25.2	25.2	80	60
par8	10,000	5000	0.5	0.02	0.03	37.4	31.2	30.5	33.4	37.3	34.1	43.2	38.7	37.1	43.2	45.0	95	75

K_G *L. granatensis* deme carrying capacity, K_T *L. timidus* deme carrying capacity, *G* intrinsic growth rate (same for *L. timidus* and *L. granatensis*), *M* migration rates between adjacent demes (same for *L. timidus* and *L. granatensis*), *A* bidirectional admixture. Population names are as in Fig. 1a and Additional file 1: Table S1.
[a]Maximum introgression frequency (across the 10 individuals) in percentage. [b]Introgression frequency (in percentage) above which empirical introgression frequencies are significantly higher than expected according to simulations

Seixas *et al. Genome Biology* (2018) 19:91

Page 5 of 20

Fig. 2 Geographic partitioning of *L. granatensis* genetic variation. Principal component analysis of genetic variation in *L. granatensis* polarized by **a, b** *L. timidus* (based on 40,902 independent SNPs) or **c, d** *L. americanus* (based on 40,961 independent SNPs). The coordinates of these outgroups are out of the represented range along PC1, in the direction of the *arrows*, whose positions give their coordinates along PC2. The percentage of variation explained by each axis is given in parentheses. The *central* and *right panels* show the correlations between PC1 and PC2 coordinates, respectively, and geographic distance to the southernmost sample. *Dashed lines* indicate linear regression trendlines. Population names are as in Fig. 1a and Additional file 1: Table S1

north, with the distance to the presumed origin of expansion (Spearman's rank correlation test *p* value = 0.0027; Fig. 3b), suggesting that introgression is more recent in the north.

Simulations of introgression during a range replacement

Patterns of genetic variation in *L. granatensis*, higher impact of introgression towards the north (found here for the nuclear genome and previously for mtDNA), and the northward increase in introgression tract lengths are compatible with introgression occurring during a northward range expansion of the species into the historical range of *L. timidus* in northern Iberia. However, while mtDNA introgression is strongly structured, being absent in southern Iberia and reaching high frequencies in the north [29], nuclear DNA introgression is generally rare (Fig. 4b) and present all over the species range (Fig. 3a). In order to appraise whether these apparently discordant patterns could be generated by a single underlying demographic

model, we simulated this process using SPLATCHE2 [44]. *L. granatensis* was simulated to expand from south-western Iberia 20 kya [31], and to replace *L. timidus* where it was present in northern Iberia at the last glacial maximum, as inferred from ecological niche modeling [45] (Fig. 4a). We simulated the demographic process over the species range and then the coalescent process to determine the proportions of ancestry among 50,000 independent genomic regions from each of ten individuals from the same geographic locations as the ten real samples. We varied carrying capacity, intraspecific migration, and interspecific admixture rates and inferred the resulting proportions of introgression in the ten simulated genomes. Introgression proportions in the invading species are expected to increase with higher carrying capacities, lower intraspecific migration, and higher admixture rates [18]. In keeping, we found that low levels of introgression, with strong predominance of markers with low introgression frequencies across the sampled specimens comparable to the empirical estimates,

Seixas *et al. Genome Biology* (2018) 19:91
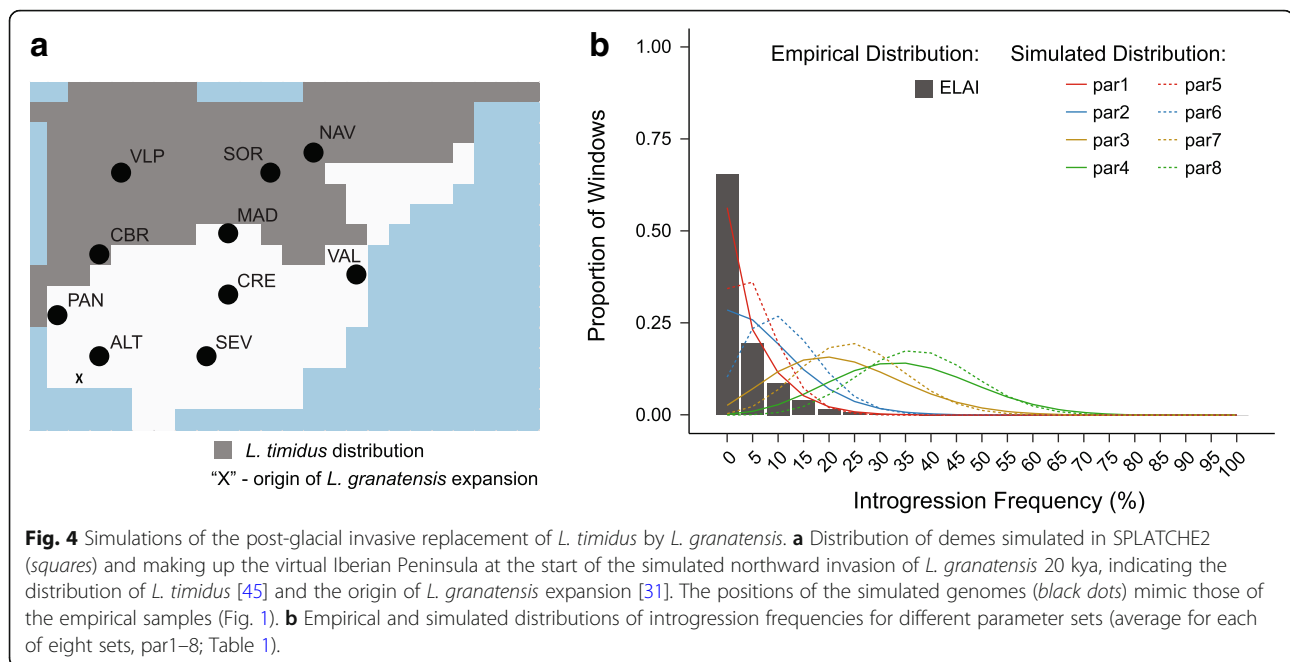
Page 6 of 20
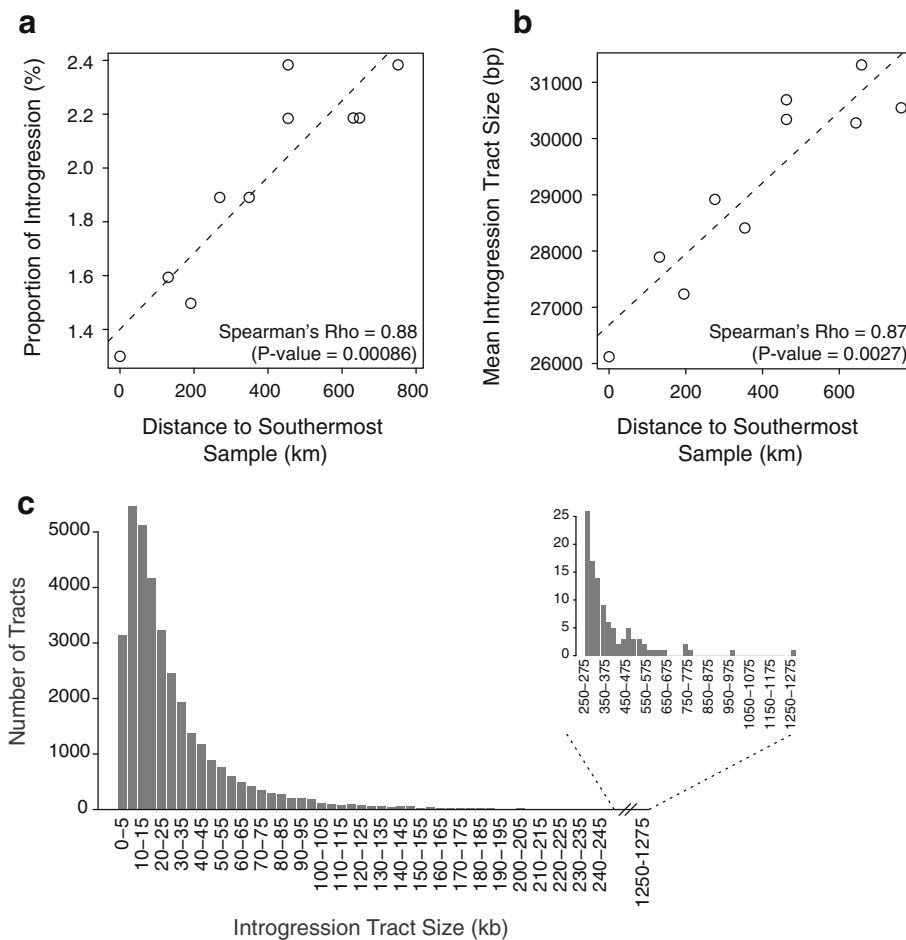
Fig. 3 Geographic variation of the inferred introgression from *L. timidus* to *L. granatensis*. Variation of the overall level of introgression (**a**) and mean introgression tract size (**b**) among the ten *L. granatensis* samples, according to their geographical distance to the southernmost sample, inferred using ELAI; *dashed lines* indicate linear regression trendlines. **c** Distribution of introgression tract sizes (in 5-kb bins) across all individuals; mean tract size is 29,364 bp

were retrieved with lower rates of admixture (Table 1; Fig. 4b). Conversely, extremely high average proportions of admixture were recovered with higher admixture rates, with important shifts towards a predominance of markers with intermediate frequencies of introgression across the ten sampled individuals (Table 1; Fig. 4b). Lower intraspecific migration rates accounted for northward gradients of introgression prevalence, similar to the empirical inferences both overall and considering separately the southern or northern samples (Table 1; Fig. 5a).

In order to test whether the empirical geographic patterns of mtDNA introgression could be recovered under the same demographic model, we repeated the simulations using the combination of parameter values that recovered geographical gradients of nuclear introgression with the lowest overall proportion of introgression (par2; Table 1). However, carrying capacity was adjusted to the effective population size of mtDNA (1/4 of the nuclear genome). Steep northward clines of increasing mtDNA

introgression were obtained when decreasing inter-deme migration to a minimum (mimicking female philopatry) and setting predominant gene flow from *L. timidus* to *L. granatensis* (a consequence of predominant male-mediated dispersal, implying that colonizers are predominantly males) (Fig. 5b). In 30.4% of the simulations, we found a significant and positive correlation between simulated and empirical frequencies of mtDNA introgression per population (Spearman's rank correlation test *p* value < 0.05). Furthermore, for each mtDNA simulation, we recorded the difference in introgression frequency between northern and southern samples and found that the empirical measure (55.4%) lies within the simulated distribution (Additional file 2: Figure S1). These results suggest that a single demographic history of northern range expansion with hybridization can reconcile contrasted patterns of nuclear and cytoplasmic introgressions, after accounting for the reduced effective population size of mtDNA, and female philopatry/male-biased migration.
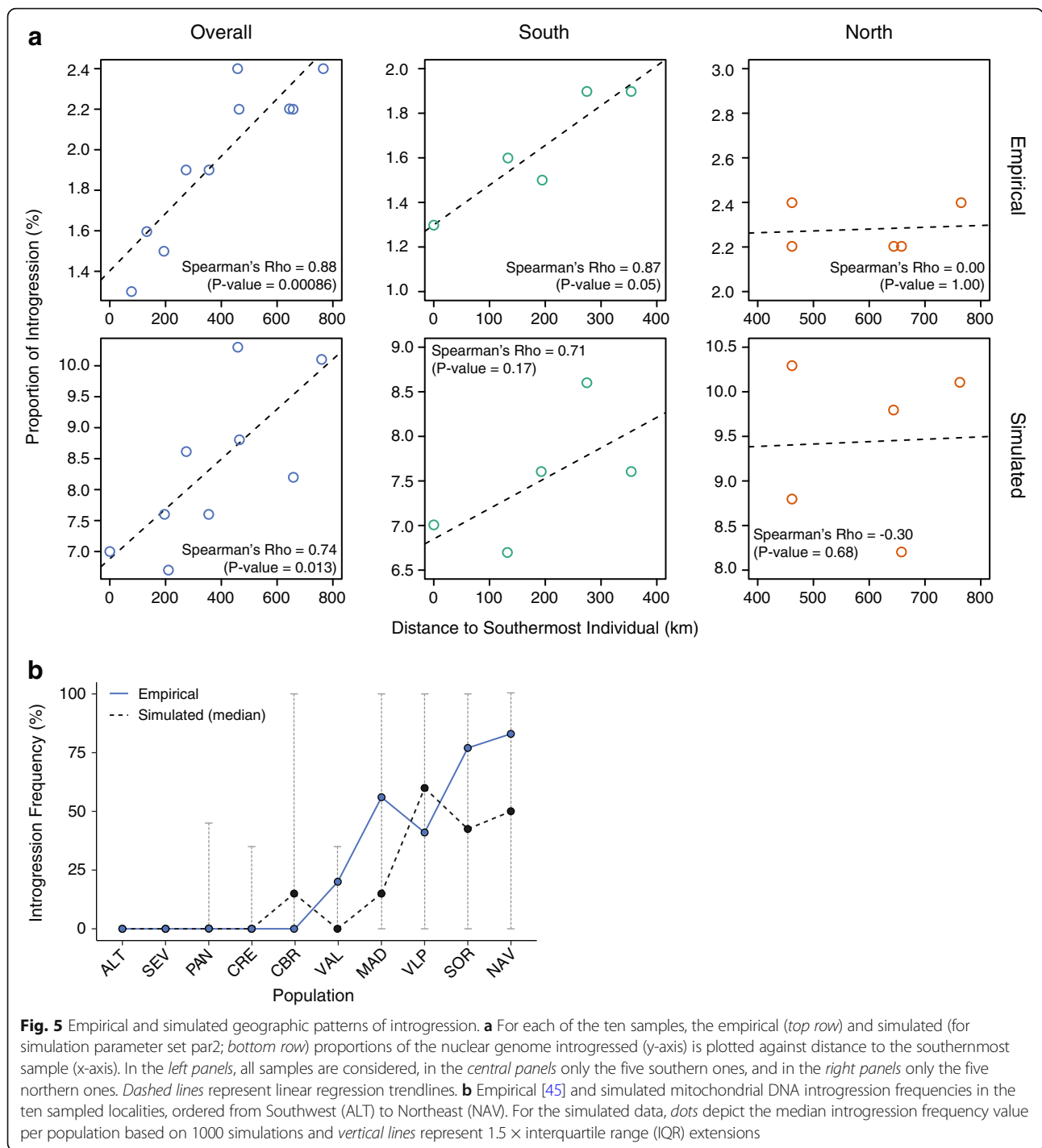
Seixas *et al. Genome Biology* (2018) 19:91

Page 7 of 20

Fig. 4 Simulations of the post-glacial invasive replacement of *L. timidus* by *L. granatensis*. **a** Distribution of demes simulated in SPLATCHE2 (*squares*) and making up the virtual Iberian Peninsula at the start of the simulated northward invasion of *L. granatensis* 20 kya, indicating the distribution of *L. timidus* [45] and the origin of *L. granatensis* expansion [31]. The positions of the simulated genomes (*black dots*) mimic those of the empirical samples (Fig. 1). **b** Empirical and simulated distributions of introgression frequencies for different parameter sets (average for each of eight sets, par1–8; Table 1).

Outlier high-frequency introgression

We were interested in detecting nuclear regions that introgressed at high frequencies, since they could have been driven by selection, eventually in relation to mtDNA introgression. Most introgressions detected by ELAI occur at low frequencies, with a majority found only in one of the 20 haploid genomes sampled (Fig. 4b). However, because ELAI was implemented to infer native *L. granatensis* variation from the admixed population, regions with high frequency introgression are expected to remain undetected by the method. In order to identify genome segments with extensive nuclear DNA introgression, we therefore used RND (Relative Node Depth [46]), which does not have this limitation. We estimated the sequence divergence (Dxy) in sliding windows along the nuclear genome between all pairs of statistically phased haplotypes containing one from the focal species (*L. granatensis*) and the other from the donor (*L. timidus*), standardized by the average divergence to the outgroup (*L. americanus*), in order to control for mutation rate variations. We then recorded the minimum of such values in each window (RNDmin [47]). Regions of introgression are expected to produce exceptionally low RNDmin values, independently of the introgression frequency [47]. Using the inferences from ELAI, we were able to verify that phasing appeared correct in regions of introgression, where linkage disequilibrium is enhanced, and allowed recovering in-phase parental haplotypes (not shown). We then used the ELAI results to predict the power and false discovery rate (FDR) of the RNDmin approach, focusing only on the range of relatively low introgression frequencies, in which ELAI is expected to

have maximum efficiency. Using an RNDmin threshold predicting an FDR of 10% resulted in a low estimated power of RND for detecting introgression (16.9, 25.7, and 42.6% for 10, 20, and 50 kb RND windows, respectively; Additional file 2: Figure S2). The distribution of RND-inferred introgression frequencies across the ten *L. granatensis* genomes was more skewed towards low frequencies than with ELAI inferences (Additional file 2: Figure S3a); however, the bulk of introgressed fragments at very high frequencies were recovered (Additional file 2: Figure S3b).

We then questioned whether such a high frequency of introgression of a few markers could be generated by the demographic range replacement process. Simulations with low levels of admixture never recovered a single marker introgressed at frequencies higher than 80% (par1–2, 5–6; Table 1). Focusing on parameter combinations that maximize the probability of introgression (par3–4, 7–8; Table 1), we conservatively identified 80% as the frequency threshold above which the empirical proportion of markers inferred as introgressed is always higher than in 95% of the simulated replicates per parameter set (Table 1). We found 139 genomic regions with outlier empirical introgression frequencies (i.e., > 80%), which contained 123 genes (Additional file 1: Table S4).

We then inspected the functions of these genes highly introgressed from *L. timidus* into *L. granatensis*. A Gene Ontology (GO) analysis revealed enrichment in several biological processes, including positive regulation of leukocyte-mediated immunity, macroautophagy, and spermatogenesis (Additional file 1: Tables S5 and S6).

Seixas *et al. Genome Biology* (2018) 19:91

Page 8 of 20

Fig. 5 Empirical and simulated geographic patterns of introgression. **a** For each of the ten samples, the empirical (*top row*) and simulated (for simulation parameter set par2; *bottom row*) proportions of the nuclear genome introgressed (y-axis) is plotted against distance to the southernmost sample (x-axis). In the *left panels*, all samples are considered, in the *central panels* only the five southern ones, and in the *right panels* only the five northern ones. *Dashed lines* represent linear regression trendlines. **b** Empirical [45] and simulated mitochondrial DNA introgression frequencies in the ten sampled localities, ordered from Southwest (ALT) to Northeast (NAV). For the simulated data, *dots* depict the median introgression frequency value per population based on 1000 simulations and *vertical lines* represent 1.5 × interquartile range (IQR) extensions

Two genes showed dN/dS ratios above 1 in the divergence to *L. americanus* ("E230025N22Rik" and HERC6), and thus potentially evolved under positive selection in hares. We found 309 non-synonymous variants between *L. americanus* and *L. timidus* in 58 of these genes, among which 30 were predicted to potentially affect protein function (Additional file 1: Table S7), according to SIFT [48]. These included two of the spermatogenesis genes (ALMS1 and NEK1) and two immune-related genes (OPTN and MSH6), which were part of the enriched GO terms.

Introgression of nuclear genes with mitochondrial functions

We investigated in more detail patterns of introgression for genes with known or potential mitochondrial functions

(hereafter "mitonuc"). Such genes with high frequencies of introgression, paralleling that for mtDNA, would be of particular interest, so we used here the results of the RND test, more amenable to detect high frequency introgression. Of the 1211 mitonuc genes reported in databases [49, 50], 1178 were covered by at least one RND window passing our threshold of information content (see "Methods"). Among the 3312 genes overlapping introgressed regions (in at least one individual), 166 were mitonuc genes, which does not reflect an enrichment (Pearson's Chi-squared test p value = 0.554). Introgression frequency of mitonuc genes followed the general genomic pattern, being mostly rare (Additional file 2: Figure S4). However, six mitonuc genes (TYMP, TMLHE, L2HGDH, ATG5, SDHAF4, and RARS2) were found introgressed at high frequencies (> 80%; Additional file 1: Table S8). Furthermore, 17 mitonuc genes showed a pattern of introgression that resembles that of mtDNA (absence of introgression in the ten southern haploid genomes and at least 20% of introgression in the ten northern ones) (Additional file 1: Table S9). For these 23 genes, we inspected rates of synonymous to non-synonymous substitutions and the impact of amino acid replacements between the alleles of *timidus* and *granatensis* (or *americanus* in the case of genes introgressed at high frequencies). No dN/dS value above 1, which would have indicated evolution under positive selection, was found. There were 11 non-synonymous variants in four genes, but only in two (SDHA4 and TMHLE) were these variants predicted to potentially influence protein function (Additional file 1: Table S10), according to SIFT.

Heterogeneity of introgression across the genome

The mean proportion of introgression across individuals was significantly lower on the X chromosome (0.24%) than on the autosomes according to the ELAI inferences (2.04%; Mann-Whitney U test p value << 0.01; Fig. 6a). The pattern was also observed using RND (Additional file 2: Figure S5). Based on the chromosomal position of informative SNPs, we examined variations along the chromosomes of the prevalence of introgression, measured as the number of ELAI introgression segments across all individuals overlapping a given SNP. We found no correlation with the distance to the centromere (Additional file 2: Figure S6). However, when separating metacentric/submetacentric and telocentric/acrocentric/subtelocentric chromosomes, such correlation was found for the former but not the latter (Additional file 2: Figure S7a). Such a pattern is suggestive of a correlation with distance to the chromosome center, which roughly coincides with the centromere in the metacentric/submetacentric chromosomes. Indeed, we confirmed that introgression frequency increased significantly with distance to the chromosome center (Spearman's rank correlation test p value << 0.01, $\rho = 0.74$; Fig. 6b), independently of the position of the centromere (Additional file 2: Figure S7b). Using LDhat [51, 52], we estimated the variations of population recombination rate along the chromosomes and also found a significant positive correlation with distance to the chromosome center, though with a lower coefficient (Spearman's rank correlation test p value << 0.01, $\rho = 0.14$; Fig. 6c).

Discussion

A null demographic model explains geographic patterns of nuclear introgression

A northward post-glacial expansion of *L. granatensis* into a territory occupied by *L. timidus*, where hybridization occurred, would leave distinctive traces in

Fig. 6 Variation of introgression prevalence along the genome. **a** Distribution of the proportion of introgression across individuals for autosomes (*Aut*) and the X chromosome (*X*) (Mann-Whitney U test $p = 0.00$). **b** Correlation between prevalence of introgression (number of introgressed ELAI segments overlapping a given position) and relative distance to the chromosome center (Spearman's rank correlation $p = 0.00$). **c.** Correlation between population recombination rate (ρ) and relative distance to chromosome center (Spearman's rank correlation $p = 0.00$). *Dashed line* indicates linear regression trendlines

Seixas et al. Genome Biology (2018) 19:91

Page 10 of 20

genomic variation. First, there should be traces of a demographic expansion of *L. granatensis*, concomitant with a contraction of *L. timidus*. Our PSMC analyses indeed suggest inversely related past demographic profiles of the two species, expansion of one being contemporaneous with retraction of the other (Fig. 1c). The method was unable to recover reliably demographic profiles at the presumed recent time of contact between the two species. However, the demography of past Iberian populations of *L. timidus* could not have been estimated since the present-day samples used in the inference are not descendants of populations from this region, which are now extinct. Still, past demographic profiles strongly favor contrasting demographic consequences for the species, given their adaptation to distinct environments, one temperate and the other boreal [53].

A second prediction of the invasion with replacement model is a gradient of genetic variation, correlated with distance to the origin of the range expansion. Our PCA analysis revealed such a gradient independently of the differentiation with *L. timidus* (Fig. 2), in keeping with previous inferences of range expansion from southwest Iberia. Third, we predict increased introgression in the direction of the expansion, which we clearly confirmed (Fig. 3a). Fourth, we predict that the age of introgression corresponds to the last de-glaciation. We obtained different estimates depending on the method used (IBS tract length distributions or average introgression tract length; 24–7 kya) but they are compatible with hybridization occurring at the end of the last glacial period and possibly persisting towards the Holocene. Independently of the absolute age of the introgression, the invasion model would predict a gradient of introgression age, from most ancient at the initial front of invasion to more recent in more recently invaded territories. This exactly matches the inferred gradient of northward increase of average introgression tract sizes, longer tracts reflecting more recent introgression (Fig. 3b).

By explicitly simulating the proposed invasion-replacement model, we were able to reproduce the empirical patterns of prevalence of introgression observed in our nuclear data (Fig. 4b and Table 1). Simulations resulted in low introgression frequencies, with a strong bias towards rare introgression, and a northward increase of introgression frequencies. Interestingly, when dividing the profiles between the northern and southern samples, empirical and simulated results strikingly coincide, with a steep cline in the south and a shallower transition in the north (Fig. 5a). In the simulated scenario, the cline in the south can only be produced by diffusion of introgressed variants from the region of hybridization into the native range, whereas the northern cline is produced by the invasion-hybridization process. Analyzing the same divide for the introgression tract lengths, we found similar profiles, suggesting slow diffusion of introgression towards the

south and rapid, repeated hybridization during the northern invasion (Additional file 2: Figure S8). These results also strongly oppose a competing hypothesis that introgression would have occurred across a static hybrid zone between the two species in northern Iberia, with southward diffusion of introgressed variants, because this should produce similar gradients in the entire range.

The overall empirical proportion of nuclear introgression was lower than in the simulations, but this could be due to the power to inventory all introgression tracts, and also to hybrid incompatibilities (not accounted for in the simulations) limiting introgression. Indeed, we found non-random patterns of introgression along the genome, suggesting the impediment of introgression by selection. Introgression is significantly reduced for the X chromosome compared to the autosomes (Fig. 6a), which suggests a disproportionate effect of the X in the establishment of reproductive isolation (large X effect [54]), resulting in reduced X-linked introgression [14, 55–61]. We also found that introgression prevalence and recombination rates increase from the center of the chromosomes to their end (Fig. 6b, c). Such correlation between introgression and chromosomal position, possibly linked to recombination [62], is typically observed across a range of hybridizing taxa [63–67]. It could result from the existence of numerous incompatibility loci spread along the genome, or from the consequences of the expected higher density of deleterious mutations in low recombining regions, which can be especially prevalent if the donor species has a higher genetic load than the recipient one ([68] and references therein).

Though simplistic, our simulations capture in a reasonably realistic manner important characteristics of the demography of the species, a prominent source of stochastic variation that is expected to affect the whole genome alike. Hybridization during a range replacement appears to be the major determinant of average geographic patterns of variation of nuclear introgression in our system and establishes a null demographic framework within which discordant introgression patterns can be mined.

Mitochondrial DNA introgression conforms to the null demographic model

Overall, our results are compatible with the invasion-replacement hypothesis and the nuclear and mitochondrial genomes share similar patterns of increased introgression towards the north. However, levels of nuclear introgression are much lower than those found for mtDNA, and the northward gradient is much shallower (Fig. 3b). We found that mimicking the haploid nature and maternal transmission of mtDNA, and female philopatry, we were able to reproduce these empirical mtDNA introgression patterns (Fig. 5b). These settings represent commonly invoked causes for

Seixas *et al. Genome Biology* (2018) 19:91

Page 11 of 20

preferential mtDNA introgression. First, the lower effective population size of mtDNA increases the probability for introgressed variants to reach high frequencies occasionally. Second, lower intra-specific migration resulting from female philopatry decreases the probability that introgressed variants in the invasion front are diluted by migration of native alleles from the parental populations [18, 24]. Male hares, as commonly described for many other mammals, tend to disperse farther than females [69, 70]. This causes interspecific crosses to occur preferentially between *L. timidus* females and *L. granatensis* males at the invasion front, and thus asymmetric mtDNA introgression [71]. Such frequency-dependent female assortative mating would also explain the absence of *L. timidus* introgression for the Y chromosome in *L. granatensis* reported by Melo-Ferreira et al. [28] based on extensive sampling.

In a recent study, Bonnet et al. [72] simulated under a multi-locus framework several demographic and selective scenarios to test cytonuclear discordance in patterns of introgression, including sex-related asymmetries, spatial invasion-replacement, and selection either promoting mtDNA introgression or impeding introgression at nuclear loci. They conclude that only positive selection on mtDNA could produce its massive introgression with low levels of nuclear gene flow. The apparent discordance with the present work can nevertheless be explained by two simple factors. First, Bonnet et al. [72] focused on global introgression frequencies, not only at the invasion front. mtDNA introgression in *L. granatensis* is predominant at the invasion front (the north) but not over the species range. Second, asymmetric gene flow was not considered in a scenario of range invasion, and we show here that it is required to reproduce the mtDNA pattern of introgression. Our results thus suggest that selection does not need to be invoked to account for this type of cytonuclear discordance, at least in our study species.

Consequences of mitochondrial DNA introgression
Our work suggests that the massive but geographically limited mtDNA introgression from *L. timidus* into *L. granatensis* may have been an accident of the demographic dynamics of a range replacement. mtDNA introgression could therefore lead to incompatibilities of heterospecific combinations of nuclear and mitochondrial genes co-controlling a given phenotype. Cases of cytonuclear incompatibilities have been reported in a variety of organisms, between closely related species [73, 74], or even between populations of the same species [75–77], including in hares [78]. There are theoretical reasons to predict rapid cytonuclear coevolution. One is the rapid rate of evolution of the animal mitochondrial genome and its reduced effective population size and absence of recombination [73, 79–81]. There is, however, no clear evidence for

the action of Muller's Ratchet on the animal mitochondrial genome [73]. The other reason for rapid cytonuclear coevolution is maternal mtDNA transmission and consequent spread of neutral or beneficial mtDNA mutations for females, even if harmful for males, which do not transmit them to later generations (the mother's curse [25, 82]). Such a phenomenon is expected to be counteracted by compensatory mutations in nuclear genes, which are transmitted by both sexes, thus causing rapid coevolution of the two genomes.

Regarding nuclear genes reported to be involved in the mitochondria ("mitonuc" genes), we found no significant differences of pairwise species dN/dS compared to background genes (Wilcoxon rank sum test, p value > 0.05). However, we identified six genes with high frequency introgression (i.e., outliers in our most relaxed demographic model) and 17 with a geographic distribution of introgression resembling that of mtDNA (in geographical and frequency pattern). Two genes (TMLHE and SDHF4) of the former category showed amino acid differences between the *timidus* and *americanus* or native *granatensis* sequences, respectively, which were predicted to have a strong functional impact, taking into account the conservation levels of the residues at deep evolutionary scales. SDHAF4 is essential for the assembly of succinate dehydrogenase (SDH; respiratory complex II), which participates in the tricarboxylic acid (TCA) cycle and in the mitochondrial electron transport chain. This gene is also possibly required to protect against ROS (reactive oxygen species) toxicity, i.e., oxidative stress [83]. TMLHE is involved in carnitine biosynthesis, an antioxidant that might protect mitochondria from oxidative stress [84]. The control of oxidative stress is an important component of many aspects of physiology and reproduction, and its disruption has been reported to occur in situations of hybridization [74, 85, 86]. These genes are thus candidates to have been affected by cytonuclear co-evolution during and after the hybridization events. However, the functional relevance of these differences must be addressed in future functional assays.

It is striking that among our set of 123 genes with outlying introgression frequencies, we found enrichment of functions related to spermatogenesis, concerning seven genes: ALMS1, ARID4B, SPATA6, SLC9C1, KIAA1109, GMCL1, and NEK1. Proving selection-driven introgression continues to be a major challenge [87], as introgression alone may lead to patterns that can be interpreted as resulting from selection using population genetic statistics designed to detect intraspecific selective sweeps (e.g., extended LD, shift in allele frequencies). However, these genomic regions were shown here to introgress at higher rates than our conservative neutral demographic expectations. These are compelling candidates for selection-driven introgression, especially given the functions with possible

impact on male fertility. The disruption of mtDNA and nuclear DNA co-evolved combinations has been shown to affect male fertility in several biological systems [75–77], including in hares [78]. Studies evaluating the fertility of *L. granatensis* males with distinct mitochondrial and nuclear backgrounds would help to test this new hypothesis.

Adaptive introgression between the two species

Independently of mitochondrial introgression, we searched for evidence of adaptive introgression in our system, in the frame of our demographic model. Evidence of adaptive introgression has now been suggested from the analyses of genomic datasets in several animal species, for instance, in humans [88], mice [14, 89, 90], butterflies [8, 91, 92], mosquitoes [57, 93, 94], or hares [15]. The interrogation of the functions of 123 genes for which introgression frequencies could not be predicted by our simulations (Fig. 4b and Table 1) revealed, in addition to spermatogenesis, enrichment in innate immune response functions. Adaptive introgression of immune-related genes has been inferred, for instance, in humans [4, 95–100], mosquitoes [101], the Alpine *Ibex* [102], and house mice [103, 104]. Viral diseases, such as rabbit hemorrhagic disease (RHDV) and myxomatosis (Myxoma virus) for rabbits, and the European brown hare syndrome (EBHSV) for hares, strongly affect the Iberian populations of lagomorphs. Variants of these viruses are known to change host-specificity and affect other species, such as RHDV2 that affects hares [105, 106] or EBHSV that affects American rabbits (*Sylvilagus*) [107]. Interestingly, one of the genes found here introgressed at high frequencies; interleukin 12B (IL12B) has been implicated in the inflammatory process and immune response to RHDV and Myxoma virus in rabbits [108], and to have adaptively introgressed from Neanderthals to modern humans in Europe [4]. These inferences thus strongly suggest that the invasion of new territories with new pathogenic pressures may have been facilitated by the incorporation of adapted genetic variants through introgression.

Conclusions

Speciation research has traditionally focused on processes leading to species divergence and isolation. In this respect, our results are in line with several other studies, i.e., reduced admixture of the X chromosome compared to the autosomes. We were able to demonstrate the genome-wide positive relationship between recombination and admixture without relying on the often used but potentially misleading differentiation proxy [109]. Altogether, our results indicate that selection spread over many genomic regions, and particularly on the X, is preventing free admixture of the genomes of these species, although as in many other systems the exact causes of selection are unknown.

However, we were particularly focused on general evolutionary mechanisms that promote admixture between partially reproductively isolated species. We provide evidence quantitatively evaluated by simulations that demographic processes accompanying invasive replacement of one species by the other, with male-biased migration, can determine introgression patterns genome-wide, including strong cytonuclear discordance of admixture levels. This provides an important general null framework to interpret numerous instances of cytonuclear introgression discordance (reviewed, e.g., by Toews and Brelsford [20]).

Having set this framework, we could pinpoint outlier candidate genes for selection-driven introgression, some of which have suggestive functions. For innate immunity genes, adaptation to the environment is an obvious cause of positive selection. For spermatogenesis genes, a role of genetic conflicts, thus having nothing to do with the environment, can be suspected. It cannot be excluded that the candidate mitonuc genes are also involved in cytonuclear conflicts. Therefore, in all cases selection-driven introgression could result from the consequences of genetic conflicts, either between different species (with parasites), or different parts of the same genome (cytoplasmic and nuclear). Genetic conflicts are often invoked to explain the accumulation of interspecific incompatibilities (reviewed by Crespi and Nosil [110]), but our results suggest they could also create the conditions for extensive admixture. The functional interpretations proposed here will have to be tested by phenotypic assays. Progress in their validation could also come from the analysis of replicated cases of introgression—in the two other situations of massive mtDNA introgression in Iberia, with *L. europaeus* and *L. castroviejoi*.

Methods

Sampling, genomic DNA extraction, library construction, and sequencing

We performed whole genome sequencing of ten Iberian hares (*L. granatensis*) and three mountain hares (*L. timidus*), the geographical origins of which are shown in Fig. 1a, b, as well as one snowshoe hare (*L. americanus*) (Additional file 1: Table S1). All specimens were females and samples were donated from hunting campaigns or collected from individuals found dead. We used the JETquick Tissue DNA Spin Kit (GENOMED) to extract genomic DNA from ear or internal organ tissues that had been preserved in RNAlater or ethanol. Illumina TruSeq DNA v2 genomic libraries with inserts of 600 bp were prepared for the 14 samples and pair-end sequenced (2 ×

Seixas *et al. Genome Biology* (2018) 19:91

Page 13 of 20

100bp) on an Illumina HiSeq 2500 platform at The Genome Analysis Centre (TGAC, Norwich, now Earlham Institute). We also used 30.7 Gb of further sequence data previously generated for the same *L. americanus* individual [111].

Data filtering, read mapping, genotype calling, and iterative mapping

Raw sequence reads were filtered by removing the first 5 bp and adapters at the end of reads using Cutadapt version 1.8 [112]. Low quality bases were removed using Trimmomatic v0.33 [113] by trimming bases with a quality score lower than 20 at the end of the reads and using a sliding window of 4 bp for a minimum average quality of 30. Reads shorter than 36 bp were discarded. Trimmed reads were mapped to the rabbit reference genome available from Ensembl (OryCun2.0, release 80) using the BWA-MEM algorithm with default parameters [114]. Correction of read pairing information and flags and sorting of mapped reads by coordinates were performed with Samtools v1.3 [115]. Soft clipped bases were further removed using NGSutils version 0.5.7 [116]. Reads were then realigned around INDELs using the Genome Analysis Toolkit (GATK v3.2–2 [117, 118]). Finally, Picard Markduplicates (http://broadinstitute.github.io/picard/) was used to remove read duplicates.

Multi-sample SNP/genotype calling was carried out using the algorithm implemented in Samtools v1.3 for each species independently, requiring minimum base and mapping qualities of 20. Species VCF files were then merged and genotypes filtered using a minimum site quality (QUAL) of 20, RMS minimum mapping quality (MQ) of 20, minimum individual coverage (FMT/DP) of 8X, and maximum overall coverage (DP) of 430X. For variable sites, a minimum genotype quality (FMT/GQ) of 20 was required. All sites failing any of the filtering criteria were coded as missing data. Furthermore, genotypes closer than 10 bp from INDELs were excluded.

In order to improve mapping efficiency, we used the first round of mapping and SNP calling to build a hare pseudo-reference genome, by replacing each base in the rabbit reference by that inferred in hares whenever the latter was found fixed for a state different from the rabbit reference. We used the resulting pseudo-reference to redo the mapping and SNP calling steps. Insertion-deletions were not considered to build the pseudo-reference, so that the rabbit genome coordinates were kept. This iterative mapping procedure has been shown to improve mapping efficiency when using a divergent reference genome [32, 119] (diverging by 5% in this case).

Haplotype phasing
We used SHAPEITv2.r837 [120] to perform read-aware phasing, including both *L. granatensis* and *L. timidus*

specimens, as we were particularly interested in phasing introgressed regions. Phase informative reads (PIRs), i.e., those that span at least two heterozygous sites and thus help local phasing [121], were extracted from the individual bam files, and phasing was performed using only bi-allelic sites with no more than two individuals with missing information. We ran SHAPEIT for each chromosome using a window size of 0.5 Mb (as recommended in the manual) with a MCMC run of 50 main iterations, with ten burn-in and ten pruning iterations. We specified an effective population size of 100,000, following the estimates derived in the present paper and by Melo-Ferreira et al. [34] and a recombination rate of 1 cM/Mb, as inferred for rabbits [122].

Estimate of mutation rate
We estimated mutation rate (μ) based on the sequence divergence between *L. americanus* and rabbit assuming $\mu = D_{XY}/(2T_D + 4Ne)$ [123], where D_{XY} [124] is the distance between hares and rabbits averaged across autosomes, T_D is the time of divergence (11.8 million years, following Matthee et al. [125]), and Ne the ancestral effective population size. We assumed a generation time of 2 years [126] and an ancestral effective population size of 1,000,000.

Inference of introgression—Efficient Local Ancestry Inference (ELAI)
In order to infer genomic segments of *L. timidus* origin introgressed in *L. granatensis* we used the Efficient Local Ancestry Inference (ELAI) method [38]. This method implements a two-layer HMM (hidden Markov model) to infer local ancestry of admixed individuals without prior definition of window sizes, by looking at two layers of linkage-disequilibrium—within and among defined groups. It returns at each variable position in the genome the most likely proportions of ancestries (true values being expected to take values 0, 1, or 2 in two-way admixture). We ran ELAI on the unphased dataset and two population samples: *L. granatensis* defined as the admixed population, and *L. timidus* defined as one of the donors in the admixture. We did not have a pure *L. granatensis* population and therefore let ELAI infer this second ancestry from the data of the admixed population. We set the number of upper-layer groups to 2, representing *L. timidus* and *L. granatensis*, and that of lower-layer clusters to 10 (five times the number of upper-layer clusters, as recommended). We performed three different expectation maximization (EM) runs of 20 steps with mixture generation values of 5000, 10,000, and 20,000 and different random seeds. ELAI results were averaged over the three independent runs. Sites with a proportion of *L. timidus* ancestry between 0.8 and 1.8 were considered heterozygous for introgression and those with values over 1.8 homozygous for introgression.

Seixas *et al. Genome Biology* (2018) 19:91

Page 14 of 20

For each individual, introgression fragments where defined as consecutive sites defined as introgressed according to the above criteria.

To evaluate the power to detect introgression using ELAI we artificially introgressed random portions of chromosome 1 from *L. timidus* into *L. granatensis* using our phased data. Several introgression fragment sizes—10, 30, and 50 kb—were used. For each introgression tract length, we artificially "introgressed" 200 non-overlapping sequence tracts taken from one *L. timidus* haplotype (from the Alps), replacing the orthologous tract in a randomly chosen *L. granatensis* haplotype. The "introgressed" fragments had a minimum of 100 informative sites and did not span the centromere. Five artificially introgressed datasets were generated for each fragment length (i.e., 1000 fragments per length) and ELAI was run as described above for the real data. We expressed the power of ELAI to detect introgression for each fragment length as the proportion of artificially introgressed fragments for which the average ancestry of informative SNPs within the fragment was at least 0.8.

Dating introgression

To infer the age of introgression we used an approach based on identical by state (IBS) tracts of DNA shared within and between populations [43]. We used the phased dataset for the ten *L. granatensis* individuals and the two *L. timidus* individuals sampled in the Alps to minimize potential effects of substructure within our geographically widespread *L. timidus* sample (Fig. 1b). Only sites segregating in this subset were considered. Furthermore, sites with missing genotypes in *L. timidus* or more than 40% missing genotypes in *L. granatensis* were removed. We generated sets of IBS tracts shared within *L. granatensis*, within *L. timidus*, and between the species for the 21 autosomes. We excluded regions of low SNP density (centromeric regions, regions with more than 10,000 consecutive 'N' bases in the reference genome, or regions between SNPs that are 5000 bp or more apart) in order to avoid erroneously inferring large IBS tracts that span these regions. IBS tracts shared between haplotypes from the same species are informative about the species demographic history while IBS tracts shared between species are informative about their divergence times and the fraction and timing of past genetic exchanges. We inferred demographic parameters under several demographic models, considering one or four pulses of introgression, and either constant or variable population size (Additional file 1: Table S3). IBS tract length distributions within species and between species were computed and jointly fit to the observed data. In order to improve computation time and numeric stability, we binned the IBS tract

length data by computing the expected abundance of tracts between $(3/2)^n$ and $(3/2)^{n+1}$ bp. We further excluded IBS tracts shorter than 300 bp (following Liu et al. [127]) or 10,000 bp since longer tracts are presumably more informative regarding introgression time [43].

We also estimated introgression time from the distribution of introgression tract lengths, as inferred with ELAI for the ten *L. granatensis* genomes, assuming that the distribution is exponential with mean $1/rt$, where t is the number of generations since the admixture event and r is the recombination rate per base pair [40]. We considered a generation time of 2 years and used estimates of recombination rate in rabbits ($r = 1.0 \times 10^{-8}$) [122].

Long-term demographic profiling of the species

We inferred the long-term demographic histories of *L. granatensis* and *L. timidus* with the Pairwise Sequentially Markovian Coalescent (PSMC) method [39], applied to the diploid genome sequence of each individual. Individuals' diploid consensus sequences were generated for each autosome with Samtools v1.3 mpileup, requiring minimum base and mapping qualities of 20, and coverage between 8 and 50X. Generation time was set to 2 years and the mutation rate (μ) to 2.8×10^{-9} substitutions/site/generation, estimated as described above. The atomic time intervals were set to $4 + 50*2 + 2 + 4$, meaning that the first parameter spans the first four atomic intervals, each of the next 50 parameters spans two atomic intervals, while the last two parameters span two and four atomic intervals, respectively.

Principal component analysis

We explored population structure in *L. granatensis* using principal component analysis (PCA), as implemented in PLINK 1.9 [128, 129], based on a subsample of bi-allelic SNPs at least 50 kb apart and without missing genotypes. The PCA analysis was performed on *L. granatensis* together with either a *L. timidus* or a *L. americanus* individual.

Spatially explicit coalescent simulations of demographic expansion and introgression

Using the spatially explicit coalescent simulator SPLATCHE2 [44], we simulated the presumed history of the interaction between *L. timidus* and *L. granatensis*. The Iberian Peninsula was subdivided in demes of 50×50 km, and *L. granatensis* was simulated to expand from a deme located in southwest Portugal [31] 20,000 years ago, progressively replacing the resident *L. timidus* in the northern half of Iberia. The range of *L. timidus* in the Northern demes was determined based on a minimum probability of presence of 0.8 at the last glacial maximum, as predicted by ecological niche modeling [45]. All simulations were performed using a density-independent competition

Seixas *et al. Genome Biology* (2018) 19:91

Page 15 of 20

model (model 6) in two layers (as used in Currat et al. [18]), corresponding to the two species, and implied the complete replacement of *L. timidus* by *L. granatensis* at the time of sampling. Admixture between layers was allowed in co-occupied demes. As in Currat et al. [18], the intrinsic growth rate was set to a fixed value (0.5) and different carrying capacities, migration rates, and admixture rates were tested, totaling eight combinations of parameter values. Two values of deme carrying capacity (K) of *L. granatensis* were considered, K = 1000 and K = 10,000. The first corresponds to an inferred effective population size of ~ 100,000 (this work and Melo-Ferreira et al. [34]) divided by the ~ 200 demes in our grid covering species distribution. The second value of K used increases by ten times the estimates of effective population size to evaluate the influence of this parameter on proportions of introgression. During the replacement, the carrying capacity of *L. timidus* was considered half of that for *L. granatensis*. Two migration rates between adjacent demes were tested—M = 0.02 and M = 0.2—and bidirectional admixture at two distinct rates was assumed—gamma = 0.005 and gamma = 0.03. Larger carrying capacities and admixture rates and lower migration rates were expected to result in higher levels of introgression [18]. We simulated 100 replicates of genomic introgression (forward demographic and backwards coalescent simulations) per set of parameter values, each corresponding to 50,000 independent markers. We recorded the proportion of introgressed loci for each of ten *L. granatensis* simulated individuals, located in demes corresponding to the geographical locations of the empirical samples.

To evaluate the expected mitochondrial DNA introgression patterns under these simulated demographic scenarios, we also simulated mitochondrial introgression under the same conditions, but modifying some of the parameters to fit the specific ploidy and transmission characteristics of this genome. We reduced the carrying capacity (K) to ¼ of that of the nuclear genome (250 and 125 for *L. granatensis* and *L. timidus,* respectively). We also set inter-deme migration to the minimum (M = 0.005) to mimic female philopatry. Gene flow was set to be predominant from *L. timidus* into *L. granatensis* (A = 0.025 from *L. timidus* to *L. granatensis* and 0.001 in the other direction) to mimic the consequences of male-mediated migration during the northward colonization of *L. granatensis*. An intrinsic growth rate of 0.5 was maintained. We simulated 10,000 replicates for each of the other parameter sets with only one marker per simulation, sampling 20 individuals per locality. The frequency of introgression was recorded per locality per simulation replicate.

Inference of outlier regions of introgression
In order to detect genomic regions with high frequencies of introgression, we could not use ELAI because we did not have a pure *L. granatensis* reference population. We therefore analyzed variations of the relative node depth (RND) [46] along the genome. Using mvftools [130] and custom R scripts, we calculated RND from the phased data on non-overlapping windows of 10, 20, or 50 kb, with at least 50 informative sites. We calculated for each *L. granatensis* haplotype its average nucleotide divergence (Dxy) [124] to all *L. timidus* haplotypes, which we divided by the divergence between *L. timidus* and *L. americanus* in order to standardize for potential variations of mutation rates across windows.

Introgression events (whatever the introgression frequency) are expected to produce exceptionally low RNDmin values (minimum RND value among haplotypes in each window [47]), but defining thresholds based on empirical distributions can be arbitrary. Therefore, we used ELAI inferences as reference to perform power and false discovery rate (FDR) analyses of the RNDmin method. This analysis was restricted to introgression frequencies in the range that could be detected by ELAI (maximum 65%). RND windows only partially overlapping ELAI segments were not considered. On this basis, we estimated the FDR and power of the detection of introgression by RND as a function of the RNDmin threshold. A threshold predicting a FDR of 10% was used (Additional file 2: Figure S2).

Regions of the genome with outlier high frequencies of introgression were defined based on the simulated demographic scenario using SPLATCHE2. For each parameter set, we recorded the minimum frequency of introgression at which at least 95% of the simulation replicates suggest a lower proportion of introgressed markers than was inferred in the empirical dataset (for all three RND window lengths). We conservatively chose the highest threshold among our eight simulated sets of parameters to define outlier regions of introgression frequency in the empirical data.

GO enrichment analyses
We tested for functional enrichment of genes with high introgression frequencies (combining the evidence from the three RND window lengths) using the g:Profiler R package [131, 132]. Categories with less than five genes were excluded and the Benjamini-Hochberg correction for multiple testing was applied. Only genes within or overlapping RND windows with more than 50 informative sites in any of the three RND window length analyses were considered for the background list of genes. We used both the rabbit GO term annotation and the more complete mouse one. For the latter, only one-to-one rabbit to mouse orthologous genes were considered. GO terms were summarized using REVIGO [133].

Seixas et al. Genome Biology (2018) 19:91

Page 16 of 20

Analyses of nuclear genes with mitochondrial functions

We generated a list of nuclear genes with mitochondrial functions (mitonuc genes) by combining two public databases: InterMitoBase [49] and MitoCarta2.0 [50]. These databases provide lists of human annotated genes encoding proteins that are present in the mitochondria. We identified rabbit orthologous genes using the Ensembl Biomart query tool [134]. Of the 708 human annotated nuclear genes in InterMitoBase, 615 were found annotated in the rabbit, while 1030 genes from the 1147 nuclear genes from Mitocarta2.0 were annotated in the rabbit genome. The union of the two databases resulted in 1210 mitonuc rabbit annotated genes. We further added one OXPHOS gene (NDUFA4L2) that was missing from both databases.

From the sets of mitonuc genes, we verified those showing a geographic introgression pattern mimicking that of mtDNA: i) absence of introgression in southern individuals (no mtDNA introgression is found in the south) [21, 30]; ii) at least two introgressed haplotypes in the five northernmost samples. This is the expected frequency if introgression frequencies at these genes were at least as high as those documented for mtDNA in the northern populations [45] (Additional file 2: Figure S9). For each gene, the window with the highest total frequency of introgression was retained.

Gene variation statistics and functional impact of amino acid differences

We produced species pairwise alignments (between *L. timidus* and *L. granatensis* and *L. timidus* and *L. americanus*) from the phased genomes for all rabbit annotated genes (19,280). For each gene, we obtained the exon coordinates of the largest transcript from the Ensembl Biomart query tool. We excluded from the alignments sites with more than two alleles. Alignments including SNPs with allele frequencies markedly deviating from Hardy-Weinberg proportions in either *L. timidus* or *L. granatensis* (exact test p value < 0.01; using Plink 1.9) were discarded, as it may result from the inclusion of paralogs. Sequences with more than 50% missing data were removed from the alignments. Furthermore, haplotypes in *L. granatensis* inferred to be of *L. timidus* origin were excluded from the *L. granatensis* alignment. Sites with less than four haplotypes with information in either *L. timidus* or *L. granatensis* or with no information in *L. americanus* were masked with Ns. Finally, alignments with less than 100 codons or with premature stop codons were removed. We estimated dN and dS (Jukes-Cantor; rates of non-synonymous and synonymous substitutions, respectively) using the Bioperl DNAStatistics module (available in http://search.cpan.org/dist/BioPerl/Bio/Align/DNAStatistics.pm) and dN/dS was calculated as the average of dN/dS pairwise estimates.

In order to examine the potential functional impact of amino acid differences, we used the SIFT Aligned Sequences tool implemented in SIFT v1.03 [48] (available at http://sift.jcvi.org). This method assumes that amino acid changes occurring in a given lineage at positions otherwise conserved at a deeper phylogenetic scale likely affect protein function. Alignments of chordate orthologous sequences for candidate genes with amino acid changes were obtained from the EggNOG 4.5.1 database [135] and aligned to our *Lepus* translated sequences using MUSCLE v3.8.31 [136]. Functional changes were assumed for normalized probabilities of tolerated change ≤ 0.05.

Relationship between chromosome position and introgression

We tested the correlation of introgression and recombination with position along the chromosomes, expressed either by the relative distance to the centromere or to the chromosome center. The population-scaled recombination rate coefficient (ρ) was estimated along the *L. granatensis* genome using the reversible-jump MCMC algorithm *interval* implemented in LDhat v2.2 [51, 52]. The method fits a uniform recombination rate over a region from patterns of linkage disequilibrium across genotypes. We selected only variable sites without missing information with VCFtools v0.1.15 [137] to create LDhat input files. We calculated ρ along the chromosomes in segments of up to 2000 variable sites, as recommended for the method. The interval algorithm was run for 1,000,000 iterations, sampling every 5000 iterations, discarding the first 10% as burn-in. We specified a block penalty of 5 in all analyses. We then attributed to each SNP the ρ value of the LDhat fragment in which it was included. Introgression prevalence at a given SNP position in the genome was measured as the number of ELAI introgressed fragments across individuals overlapping that SNP. The relative distance of a SNP to either the centromere or the chromosome center was calculated by dividing the distance to this reference point (in base pairs) by the length of the chromosome arm or chromosome length, respectively.

To ensure independence, we subsampled SNPs that were at least 50 kb apart. Rabbit chromosomes 1 and 2 were excluded given their known structural differences between rabbits and hares (both are split in hares [33]). Chromosomes were classified as metacentric, submetacentric, subtelocentric, acrocentric, and telocentric according to arm ratio estimates [138], based either on karyotype measurements of the long and short arms (see [139]) or sequence lengths of the two arms in the rabbit reference genome (Additional file 1: Table S11). When analyzed separately based on centromere position, only chromosomes with consistent classification following these criteria were considered. Spearman's rank correlation was used to test the correlation between prevalence

Seixas *et al. Genome Biology* (2018) 19:91

Page 17 of 20

of introgression and recombination with relative distance to the centromere or to the chromosome center. In the first case, SNPs were grouped by bins of distance and the prevalence of introgression re-calculated as the sum of introgression frequencies across SNPs within a bin, while in the latter the correlation was tested with all subsampled SNPs.

Additional files

Additional file 1: Table S1. Basic information of specimens sequenced in this study. **Table S2.** ELAI's Power for inferring introgression according to our simulations of artificial introgression. **Table S3.** Demographic inferences from IBS tracts. **Table S4.** List of nuclear genes overlapping regions with outlier introgression frequencies. **Table S5.** Gene Ontology functional enrichment analyses of genes overlapping regions with outlier frequencies of introgression. **Table S6.** Summary of GO functional categories significantly enriched in the set of genes with outlier introgression frequencies. **Table S7.** Nonsynonymous mutations detected within 123 high frequency introgression genes and their potential functional impact inferred using SIFT. **Table S8.** List of mitonuc genes with outlier frequencies of introgression. **Table S9.** List of mitonuc genes with geographic patterns similar to that of mtDNA. **Table S10.** Nonsynonymous mutations detected within mitonuc gene candidates to have co-introgressed with mitochondrial. **Table S11.** Classification of chromosomes according to centromere position. (XLSX 58 kb)

Additional file 2: Figure S1. Distribution of differential levels of average introgression between the five northern and five southern individuals across the 1000 simulations of mitochondrial introgression. **Figure S2.** Power and false discovery rate of the relative node depth method for inferring introgression. **Figure S3.** Empirical distribution of nuclear introgression frequencies inferred with RND. **Figure S4.** Introgression frequency distribution of mitonuc and background genes. **Figure S5.** Variation of the proportion of introgression across individuals for autosomes and the X chromosome. **Figure S6.** Correlation between prevalence of introgression (estimated with the ELAI method) and relative distance to the centromere for all chromosomes. **Figure S7.** Correlation between prevalence of introgression (estimated with the ELAI method) and relative distance to **a** the centromere and **b** the chromosome center for each chromosome category. **Figure S8.** Correlation between introgression tract size and geography. **Figure S9.** Expected introgression frequency distribution in a sample of ten *L. granatensis* individuals with the same geographic origin as the ten samples used in this study, considering empirical mtDNA introgression frequencies. (PDF 1177 kb)

Abbreviations

ALMS1: ALMS1, centrosome and basal body associated protein; ARID4B: AT-rich interaction domain 4B; ATG5: Autophagy related 5; E230025N22Rik: Riken cDNA E230025N22 gene; GMCL1: Germ cell-less, spermatogenesis associated 1; Herc6: Hect domain and RLD 6; IL12B: Interleukin 12B; L2HGDH: L-2-hydroxyglutarate dehydrogenase; MSH6: mutS homolog 6; NEK1: NIMA related kinase 1; OPTN: Optineurin; RARS2: Arginyl-tRNA synthetase 2, mitochondrial; SDHAF4: Succinate dehydrogenase complex assembly factor 4; SLC9C1: Solute carrier family 9 member C1; SPATA6: Spermatogenesis associated 6; TMLHE: Trimethyllysine hydroxylase, epsilon; TYMP: Thymidine phosphorylase

Availability of data and materials

The dataset supporting the conclusions of this article is available in the Sequencing Read Archive (https://www.ncbi.nlm.nih.gov/sra) repository as study PRJNA399194 (accession numbers SRX3108378-SRX3108391) [140]. Additional whole genome sequencing data for the *L. americanus* individual were retrieved from NCBI SRA [141]. The rabbit genome reference assembly used to build the hare pseudo-reference assembly was downloaded from Ensembl [142].

Authors' contributions

JMF and PB designed and supervised the work. FAS performed the analyses and wrote the first draft of the manuscript. All authors interpreted the data and read, revised, and approved the final version of the manuscript.

Ethics approval

No animals were killed for the purpose of this research. The samples were donated from animals shot during regular hunting campaigns or that were found dead. The samples were collected before the Nagoya Protocol entered into force on 12 October 2014. The sampled species have a "least concern" status in the IUCN Red List of Threatened Species.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal. [2]Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Rua Campo Alegre s/n, 4169-007 Porto, Portugal. [3]Institut des Sciences de l'Évolution, Université de Montpellier, CNRS, IRD, EPHE, Place Eugène Bataillon, 34095 Montpellier, France.

Seixas *et al. Genome Biology* (2018) 19:91

Page 18 of 20

References

1. Hedrick PW. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol Ecol. 2013;22:4606–18.
2. Tigano A, Friesen VL. Genomics of local adaptation with gene flow. Mol Ecol. 2016;25:2144–64.
3. Rieseberg LH, Kim S-C, Randell RA, Whitney KD, Gross BL, Lexer C, et al. Hybridization and the colonization of novel habitats by annual sunflowers. Genetica. 2007;129:149–65.
4. Quach H, Rotival M, Pothlichet J, Loh Y-HE, Dannemann M, Zidane N, et al. Genetic adaptation and Neandertal admixture shaped the immune system of human populations. Cell. 2016;167:643–56. e17
5. Macholán M, Baird SJ, Munclinger P, Dufková P, Bímová B, Piálek J. Genetic conflict outweighs heterogametic incompatibility in the mouse hybrid zone? BMC Evol Biol. 2008;8:271.
6. Albrechtova J, Albrecht T, Baird SJE, Macholán M, Rudolfsen G, Munclinger P, et al. Sperm-related phenotypes implicated in both maintenance and breakdown of a natural species barrier in the house mouse. Proc R Soc B Biol Sci. 2012;279:4803–10.
7. Beck EA, Thompson AC, Sharbrough J, Brud E, Llopart A. Gene flow between *Drosophila yakuba* and *Drosophila santomea* in subunit V of cytochrome *c* oxidase: a potential case of cytonuclear cointrogression. Evolution. 2015;69:1973–86.
8. The Heliconius Genome Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature. 2012;487:94–8.
9. Nadeau NJ, Martin SH, Kozak KM, Salazar C, Dasmahapatra KK, Davey JW, et al. Genome-wide patterns of divergence and gene flow across a butterfly radiation. Mol Ecol. 2013;22:814–26.
10. Huerta-Sánchez E, Jin X, Asan BZ, Peter BM, Vinckenbosch N, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 2014;512:194–7.
11. Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. Nature. 2014;507:354–7.
12. Vernot B, Akey JM. Resurrecting surviving Neandertal lineages from modern human genomes. Science. 2014;343:1017–21.
13. Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature. 2015;518:371–5.
14. Liu KJ, Steinberg E, Yozzo A, Song Y, Kohn MH, Nakhleh L. Interspecific introgressive origin of genomic diversity in the house mouse. Proc Natl Acad Sci U S A. 2015;112:196–201.
15. Jones MR, Mills LS, Alves PC, Callahan CM, Alves JM, Lafferty DJR, et al. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. Science. 2018;360:1355–8.
16. Stankowski S, Streisfeld MA. Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. Proc R Soc B Biol Sci. 2015;282:20151666.
17. Richards EJ, Martin CH. Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. PLoS Genet. 2017;13:e1006919.
18. Currat M, Ruedi M, Petit RJ, Excoffier L. The hidden side of invasions: massive introgression by local genes. Evolution. 2008;62:1908–20.
19. Excoffier L, Foll M, Petit RJ. Genetic consequences of range expansions. Annu Rev Ecol Evol Syst. 2009;40:481–501.
20. Toews DPL, Brelsford A. The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol. 2012;21:3907–30.
21. Melo-Ferreira J, Boursot P, Suchentrunk F, Ferrand N, Alves PC. Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. Mol Ecol. 2005;14:2459–64.
22. Good JM, Hird S, Reid N, Demboski JR, Steppan SJ, Martin-Nims TR, et al. Ancient hybridization and mitochondrial capture between two species of chipmunks. Mol Ecol. 2008;17:1313–27.
23. Sequeira F, Sodré D, Ferrand N, Bernardi JA, Sampaio I, Schneider H, et al. Hybridization and massive mtDNA unidirectional introgression between the closely related Neotropical toads *Rhinella marina* and *R. schneideri* inferred from mtDNA and nuclear markers. BMC Evol Biol. 2011;11:264.
24. Petit RJ, Excoffier L. Gene flow and species delimitation. Trends Ecol Evol. 2009;24:386–93.
25. Gemmell NJ, Metcalf VJ, Allendorf FW. Mother's curse: the effect of mtDNA on individual fitness and population viability. Trends Ecol Evol. 2004;19:238–44.
26. Altuna J. Hallazgo de una liebre artica (*Lepus timidus*) en el yacimiento prehistorico de Urtiga (Guipuzcoa). Munibe. 1970;22:165–8.
27. Melo-Ferreira J, Boursot P, Randi E, Kryukov A, Suchentrunk F, Ferrand N, et al. The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian peninsula. Mol Ecol. 2007;16:605–18.
28. Melo-Ferreira J, Alves PC, Freitas H, Ferrand N, Boursot P. The genomic legacy from the extinct *Lepus timidus* to the three hare species of Iberia: contrast between mtDNA, sex chromosomes and autosomes. Mol Ecol. 2009;18:2643–58.
29. Melo-Ferreira J, Alves PC, Rocha J, Ferrand N, Boursot P. Interspecific X-chromosome and mitochondrial DNA introgression in the Iberian hare: selection or allele surfing? Evolution. 2011;65:1956–68.
30. Alves PC, Melo-Ferreira J, Freitas H, Boursot P. The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus *Lepus*. Philos Trans R Soc Lond Ser B Biol Sci. 2008;363:2831–9.
31. Marques JP, Farelo L, Vilela J, Vanderpool D, Alves PC, Good JM, et al. Range expansion underlies historical introgressive hybridization in the Iberian hare. Sci Rep. 2017;7:40788.
32. Halligan DL, Kousathanas A, Ness RW, Harr B, Eöry L, Keane TM, et al. Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. PLoS Genet. 2013;9:e1003995.
33. Robinson TJ, Yang F, Harrison WR. Chromosome painting refines the history of genome evolution in hares and rabbits (order Lagomorpha). Cytogenet Genome Res. 2002;96:223–7.
34. Melo-Ferreira J, Boursot P, Carneiro M, Esteves PJ, Farelo L, Alves PC. Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulations. Syst Biol. 2012;61:367–81.
35. Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, et al. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLoS Genet. 2009;5:e1000519.
36. Liu KJ, Dai J, Truong K, Song Y, Kohn MH, Nakhleh L. An HMM-based comparative genomic framework for detecting introgression in eukaryotes. PLoS Comput Biol. 2014;10:e1003649.
37. Martin SH, Davey JW, Jiggins CD. Evaluating the use of ABBA-BABA statistics to locate Introgressed loci. Mol Biol Evol. 2014;32:244–57.
38. Guan Y. Detecting structure of haplotypes and local ancestry. Genetics. 2014;196:625–42.
39. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011;475:493–6.
40. Pool JE, Nielsen R. Inference of historical changes in migration rate from the lengths of migrant tracts. Genetics. 2009;181:711–9.
41. Gravel S. Population genetics models of local ancestry. Genetics. 2012;191:607–19.
42. Liang M, Nielsen R. The lengths of admixture tracts. Genetics. 2014;197:953–67.
43. Harris K, Nielsen R. Inferring demographic history from a Spectrum of shared haplotype lengths. PLoS Genet. 2013;9:1–18.
44. Ray N, Currat M, Foll M, Excoffier L. SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. Bioinformatics. 2010;26:2993–4.
45. Acevedo P, Melo-Ferreira J, Farelo L, Beltran-Beck B, Real R, Campos R, et al. Range dynamics driven by quaternary climate oscillations explain the distribution of introgressed mtDNA of *Lepus timidus* origin in hares from the Iberian peninsula. J Biogeogr. 2015;42:1727–35.
46. Feder JL, Xie X, Rull J, Velez S, Forbes A, Leung B, et al. Mayr, Dobzhansky, and bush and the complexities of sympatric speciation in *Rhagoletis*. Proc Natl Acad Sci U S A. 2005;102:6573–80.
47. Rosenzweig BK, Pease JB, Besansky NJ, Hahn MW. Powerful methods for detecting introgressed regions from population genomic data. Mol Ecol. 2016;25:2387–97.
48. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4:1073–81.
49. Gu Z, Li J, Gao S, Gong M, Wang J, Xu H, et al. InterMitoBase: an annotated database and analysis platform of protein-protein interactions for human mitochondria. BMC Genomics. 2011;12:335.

Seixas *et al. Genome Biology* (2018) 19:91

Page 19 of 20

50. Calvo SE, Clauser KR, Mootha VK. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. 2016;44:D1251–7.
51. McVean G, Awadalla P, Fearnhead P. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics. 2002;160:1231–41.
52. Auton A, McVean G. Recombination rate estimation in the presence of hotspots. Genome Res. 2007;17:1219–27.
53. Stewart JR, Lister AM, Barnes I, Dalen L. Refugia revisited: individualistic responses of species in space and time. Proc R Soc B Biol Sci. 2010;277:661–71.
54. Coyne JA, Orr H. Speciation. Sunderland, MA: Sinauer Associates; 2004.
55. Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, et al. The genomic landscape of species divergence in *Ficedula flycatchers*. Nature. 2012;491:756–60.
56. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al. Genome-wide evidence for speciation with gene flow in *Heliconius butterflies*. Genome Res. 2013;23:1817–28.
57. Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, et al. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science. 2015;347:1258524.
58. Sankararaman S, Mallick S, Patterson N, Reich D. The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. Curr Biol. 2016;26:1241–7.
59. Carneiro M, Albert FW, Afonso S, Pereira RJ, Burbano H, Campos R, et al. The genomic architecture of population divergence between subspecies of the European rabbit. PLoS Genet. 2014;10:e1003519.
60. Macholán M, Munclinger P, Šugerková M, Dufková P, Bímová B, Božíková E, et al. Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. Evolution. 2007;61:746–71.
61. Payseur BA, Krenz JG, Nachman MW. Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. Evolution. 2004;58:2064–78.
62. Haenel Q, Laurentino TG, Roesti M, Berner D. Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. Mol Ecol. 2018;27:2477–97.
63. Gante HF, Matschiner M, Malmstrøm M, Jakobsen KS, Jentoft S, Salzburger W. Genomics of speciation and introgression in princess cichlid fishes from Lake Tanganyika. Mol Ecol. 2016;25:6143–61.
64. Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, et al. Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science. 2018; https://doi.org/10.1126/science.aar3684.
65. Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL. Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*. PLoS Genet. 2014;10:e1004410.
66. Geraldes A, Basset P, Smith KL, Nachman MW. Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. Mol Ecol. 2011;20:4722–36.
67. Janoušek V, Munclinger P, Wang L, Teeter KC, Tucker PK. Functional Organization of the Genome may Shape the species boundary in the house mouse. Mol Biol Evol. 2015;32:1208–20.
68. Martin SH, Jiggins CD. Interpreting the genomic landscape of introgression. Curr Opin Genet Dev. 2017;47:69–74.
69. Bray Y, Devillard S, Marboutin E, Mauvy B, Péroux R. Natal dispersal of European hare in France. J Zool. 2007;273:426–34.
70. Avril A, Léonard Y, Letty J, Péroux R, Guitton JS, Pontier D. Natal dispersal of European hare in a high-density population. Mamm Biol. 2011;76:148–56.
71. Chan KMA, Levin SA. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution. 2005;59:720–9.
72. Bonnet T, Leblois R, Rousset F, Crochet PA. A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. Evolution. 2017;71:2140–58.
73. Sloan DB, Havird JC, Sharbrough J. The on-again, off-again relationship between mitochondrial genomes and species boundaries. Mol Ecol. 2017;26:2212–36.
74. Dion-Côté A-M, Barbash DA. Beyond speciation genes: an overview of genome stability in evolution and speciation. Curr Opin Genet Dev. 2017;47:17–23.
75. Dowling DK, Nowostawski AL, Arnqvist G. Effects of cytoplasmic genes on sperm viability and sperm morphology in a seed beetle: implications for sperm competition theory? J Evol Biol. 2007;20:358–68.
76. Froman DP, Kirby JD. Sperm mobility: phenotype in roosters (*Gallus domesticus*) determined by mitochondrial function. Biol Reprod. 2005;72:562–7.
77. Yee WKW, Sutton KL, Dowling DK. In vivo male fertility is affected by naturally occurring mitochondrial haplotypes. Curr Biol. 2013;23:R55–6.
78. Smith S, Turbill C, Suchentrunk F. Introducing mother's curse: low male fertility associated with an imported mtDNA haplotype in a captive colony of brown hares. Mol Ecol. 2010;19:36–43.
79. Dowling DK, Friberg U, Lindell J. Evolutionary implications of non-neutral mitochondrial genetic variation. Trends Ecol Evol. 2008;23:546–54.
80. Burton RS, Barreto FS. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? Mol Ecol. 2012;21:4942–57.
81. Levin L, Blumberg A, Barshad G, Mishmar D. Mito-nuclear co-evolution: the positive and negative sides of functional ancient mutations. Front Genet. 2014;5:1–11.
82. Evolution FSA. Mitochondrial burden on male health. Curr Biol. 2012;22:797–9.
83. Van Vranken JG, Bricker DK, Dephoure N, Gygi SP, Cox JE, Thummel CS, et al. SDHAF4 promotes mitochondrial succinate dehydrogenase activity and prevents neurodegeneration. Cell Metab. 2014;20:241–52.
84. Savitha S, Tamilselvan J, Anusuyadevi M, Panneerselvam C. Oxidative stress on mitochondrial antioxidant defense system in the aging process: role of DL-α-lipoic acid and L-carnitine. Clin Chim Acta. 2005;355:173–80.
85. Barreto FS, Burton RS. Elevated oxidative damage is correlated with reduced fitness in interpopulation hybrids of a marine copepod. Proc R Soc B Biol Sci. 2013;280:20131521.
86. Du SNN, Khajali F, Dawson NJ, Scott GR. Hybridization increases mitochondrial production of reactive oxygen species in sunfish. Evolution. 2017;71:1643–52.
87. Racimo F, Marnetto D, Huerta-Sánchez E. Signatures of archaic adaptive introgression in present-day human populations. Mol Biol Evol. 2017;34:296–317.
88. Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. Evidence for archaic adaptive introgression in humans. Nat Rev Genet. 2015;16:359–71.
89. Song Y, Endepols S, Klemann N, Richter D, Matuschka FR, Shih CH, et al. Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. Curr Biol. 2011;21:1296–301.
90. Staubach F, Lorenc A, Messer PW, Tang K, Petrov DA, Tautz D. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (*Mus musculus*). PLoS Genet. 2012;8:e1002891.
91. Pardo-Díaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, Joron M, et al. Adaptive introgression across species boundaries in *Heliconius* butterflies. PLoS Genet. 2012;8:e1002752.
92. Zhang W, Dasmahapatra KK, Mallet J, Moreira GRP, Kronforst MR. Genome-wide introgression among distantly related *Heliconius* butterfly species. Genome Biol. 2016;17:25.
93. Clarkson CS, Weetman D, Essandoh J, Yawson AE, Maslen G, Manske M, et al. Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. Nat Commun. 2014;5:4248.
94. Norris LC, Main BJ, Lee Y, Collier TC, Fofana A, Cornel AJ, et al. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. Proc Natl Acad Sci U S A. 2015;112:815–20.
95. Mendez FL, Watkins JC, Hammer MF. A haplotype at STAT2 Introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. Am J Hum Genet. 2012;91:265–74.
96. Mendez FL, Watkins JC, Hammer MF. Neandertal origin of genetic variation at the cluster of OAS immunity genes. Mol Biol Evol. 2013;30:798–801.
97. Dannemann M, Andrés AM, Kelso J. Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human toll-like receptors. Am J Hum Genet. 2016;98:22–33.
98. Deschamps M, Laval G, Fagny M, Itan Y, Abel L, Casanova J, et al. Genomic signatures of selective pressures and introgression from archaic hominins at human innate immunity genes. Am J Hum Genet. 2016;98:5–21.
99. Nédélec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, et al. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. Cell. 2016;167:657–669.e21.
100. Sams AJ, Dumaine A, Nédélec Y, Yotova V, Alfieri C, Tanner JE, et al. Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. Genome Biol. 2016;17:246.
101. Mancini E, Spinaci MI, Gordicho V, Caputo B, Pombi M, Vicente JL, et al. Adaptive potential of hybridization among malaria vectors: introgression at the immune locus TEP1 between *Anopheles coluzzii* and *A. gambiae* in "far-west" Africa. PLoS One. 2015;10:1–13.

Seixas *et al. Genome Biology* (2018) 19:91

Page 20 of 20

102. Grossen C, Keller L, Biebach I, Croll D. Introgression from domestic goat generated variation at the major histocompatibility complex of alpine ibex. PLoS Genet. 2014;10:e1004438.
103. Hasenkamp N, Solomon T, Tautz D. Selective sweeps versus introgression - population genetic dynamics of the murine leukemia virus receptor *Xpr1* in wild populations of the house mouse (*Mus musculus*). BMC Evol Biol. 2015;15:248.
104. Ullrich KK, Linnenbrink M, Tautz D. Introgression patterns between house mouse subspecies and species reveal genomic windows of frequent exchange. bioRxiv. 2017; https://doi.org/10.1101/168328.
105. Camarda A, Pugliese N, Cavadini P, Circella E, Capucci L, Caroli A, et al. Detection of the new emerging rabbit haemorrhagic disease type 2 virus (RHDV2) in Sicily from rabbit (*Oryctolagus cuniculus*) and Italian hare (*Lepus corsicanus*). Res Vet Sci. 2014;97:642–5.
106. Velarde R, Cavadini P, Neimanis A, Cabezón O, Chiari M, Gaffuri A, et al. Spillover events of infection of Brown hares (*Lepus europaeus*) with rabbit Haemorrhagic disease type 2 virus (RHDV2) caused sporadic cases of an European Brown hare syndrome-like disease in Italy and Spain. Transbound Emerg Dis. 2017;64:1750–61.
107. Lavazza A, Cavadini P, Barbieri I, Tizzani P, Pinheiro A, Abrantes J, et al. Field and experimental data indicate that the eastern cottontail (*Sylvilagus floridanus*) is susceptible to infection with European brown hare syndrome (EBHS) virus and not with rabbit haemorrhagic disease (RHD) virus. Vet Res. 2015;46:1–10.
108. Neves F, Abrantes J, Almeida T, de Matos AL, Costa PP, Esteves PJ. Genetic characterization of interleukins (IL-1α, IL-1β, IL-2, IL-4, IL-8, IL-10, IL-12A, IL-12B, IL-15 and IL-18) with relevant biological roles in lagomorphs. Innate Immun. 2015;21:787–801.
109. Wolf JBW, Ellegren H. Making sense of genomic islands of differentiation in light of speciation. Nat Rev Genet. 2017;18:87–100.
110. Crespi B, Nosil P. Conflictual speciation: species formation via genomic conflict. Trends Ecol Evol. 2013;28:48–57.
111. Carneiro M, Rubin C-J, Di Palma F, Albert FW, Alfoldi J, Barrio AM, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science. 2014;345:1074–9.
112. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17:10.
113. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
114. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25:1754–60.
115. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
116. Breese MR, Liu Y. NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. Bioinformatics. 2013;29:494–6.
117. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
118. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.
119. Sarver BAJ, Keeble S, Cosart T, Tucker PK, Dean MD, Good JM. Phylogenomic insights into mouse evolution using a Pseudoreference approach. Genome Biol Evol. 2017;9:726–39.
120. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. Nat Methods. 2012;9:179–81.
121. Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. Haplotype estimation using sequencing reads. Am J Hum Genet. 2013;93:687–96.
122. Chantry-Darmon C, Urien C, De Rochambeau H, Allain D, Pena B, Hayes H, et al. A first-generation microsatellite-based integrated genetic and cytogenetic map for the European rabbit (*Oryctolagus cuniculus*) and localization of angora and albino. Anim Genet. 2006;37:335–41.
123. Kimura M. Evolutionary rate at the molecular level. Nature. 1968;217:624–6.
124. Nei M. Molecular evolutionary genetics. New York: Columbia University Press; 1987.
125. Matthee CA, Van Vuuren BJ, Bell D, Robinson TJ, Sullivan J. A molecular Supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. Syst Biol. 2004;53:433–47.
126. Marboutin E, Peroux R. Survival pattern of European hare in a decreasing population. J Appl Ecol. 1995;32:809–16.
127. Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, et al. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. Cell. 2014;157:785–94.
128. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
129. 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.
130. Pease J, Rosenzweig B. Encoding data using biological principles: the multisample variant format for Phylogenomics and population genomics. IEEE/ACM Trans Comput Biol Bioinforma. https://doi.org/10.1109/tcbb.2015.2509997
131. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. G:profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. 2007;35:193–200.
132. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, et al. G: profiler—a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 2016;44:W83–9.
133. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011;6:e21800.
134. Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database. 2011;2011:bar030.
135. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. EGGNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2016;44:D286–93.
136. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.
137. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8.
138. Levan A, Fredga K, Sandberg AA. Nomenclature for centromeric position on chromosomes. Hereditas. 1964;52:201–20.
139. Standard karyotype of the laboratory rabbit, *Oryctolagus cuniculus*. Cytogenet Cell Genet. 1981;31:240–8.
140. Seixas FA, Boursot P, Melo-Ferreira J. The genomic impact of historical hybridization with massive mitochondrial DNA introgression. NCBI SRA. BioProject Accession: PRJNA399194. Available from: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA399194/. Accessed 11 June 2018.
141. Carneiro M, Rubin C-J, Di Palma F, Albert FW, Alfoldi J, Barrio AM, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. NCBI SRA. Accession Number: SRX545993. Available from: [https://www.ncbi.nlm.nih.gov/sra/SRX545993%5baccn]. Accessed 11 June 2018.
142. Carneiro M, Rubin C-J, Di Palma F, Albert FW, Alfoldi J, Barrio AM, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Ensembl. Available from: ftp://ftp.ensembl.org/pub/release-80/fasta/oryctolagus_cuniculus/dna/. Accessed 11 June 2018.
143. Mitchell-Jones AJ, Amori G, Bogdanowicz W, Krystufek B, Reijnders PJH, Spitzenberger F, et al. The atlas of European mammals. London: Academic Press; 1999.