Elsevier Editorial System(tm) for Molecular Phylogenetics and Evolution
Manuscript Draft

Manuscript Number: MPE-14-24R1

Title: The shared mitochondrial genome of Rupicapra pyrenaica ornata and Rupicapra rupicapra cartusiana: Old remains of a common past

Article Type: Short Communication

Keywords: chamois, mtDNA, phylogeny, introgression

Corresponding Author: Dr. Ana Domínguez, PhD

Corresponding Author's Institution: Universidad de Oviedo

First Author: Trinidad Pérez

Order of Authors: Trinidad Pérez; Iago Gonzalez; Sabine E Essler; Margarita Fernández; Ana Domínguez, PhD

Abstract: Mitochondrial DNA (mtDNA) has largely been used for species delimitation. However, mtDNA introgression across species boundaries can lead to inconsistent phylogenies. Partial sequences of the mitochondrial genome in the chamois, genus Rupicapra, show the presence of three well differentiated clades, West (mtW), Central (mtC) and East (mtE), each with a geographically restricted distribution. The complete mtDNAs of the clades mtW and mtE (main representatives of the two currently considered species R. pyrenaica and R. rupicapra respectively) have been reported. In the present study, we sequenced the clade mtC present in populations from both species inhabiting the central area of Europe: the Apennines (R. pyrenaica ornata) and the Chartreuse Mountains (R. rupicapra cartusiana). The phylogenetic comparison with the genomes of Caprini, ,highlights the ancient presence of chamois in Europe relative to the fossil record, and the old age of the chamois clade mtC that was split from the clade mtW in the early Pleistocene. The separation of R. pyrenaica ornata and R. rupicapra cartusiana female lineages was recent, dating of the late Pleistocene. Our data represent an example of mtDNA introgression of resident females of Chartreuse Mountains into immigrant males of R. rupicapra due to male-biased migration and female phylopatry.

Ana Domínguez
sanjurjo@uniovi.es

Molecular Phylogenetics and Evolution
Elsevier

Reply letter Oviedo, 5-26-2014

Manuscript: MPE-14-24

The shared mitochondrial genome of *Rupicapra pyrenaica ornata* and *Rupicapra rupicapra cartusiana*: Old remains of a common past

Dear Sirs,

The authors thank for the reviewers' comments and ideas for improvement of the paper. We appreciated the helpful suggestions by the reviewers and hope to fulfill all comments satisfactorily. In particular, we highlighted the novelty and interest of the paper as requested by the Associated Editor.

Following are the reviewers' comments followed by our answers and the changes/adaptations that we have made to the manuscript. Referenced line numbers correspond to the revised version of the manuscript.

We look forward to hearing from you.

Yours sincerely,

Ana Domínguez

Ms. No. MPE-14-24

The shared mitochondrial genome of *Rupicapra pyrenaica ornata* and *Rupicapra rupicapra cartusiana*: Old remains of a common past

Comments to the reviews

REVIEWER #1

1) The discussion includes an important and highly interesting paragraph (lines 196-213), where the authors explain their hypothesis regarding the mechanism of mtDNA replacement in Rupicapra, with reference to other genetic markers and biogeography. This section falls a little short in my view, and it could in part be improved in terms of the flow of thoughts/logic.

- Thank you for pointing this. We extended this section of the discussion trying to improve the flow of ideas (L210-L229).

2) Dating of recent (< ca 2. My) evolutionary events based on mtDNA is nowadays acknowledged to benefit from inclusion of not only external calibration points, but also some internal calibration. Because here only external calibration points are available, it seems most likely that the divergence times inferred for Rupicapra in this manuscript are an overestimation (see references by, e.g., Simon Ho and colleagues). This should be briefly discussed.

- We agree in that the estimation of divergence times suffers from calibration uncertainty and this was now addressed in our previous version of the manuscript. Now we included a cautionary paragraph (L197-199) to take into account the time dependency of molecular rate estimates and its effect on estimation of recent divergence times that was reported in Ho et al. 2005. In addition, we have chosen the most adequate calibration scenario as was suggested by Rev #2 instead of considering three calibration scenarios as in our first vs of the manuscript.

3) Despite the brevity of the manuscript (short communication), the text contains detailed analyses of variability and signatures of selection on Rupicapra mtDNA, which is not mentioned in the abstract, and which comes across as a side track, considering the main focus of the manuscript: introgression of mtDNA. The findings are certainly interesting and worth presenting in a paper, but in case the review process should reveal aspects of the main message of the paper that need to be developed in more detail, the aspect of variability and selection could be shortened down.

- After introducing the suggestions requested, we did not surpassed the length limit of the manuscript and consequently we did not delete this part.

4) While the manuscript is written in a mostly concise and convincing way, my impression (a non-native speaker myself) is that the language should be checked throughout for clarity and correctness. Besides grammar and other language issues, I found that - especially in the results/discussion - the wording is not clear when using "this result" or "these findings" (e.g., lines 172-192). When reading I had to check carefully whether such sentences referred to results from the present manuscript, or whether the authors are referring to findings from previous studies. This should be checked throughout - but will not require major changes.

- The language usage was checked thorough by Sara de Albornoz (Official English Translator) and by the native speaker Steve Smith, PhD, currently Head of Genetics at the Department of Integrative Biology and Evolution at the University of Veterinary Medicne Vienna. We have rewritten the paragraph signalled to improve clarity.

5) Line 112: "using different calibration scenarios and _a_ normal distribution prior": Based on Table 2 I assume that each BEAST analysis used two calibration points, one for the Bovidae divergence, and one for the diversification of Caprini. Line 112 should therefore be clarified. Also, Table 2 should state that the assumed values are mean and S.D. of the normally distributed priors.

- After a suggestion of the Rev. #2, we have evaluated the prospect of the different calibration points and selected the one that we find the more appropriate (please see our answer to comment nº 53).

6) L114: you mean a 10% burn-in was used for preliminary visualization in Tracer? From lines 113-117 it seems that for the presentation of your phylogenetic inferences (in TreeAnnotator; visualized in Figure 2) you are only using a burn-in of ca. 4% (discard 1,000 trees out of a total of 25 million:1,000=25,000 trees).This should be clarified, and you should verify whether indeed a burn-in of 4% (if true) was large enough in TreeAnnotator.

- This was a mistake, we used a burn-in of 10% (2,500 trees) in TreeAnnotator.

7) L133: maybe say "p distances" rather than "p values", to avoid confusion that this might be a likelihood, not a genetic distance?

- Of course, "p distances" is much more appropriate. We changed this thorough the manuscript.

8) L161: "direct repeat" was a little unclear to me. You mean that it is a perfect repeat, the repeat units are 100% identical?

- Yes is a perfect repeat of 10 nt. This was changed to read "perfect direct repeat" (L165).

9) L170: should be "amino acid" (with space)

- This was corrected

10) L175-178: consider explaining more clearly what "replacements" (presumably a finding of Hassanin et al.?) you are referring to, and what the significance of these findings is for your study/question. I could not quite follow this reasoning.

- We reworded this paragraph to make it clearer. Now it reads as "*Hassanin et al. (2009) have identified several amino acid replacements that are diagnostic of the different clades of Caprini. We found the molecular signatures of Rupicapra in the mitochondrial sequences of the subspecies R. p. ornata and R. r. cartusiana. The diagnostic replacement 61S->P in ATP8 appears to be erroneous in Hassanin et al. (2009) however, as this substitution does not appear in any of the sequences of Rupicapra*" (L178-182).

11) L181f: The sentence starting with "The analysis of complete mitochondrial..." should be edited for clarity.

- We have rewritten this sequence that now reads as "*The analysis of complete mitochondrial sequences, in the present study, demonstrates a closer relationship of the clade mtC to Iberian chamois (R. p. pyrenaica) than to the Eastern chamois (R. r. rupicapra)*." (L185-188)

12) L191: "... or at most since the Early Pleistocene" - consider re-writing this sentence for improved clarity.

- We have changed the paragraph to " *… or, at least, since the Early Pleistocene*" (L196).

13) Table 1: Just to make sure, should the second last line be pN/pS, or pS-pN?

- It is pN/pS that we consider more meaning that pS-pN

14) Figure 2: Consider showing the calibration points, and marking the Caprini in some way (on the tree, or as a bracket on the right of the tree)?

- We have signaled the calibration points on the tree and explained in the caption their correspondence to Bobidae and Caprini.

REVIEWER #2

15) L18: are mtW, mtC, and mtE each restricted geogrphically or all within a restricted dist?

We changed the wording to "each with a restricted distribution" (L19-20).

16) L25: Early Pleistocene does not sound very old. Especially if we consider that Caprini is about 10my old.

- We reworded the paragraph that now reads as *"the ancient presence of chamois in Europe, relative to the fossil record,.."* (L26).

17) L28: is this explicitly tested in your analysis? how?

- This is not explicitly tested in our analysis, but its occurrence was already argued in our paper (Rodriguez et al. 2010) from the comparison of mtDNA markers and microsatellites. Briefly, the individuals from Chartreuse belong to the clade central with posterior probability equal to 1 while they group with Eastern chamois for microsatellites (assignment proportion =1 in STRUCTURE analysis).

18) Keywords. L33: Avoid words that are already used in the title. I think "taxonomy" is not an appropriate keyword for this paper. How is "introgression" tested with your dataset?

- We have omitted the words that are already in the title and have cut out "taxonomy". As explained in our answer to comment (17) we have not tested introgression from this dataset but was detected from the comparison of mitochondrial and nuclear patterns of variation previously described (L33).

19) Introduction. Overall: You may include some background information about the use of mitogenomes in your type of study. Also consider adding some background on introgression and link it to your discussion. Clearly state your hypotheses, objectives, or questions to be tested.

- As suggested, we added a statement on the use of mtDNA in this kind of studies and the features associated to introgression (L36-39). In addition, we explained that the resolution of our previous phylogenetic analysis on mtDNA markers was poor (L51-53) and stated more clearly the aims of the present work at the end of the introduction (L54-59)

20) L38: You can split this into a couple of sentences to improve clarity.

The paragraph was split (L39-44) and now reads as "*Populations of chamois are classified into two species, R. pyrenaica and R. rupicapra, in most modern taxonomic revisions. The species R. pyrenaica comprises the populations from southwestern Europe, including the subspecies parva, pyrenaica and ornata. The species R. rupicapra covers the populations from northeastern Europe and contains seven subspecies: cartusiana, rupicapra, tatrica, carpatica, balcanica, asiatica and caucasica*".

21) L47: "Mitochondrial DNA is frequently used to delimit species…" Citation?

- The citation was included (L36).

22) L47-48: Revise punctuation

- The paragraph was rewritten as *"The clade mtC is shared by the two species of the genus: R. pyrenaica ornata in the Apennines and R. rupicapra cartusiana in the Massif of Chartreuse."* (L49-51).

23) L48-49: This is the result of your paper... How does having a full mt-genome make things different?

- The use of partial mtDNA sequences did not allow phylogenetic resolution due to rapid radiation of the three main clades. The node joining clade C and W formed in the Bayes, NJ and MP analysis with low support (bootstrap around 50%) and the ML analysis rendered a node joining C and E. With the complete mtDNA sequences the phylogeny relationships were resolved.

24) L49-53: this is a bit difficult to follow. consider rewording.

- We have reworded the paragraph to *"The divergence between these major clades happened in a short period of time, which precluded the resolution of phylogenetic relationships from the comparison of partial sequences (1,646 nucleotides) of the mitochondrial genome (Rodríguez et al., 2010)"* (L51-53).

25) L50: "… 1646 nucleotides …" mitochondrial or nuclear genome?

- We now explain that the previous study was based on mtDNA (L53)

26) L54-59: What is the hypothesis or objective of this study? Please state it clearly.

- As suggested, we now state more clearly the objective of our work. The new paragraph at the end of the introduction (L54-59) reads as follows: *"Here we sequenced the complete mitochondrial genome of two individuals of the clade mtC, one from the Apennines (R. pyrenaica ornata) and one from the Chartreuse Mountains (R. rupicapra cartusiana) and compare them to the mitochondrial genomes of the Tribe Caprini (Hassanin et al., 2009) with the aim of clarifying the phylogenetic relationships of the mitochondrial lineages of Rupicapra and relating them to the climatic changes of the Quaternary in Europe and the dispersion events that lead to mtDNA introgression."*

27) L57: This sentence is unclear. Where were these calibration points used? In this study? In the literature?

- Following your suggestion (please, see our answer to comment nº 53) we now focus on one calibration scenario, consequently this paragraph has been deleted.

28) Methods. Overall: Reorganizing the methods into discrete subsections might help with the organization of this part of the manuscript. Examples of subsections might be: Taxon Sampling and Data Collection, Sequence Alignment and Phylogenetics, and Estimates of Divergence.

- We reorganized the methods into two discrete subsections as recommended.

29) L75: "Sequencher 4.9" manufacturer name?

- Done (L76)

30) L76: either replace the word "programme" for "software" or delete it altogether. "BioEdit 3.2" citation?

- "alignment programme" was corrected to "alignement editor" and the citation of BioEdit was included (L78).

31) L82: "Mega 5.2" citation?

- Done (L82).

32) L85: "N-G model" citation?

- Done (L86)

33) L90: delete comma after "… by Hassanin et al (2009)"

- Done

34) L94-96: as it is written, this sentence states that you used a model of nucleotide substitution for MP, which is not true. please reword

- Thanks for spotting this. The paragraph was reworded to *"Neighbor-Joining (NJ), Maximum Parsimony (MP), Maximum-Likelihood (ML) or Bayesian approaches were used to construct phylogenetic trees."* (L97-99).

35) L100-102: did you explore various models of substitution based on multiple partitions?

- Yes, we explored the optimal substitution model for the different datasets: tRNAs and other non-coding sequences, the rRNAs, the coding sequences and the CR.

36) L104: posterior probabilities is not capitalized

- Corrected (L106).

37) L108: "… GTR+G+I …" Some authors argue that models including both gamma and a proportion of invariable sites should not be used (Ren et al. 2005. Syst. Biol 54:808-818; Yang Z. 2006. Computational molecular evolution). A gamma distribution already allows for sites with very low rates. As a result, adding a proportion of invariable sites creates a strong correlation, making it difficult to estimate both parameters reliably.

We have done an extensive search about this interesting topic on the scientific literature. In fact authors as Sullivan and Swofford, 2001 and Yang 2006 have exposed the problems associated when trying to estimate both the proportion of invariant sites and the sites with very low rates of variation. However, there is discussion about this topic, and other authors (Waddell, 2005) defend to use it as any other model if the model fits the data. Hence, and giving that the different phylogenetic analysis software maintain this model and do not alert of problems of using it, we have decided to continue using this model as it is the one that best fits our data. In any case, we did use also the simpler model of Jukes Cantor and also parsimony free analysis. The other models that best fit the data also include invariant (TN93+G+I and HKY+G+I), the proportion of invariant sites obtained is 0.56. To check the effect of including invariant sites or not, we have performed the analysis in Beast using the GTR+G model and we obtained the same topology and slightly larger divergence times.

38) L108: "AIC" It says BIC in L101

- Thanks for pointing that. We used AIC and corrected the text accordingly (L103).

39) L108: "… with the empirical…" I don't think the rest of the sentence is necessary.

- The rest of the sentence was deleted and wrote "*with the empirical base frequencies*" (L111).

40) L111: please indicate which of these are fossil vs. molecular estimates for calibration

- Following your suggestion (comment nº 53) we presented only one calibration scenario that is based on fossils, now explained in L112-119.

41) L112: "Table 3" is not part of this manuscript. Perhaps you mean Table 2?

- Yes we wanted to mean Table 2. However, after the revision regarding the calibration scenarios, this table was deleted.

42) L112: I am not convinced that a 'normal distribution prior' for all your calibrated nodes or "calibration scenarios" is appropriate for your study. For example, Bibi (2013) assumed the age of the crown clade Bovidae to be 18 my. However, the original estimate was based on fossil evidence of Eotragus sp. obtained by Solouinas et al. (1995) J. Vert. Paleo.

What are the other calibration estimates based on? Molecules? Other fossils?

Please consider what a normal distribution prior means in terms of the type of calibration you have and contrast this with other types of priors (see Ho and Phillips 2009; Syst. Bio. for a good review on prior selection for divergence dating.)

- We appreciate your suggestions regarding the calibration scenarios and the problems associated to the election of priors. After reading carefully the revision by Ho and Philips, we decided to present the calibration scenario based on Bibi, that we found the most appropriate. In addition we now explain in Mat. and Met. the reasons behind the use of normal priors as soft bounds as explained by Bibi. He says that in the calibration points of crown Bovidae and crown Caprini, the normal distribution is appropriate because the phylogenetic position of the fossil is not precisely known relative to the clade in question (see L112-119).

43) L115: "Tracer 1.5" citation? Also, delete "within the BEAST package". "the resulting samples" are trees or generations? Please be specific

- Done (L122).

44) L116-117: You can simplify this sentence by changing the sentence structure and choosing words a bit more carefully

- We changed the wording to "*The Maximum clade credibility criterion tree was obtained with TreeAnnotator using a burn-in of 2,500 and with mean node heights*" (L122-123).

45) Results L133-135: what kind of genetic distance was used for these estimates?

- The distance was the observed number of substitutions per nucleotide (L81).

46) L136: "low" is a relative term. if you calculated these differences, what are the values?

- The distances were included in L143.

47) L140-141: You ran 4 different analyses. Did all of them result in identical topologies? Often researchers report the ML and Bayes trees in their publications. Especially if you are showing node support as well as estimates of divergence. This could help clean up your results to make them easier to understand.

- The four different analyses have rendered identical Rupicapra topologies (L149-150). We now included the ML tree as supplementary material.

48) L143: "Table 2" I don't think this reference is appropriate here. Table 2 does not show evidence of support.

- This was a mistake. Thanks for pointing it out. After omitting the analysis based on different calibration scenarios (please, see our reply to comments nº40 and nº53) we did include the appropriate table.

49) L143: "… most external node…" in relation to what?

- We omitted "external node" and now the paragraph reads as: *"Phylogenetic analysis of partial datasets always grouped the node joining the two sequences of the Rupicapra clade mtC (ornata and cartusiana) with the maximum support* (L150-151)*."*

50) L146: Supplementary materials were not available for me to review

- The table given as Supp. Mat. in the first version of the manuscript is now given as Table 2. New supplementary material was provided as specified in the answers to comments nº47 and nº 59.

51) L149: "… loosely supported…" These are very low values; I would indicate them as "poorly supported" or "not supported".

- Accordingly changed to "*poorly supported*" (L155).

52) L152-154: "The calculated divergence times […] to calibrate the tree …" I am confused by the obvious nature of this statement. You tested 3 calibration scenarios and then concluded that (i.e. see lines 152-154). Alternatively, if you knew that a priori, why test all of these calibration scenarios?

- After your suggestion we used only one calibration point (see our answer to comments nº 42 and nº 53).

53) L152-155: I am not really sure of the reasoning for using 3 different "calibration scenarios". Choosing calibration points for divergence dating is not a trivial matter. You must consider your calibration points carefully based on evidence rather than test multiple "calibration scenarios" without any reason to back it up. In the end, you only show one scenario anyway in Figure 2. So, what is the point of doing the others?

- We agree, and now focus on the most appropriated calibration based on the fossil record and consequently Table 2 was deleted (please, see our answer to comment nº 42).

54) Discussion. Overall: I am unsure about how the mitogenomic analysis informs what was already known about the relationships of these populations. I think that a few carefully done tests could help you test for introgression instead of just saying that there is introgression without testing it. This could even work with a smaller sequence dataset but with more sampled individuals (perhaps like the ones presented in the Rodríguez et al 2010). See an example of this in Marshall et al. 2011 Syst. Bio. Also, consider bringing your discussion further. For example: Other than producing a better supported phylogenetic tree, how does a mitogenomic analysis help us understand the bigger picture of your study? Were your estimates of divergence any different than the previous estimates by Bibi (2013)? How could this type of analysis inform other relationships within Caprini?

- The previous analysis of partial mtDNA sequences precluded the resolution of phylogenetic relationships between clades of *Rupicapra* (please, see our answer to comment nº 17). Regarding the test of introgression, we now included a statement in the discussion (L204-211) to highlight the fact that the populations of the Massif of Chartreuse and the Apeninnes share an mtDNA lineage that diverged from the lineage of the Eastern chamois (Rupicapra) almost 2 mya, however the individuals from Chartreuse belong to the Eastern chamois without ambiguity for microsatellites, for the melacortin 1 receptor, for which only three haplotypes were

found and the chamois of Chartresuse share the haplotype with the eastern chamois, and for the Y-chromosome where the two haplotypes identified in Chartreuse were also sampled from the R. rupicapra and differ from the haplotypes sampled from R. pyrenaica.

We relate our phylogeny to that found in previous studies by Hassanin et al., 2012 and Bibi, 2013 (L190-193) and reworked the last part of the discussion to broaden the implications of our study (L220-229).

55) L174-176: I am unsure of what this sentence means. Please clarify.

- The sentence was reworded to improve clarity. Now it reads as "*Hassanin et al. (2009) have identified several amino acid replacements that are diagnostic of the different clades of Caprini. We found the molecular signatures of Rupicapra in the mitochondrial sequences of the subspecies R. p. ornata and R. r. cartusiana. The diagnostic replacement 61S->P in ATP8 appears to be erroneous in Hassanin et al. (2009) however, as this substitution does not appear in any of the sequences of Rupicapra*" (L178-182).

56) Table 1: I think this table is fine.

57) Table 2: Please write a better caption for this table. It would be nice to specify which calibrations are fossils and which are secondary calibrations.

- After your suggestion this table was deleted (please, see our answer to comment nº 53).

58) Figure 1: Please write a better caption for this figure. Make the Y axis label less criptic.

- Done, the caption was rewritten (L237-241) and the Y-axis label extended.

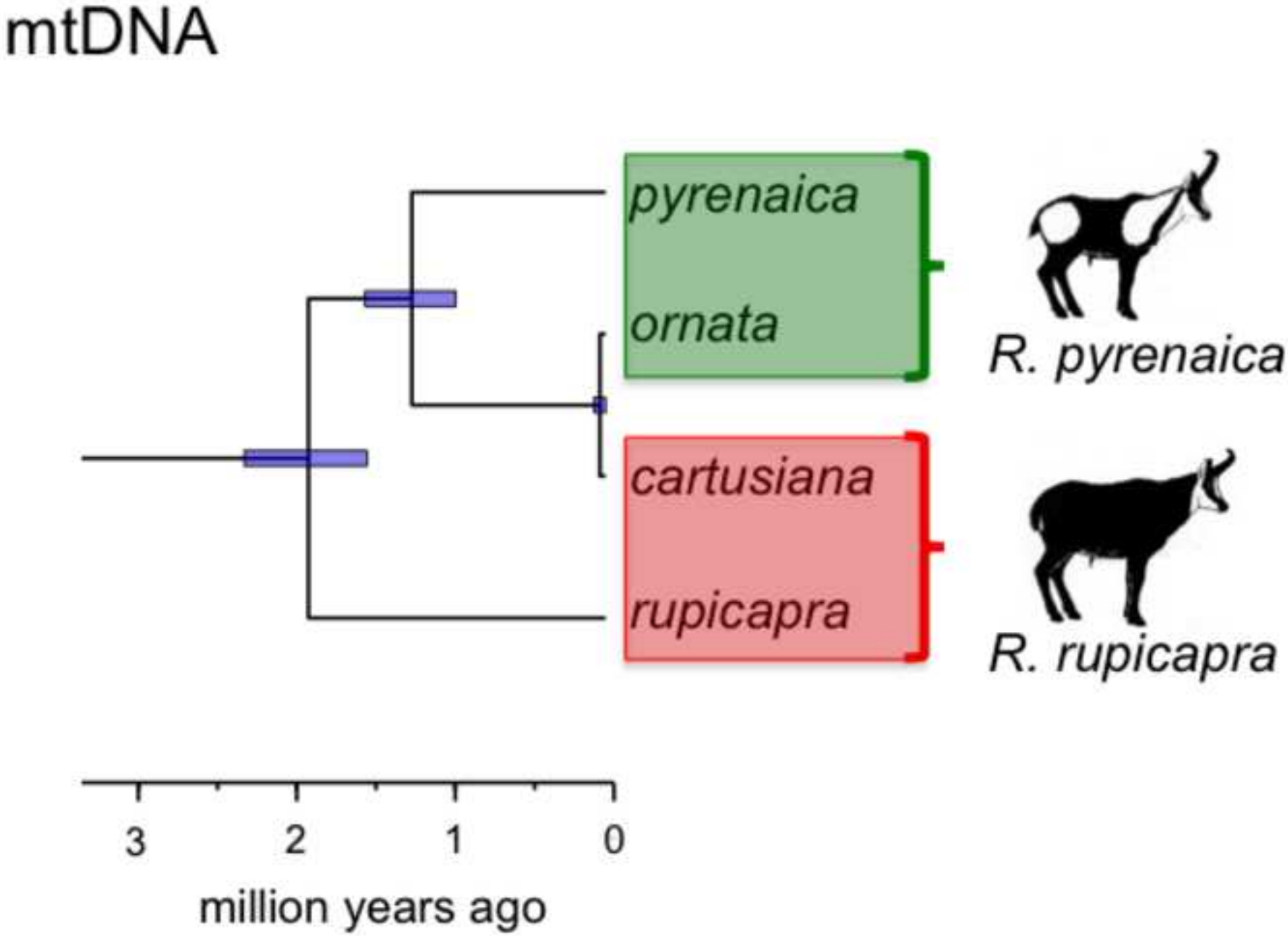
59) Figure 2: Please write a better caption for this figure. I think that the GenBank accession numbers are better shown as a table in the supplementary materials. Remove these from the caption. Overall, I am disappointed with this figure. There is very little information that can be gathered by looking at this tree.

- Following your suggestions we have rewritten the caption to this figure. The GenBank accession numbers were deleted and are provided as a table in the supplementary materials. Additional information was included in the figure as suggested in comment nº 60.

60) Consider adding the ML tree with node support for comparison. The BEAST tree could be better labeled with: 1. The calibration point(s), 2. Which group is W vs. C vs. E, 3. Specific 95% HPDs for the estimated nodes (instead of the bars), 4. A time scale bar, and 5. A description of what the scale bar at the bottom represents. Perhaps only show dates only for important nodes that you discuss to draw attention to the research question.

- We have added the ML tree as supplementary material given that the number of figures is limited to two for "Short communications". The calibration points, rupicapra clades and time scale bar were included in the figure. However, for the 95% HPDs we think the bars are easier to see in the figure. In addition, the phylogenetic tree will be included should the ms be accepted and this will allow alternate presentation forms to the readers. The clades mtW, mtC and mtE are now indicated in the figure.

*Graphical Abstract (for review)



***Highlights (for review)**

Highlights

1-The comparison of the three mitochondrial lineages of chamois highlights its presence in Europe since the early Pleistocene

2- The mtDNA phylogeny of chamois does not concur with its taxonomy

3- The mtDNA of resident females of Chartreuse Mountains was introgressed into immigrant males

*Manuscript
Click here to view linked References

The shared mitochondrial genome of *Rupicapra pyrenaica ornata* and *Rupicapra rupicapra cartusiana*: Old remains of a common past

Trinidad Pérez [a], Iago González, Sabine E. Essler [b], Margarita Fernández [a], Ana Domínguez [a*]

a: Universidad de Oviedo, Biología Funcional (Genética). Julián Clavería 6, 33071 Oviedo, Spain.

b: University of Veterinary Medicine Vienna, Department of Pathobiology, Institute of Immunology, Veterinaerplatz 1, A-1210 Vienna, Austria.

*: Corresponding author. Tel: (34)985102719; Fax: (34)985103534.

E-mail address: sanjurjo@uniovi.es (A. Domínguez)

Abstract

Mitochondrial DNA (mtDNA) has largely been used for species delimitation. However, mtDNA introgression across species boundaries can lead to inconsistent phylogenies. Partial sequences of the mitochondrial genome in the chamois, genus *Rupicapra*, show the presence of three well differentiated clades, West (mtW), Central (mtC) and East (mtE), each with a geographically restricted distribution. The complete mtDNAs of the clades mtW and mtE (main representatives of the two currently considered species *R. pyrenaica* and *R. rupicapra* respectively) have been reported. In the present study, we sequenced the clade mtC present in populations from both species inhabiting the central area of Europe: the Apennines (*R. pyrenaica ornata*) and the Chartreuse Mountains (*R. rupicapra cartusiana*). The phylogenetic comparison with the genomes of Caprini, ,highlights the ancient presence of chamois in Europe relative to the fossil record, and the old age of the chamois clade mtC that was split from the clade mtW in the early Pleistocene. The separation of *R. pyrenaica ornata* and *R. rupicapra cartusiana* female lineages was recent, dating of the late Pleistocene. Our data represent an example of mtDNA introgression of resident females of Chartreuse Mountains into immigrant males of *R. rupicapra* due to male-biased migration and female phylopatry.

Keywords: chamois, mtDNA, phylogeny, introgression

1. Introduction

Mitochondrial DNA is the marker of choice for species delimitation (Avise, 2004). However, mtDNA introgression across species boundaries has been repeatedly documented in mammals, leading to inconsistent phylogenies (Petit and Excoffier, 2009). In chamois (genus *Rupicapra*), mtDNA phylogenies do not concur with taxonomic classification. Populations of chamois are classified into two species, *R. pyrenaica* and *R. rupicapra,* in most modern taxonomic revisions. The species *R. pyrenaica* comprises the populations from southwestern Europe, including the subspecies *parva*, *pyrenaica* and *ornata*. The species *R. rupicapra* covers the populations from northeastern Europe and contains seven subspecies: *cartusiana*, *rupicapra*, *tatrica*, *carpatica*, *balcanica*, *asiatica* and *caucasica*. However, the taxonomy of the genus has been subject of continuous revisions during the twentieth century (Corlatti et al., 2011; Groves and Grubb, 2011; Valdez, 2011). Phylogenetic studies based on mtDNA (Crestanello et al., 2009; Rodríguez et al., 2009; Rodríguez et al., 2010) showed the existence of three old clades that are distributed following a clear geographical pattern and were referred to accordingly by Rodríguez et al. (2009) as west, central and east (mtW, mtC and mtE). The clade mtC is shared by the two species of the genus: *R. pyrenaica ornata* in the Apennines and *R. rupicapra cartusiana* in the Massif of Chartreuse. The divergence between these major clades happened in a short period of time, which precluded the resolution of phylogenetic relationships from the comparison of partial sequences (1,646 nucleotides) of the mitochondrial genome (Rodríguez et al., 2010).

Here we sequenced the complete mitochondrial genome of two individuals of the clade mtC, one from the Apennines (*R. pyrenaica ornata*) and one from the Chartreuse Mountains (*R. rupicapra cartusiana*) and compare them to the mitochondrial genomes of the Tribe Caprini (Hassanin et al., 2009) with the aim of clarifying the phylogenetic relationships of the mitochondrial lineages of *Rupicapra* and relating them to the climatic changes of the Quaternary in Europe and the dispersion events that lead to mtDNA introgression.

2. Materials and methods

2.1 Data Collection, Sequencing and Alignment

We obtained the complete mitochondrial genome of one specimen of *R. p. ornata* (ANo01) and one of *R. r. cartusiana* (CHAv02) using the set of twenty-three primer pairs published by Hassanin et al. (2009). The DNA of the two samples had been previously isolated (Pérez et al., 2002; Rodríguez et al., 2009). Amplifications were done in a final volume of 20 µl containing 2 µl (≈ 40–70 ng) DNA, 0.75 µM of each primer, 1x PCR Buffer, 250 µM of each dNTP, 2.5-3.75 mM $MgCl_2$ and 1 U of Taq DNA polymerase (Qiagen, Hilden, Germany). Amplification was carried out in PE GeneAmp PCR 9700 thermal cycler (Applied Biosystems, Foster City, CA) with an initial step of 3 min at 94 ºC, 30-35 cycles of 40 s at 94 ºC, 40 s at 50-60 ºC and 40 s at 72 ºC, followed by 5 min at 72 ºC. PCR products were electrophoresed along with size standards in 2% agarose gel in 1x Tris–borate–EDTA and visualized by UV. The PCR-amplified products were purified with the Exo-SAP-IT kit (USB Corporation, Cleveland, OH). Both strands of all the PCR products were sequenced with PCR primers and the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Sequencing products were purified with isopropanol

precipitation and sequenced in an ABI 310 Genetic Analyzer (Applied Biosystems). The sequence data were analyzed and assembled using Sequencher 4.9 (Gene Codes Corp., Ann Arbor, MI).

The sequences were aligned using the multiple alignment editor BioEdit 3.2 (Hall, 1999) and manually checked and edited. The features of the sequences were identified by comparison to the mitochondrial genomes of *R. pyrenaica* (Acc. Nº FJ207538.1) and *R. rupicapra* (Acc. Nº FJ207539.1). The number of substitutions per nucleotide (p) between pairs of sequences of *Rupicapra* for the different genes was computed with MEGA 5.2 (Tamura et al., 2011). The datasets corresponding to the 13 coding genes were analyzed to test for selection operating at the amino acid sequence level. The overall mean distance for synonymous substitutions per synonymous site (pS) and non-synonymous substitutions per non-synonymous site (pN) were computed using the Nei-Gojobori model (Nei and Gojobori, 1986). The difference pS-pN was estimated in MEGA 5.2 using the Nei-Gojobori method, its variance was computed using 1000 bootstrap replicates and the signification of the difference from zero (null hypothesis of strict-neutrality) tested by a Z-test.

2.2 Phylogenetics and Estimates of Divergence

The phylogenetic relationships of these sequences with the mitochondrial sequences of the Caprinae studied by Hassanin et al. (2009) (GenBank Acc. Nº in Supplementary Material 1) were investigated using the sequence of *Bos taurus* as an outgroup. Sequences were arranged in five datasets including the tRNAs and other non-coding sequences, the rRNAs, the coding sequences, the control region (CR), after excluding the repeats at 5' to ensure homology, and the complete mtDNA (the repeats at 5' of the CR were excluded). Neighbor-Joining (NJ), Maximum Parsimony (MP), Maximum-Likelihood (ML) or Bayesian approaches were used to construct phylogenetic trees. A Neighbor-Joining (NJ) tree based on Jukes-Cantor distance was constructed with MEGA 5.2. The topology of the tree was further investigated by model free Maximum Parsimony (MP) as implemented in MEGA 5.2. The MP tree was obtained using the Tree-Bisection-Reconnection algorithm with search level 3 in which the initial trees were obtained with the random addition of sequences (10 replicates). The optimal substitution model was determined with MEGA 5.2 using the Akaike Information Criterion (AIC) and was used to obtain a Maximum Likelihood (ML) tree with the Heuristic Method of the Nearest-Neighbor-Interchange. The reliability of the nodes was assessed by 1000 bootstrap replicates (Felsenstein, 1985) under NJ, MP and ML and by the posterior probability (BPP) of the nodes under the Bayesian approach. Bayesian analysis was conducted using the Monte Carlo Markov chains (MCMC) method implemented in BEAST 1.7 (Drummond and Rambaut, 2007). We used a lognormal relaxed clock with uncorrelated rates and a Yule speciation process as priors. The model of nucleotide substitution was the GTR+G+I as determined by the AIC criteria in MEGA 5.2, with the empirical base frequencies. We used a UPGMA starting tree. Divergence times were estimated with BEAST 1.7, using two calibration points based on the fossil record and using soft bounds to account for uncertainty (Ho and Philips, 2009). We used the ages and prior probability distributions given in Bibi (2013). The two calibration points were crown Bovidae with

a normal prior (mean = 18 Ma, standard deviation = 1 Ma), based on *Eotragus noyei*; and crown Caprini with a normal prior (mean = 8.9 Ma, standard deviation = 2 Ma) based on *Aragoral mudejar* (see Additional File 1 in Bibi, 2013). Following Bibi (2013), using a normal distribution prior is appropriate when the phylogenetic position of the fossil relative to the clade in question is not precisely known. No monophyletic constraints were used. All the analyses were run for 25 million generations with tree and parameter sampling every 1,000 generations. A burn-in of 10% was used and the convergence of all parameters assessed using the software TRACER 1.5 (Rambaut and Drummond, 2009). The Maximum clade credibility criterion tree was obtained with TreeAnnotator, using a burn-in of 2,500 and with mean node heights. All trees obtained by the different methods were visualized with FigTree 1.4 (Rambaut, 2006).

3. Results

The complete mitochondrial genomes of *R. pyrenaica ornata* and *R. rupicapra cartusiana* were determined and deposited in the GenBank with accession numbers KJ184173 and KJ184175, respectively. The total lengths of the sequences were 16,399 bp and 16,398 for *ornata* and *cartusiana,* respectively, with the standard composition of 13 protein-coding genes, 22 tRNAs, 2 rRNAs, the replication origin of the light strand and the CR. The overall nucleotide composition of the H strand was 26% T, 27% C, 33% A and 14% G, similar to the mtDNA of *R. rupricapra rupicapra* (Acc. no. FJ207539.1) and *R. pyrenaica pyrenaica* (Acc. no. FJ207538.1). The CR is organized into three main domains as is general in Caprini, the L (left) where the heavy strand pauses, the central conserved domain, and the R (right) domain that contains the main regulatory elements of the mitochondrial genome. The mtC clade of chamois has two repetitive sequences (RS2) of 77 and 75 bp which is also found in the sequenced *R. rupicapra* and *R. pyrenaica* but it displays a 41 bp deletion in the first RS relative to these sequences.

The number of substitutions per nucleotide between pairs of sequences was lower for RNA genes than for coding sequences and the CR showed the higher p distances (Figure 1). Distances between mtW and mtC (pairs pyr-orn and pyr-cart) are, in general, lower than distances between representatives of the clade mtC and the clade mtE (pairs rup-orn and rup-cart). The differentiation between the sequences of *ornata* and *cartusiana* is very low (p=0.0022, SE=0.0013) except for the CR (p=0.0197, SE=0.0045). The numbers of synonymous substitutions per synonymous site were significantly higher than the number of non-synonymous substitutions per non-synonymous site for every protein-coding gene (Table 1).

Phylogenetic analysis using 16,384 positions in the dataset of Caprini produced the tree presented in Figure 2. The relationships between the sequences of *Rupicapra* obtained from the analysis of the complete mtDNA dataset are supported by the four methods of tree construction (Table 2). Phylogenetic analysis of partial datasets always grouped the two sequences of the *Rupicapra* clade mtC (*ornata* and *cartusiana*) with maximum support. In the same way, all the analyses recovered the internal node grouping the four sequences of *Rupicapra* with maximum support. The node joining the clade mtC with *pyrenaica* was recovered with high support for the datasets of rRNAs and of coding sequences but was poorly supported by the analysis of tRNAs

(see Table 2). An alternative hypothesis grouping *pyrenaica* and *rupicapra* and placing the clade mtC in a basal position was supported by the analysis of the CR. The divergence times of the *Rupicapra* mitochondrial lineages were estimated from the complete dataset. The basal lineages (clades mtW-C and mtE) diverged around 1.93 mya (1.56-2.33 CI, 95% HPD) followed by the split of the branches mtC and mtW, approxiamately1.27 mya (1.00-1.57 CI, 95% HPD). Within the clade mtC, the split between *ornata* and *cartusiana* was dated to 0.09 mya (0.05-0.12 CI, 95% HPD).

4. Discussion

The mitochondrial genome of the clade mtC of *Rupicapra* is shorter than the reported sequences for *R. pyrenaica* and *R. rupicapra*. The clade mtC presents a deletion of 41 bp in the D-loop of the CR, 117 nt to the right of tRNApro. The deleted fragment is flanked by a perfect direct repeat of the sequence ACAAACCCAC. The same deletion was also present in 9 individuals of *R. pyrenaica* whose CR had been previously sequenced and correspond to the haplotypes CR14 (GU951856.1), CR18 (GU951860.1) and CR19 (GU951861.1) in Rodriguez et al. (2010). The presence of the same deletion in different mtDNA clades and the fact that it is flanked by a direct repeat suggest that it occurred by the same mechanism that could be related to slippage during replication.

Most substitutions in coding regions are synonymous as expected for genes subject to purifying selection. Among the enzymes involved in oxidative phosphorylation, the ATPases show more amino acid changes while COX genes seem to be highly conserved. Our findings are in concordance with the general pattern of variation that has been reported for Caprini (Hassanin et al, 2009). The phylogenetic analysis of complete mitochondrial genomes (see Figure 2) groups the clade mtC of the genus *Rupicapra* within the mtW branch, represented by *R. p. pyrenaica*. Hassanin et al. (2009) have identified several amino acid replacements that are diagnostic of the different clades of Caprini. We found the molecular signatures of *Rupicapra* in the mitochondrial sequences of the subspecies *R. p. ornata* and *R. r. cartusiana*. The diagnostic replacement 61S->P in ATP8 appears to be erroneous in Hassanin et al. (2009) however, as this substitution does not appear in any of the sequences of *Rupicapra*. Previous analysis of the relationships between the mitochondrial lineages of *Rupicapra* based on partial sequences provided weak resolution due to the rapid radiation of the three main clades (Crestanello et al., 2009; Rodríguez et al., 2009; Rodríguez et al., 2010). The analysis of complete mitochondrial sequences in the present study demonstrates a closer relationship of the clade mtC to Iberian chamois (*R. p. pyrenaica)* than to the Eastern chamois (*R. r. rupicapra)*. The same topology is returned by the analysis of the partial datasets of coding sequences and of rRNAs, whereas alternative topologies are obtained by assessing the tRNAs or the CR. The topology of the tree of Caprini concurs with the ones obtained in previous studies (Hassanin et al. 2012; Bibi, 2013) and is remarkable in that the distance between the clades mtC (*R. p. ornata* and *R. r. cartusiana)* and mtW (*R. p. pyrenaica*) is comparable to some distances obtained between different species of Caprini.

The divergence time estimates between chamois lineages based on the complete mtDNA are larger than the previous ones (Lalueza-Fox et al., 2005; Rodríguez et al., 2010) and confirm that the chamois inhabited Europe since the Late Pliocene or, at least, since the Early Pleistocene. It has been shown that the estimated rate of substitutions is larger when based on young calibration points than on older ones (Ho, 2005). Even taking into account that the lack of most terminal calibration points in our study can lead to some overestimation, divergence times are much older than the age of the first fossil chamois in Europe, that was recorded from the middle Pleistocene in France (Masini and Lovari, 1988). The lineages mtC and mtW also diverged well before the major glaciations of the Quaternary. The split between the two representatives of the clade mtC, *R. pyrenaica ornata* and *R. rupicapra cartusiana*, occurred much more recently, at the late Pleistocene during the Würm glaciation. The topology obtained from mtDNA, joining together the chamois populations of the Apennines and the Massif of Chartreuse that belong to different species, is in conflict with topologies obtained from different nuclear datasets. According to microsatellites (Rodríguez et al., 2010), Y chromosome (Pérez et al., 2011) and the melanocortin 1 receptor gene (Pérez et al., 2013), the population of Chartreuse, *R. rupicapra cartusiana*, groups without ambiguity with its conspecific populations of *R. rupicapra*. In particular, *R. r. cartusiana* is very close to its neighbor, the alpine chamois, *R. r. rupicapra* for all of the three above mentioned nuclear markers. Discordance of spatial patterns of nuclear and mitochondrial gene pools was also reported in a regional study in the Eastern Alps and was interpreted as a combined effect of phylogeographic background and sex-specific dispersal (Schaschl, 2003). The presence of the mitochondrial clade mtC in an otherwise *R. r. rupicapra* genome may be explained by introgression. The nuclear genome of *cartusiana* has likely been replaced or rather there was introgression of the residing mitochondrial genome of the population in the massif of Chartreuse into the immigrant population from the Eastern Alps. Probably, a small isolated population in the Chartreuse Mountains received immigrant males from the East with hybridization and drift acting such that the original nuclear genome was mostly replaced. The process of mtDNA introgression has been repeatedly reported in mammal species (Currat et al., 2008; Melo-Ferreira et al., 2014; Petit and Excoffier, 2009; Ropiquet and Hassanin, 2006) and has been explained as a consequence of male-biased dispersal and female philopatry. Hybridization between resident females and immigrant males together with strong drift early after hybridization can account for the persistence of female transmitted mtDNA in the population while the nuclear DNA is replaced (Petit and Excoffier, 2009). Our interpretation agrees with the observation of Lovari and Scala (1980) that *R. r. cartusiana* bears some intermediate morphological phenotypes between *R. pyrenaica* and *R. rupicapra*. Future work including more markers could eventually find more remnants of this old hybridization within the genome of *R. r. cartusiana*.

Acknowledgments

This work was supported by the Spanish “Ministerio de Economía y Competitividad” (grant number CGL2011-25117). The authors are indebted to Carlos Nores, Luca Rossi and Jacques Michalet for providing the samples. The chamois drawings in the Graphical Abstracts were

adapted from Lovari (1985). We want to thank Sara de Albornoz for the correction of English and Steve Smith for his valuable comments on the manuscript.

Caption to Figure 1

Number of substitutions per nucleotide (p distance), between pairs of sequences of *Rupicapra*. Distances were computed for the different mitochondrial genes, sets of genes (tRNAs, rRNAs, coding genes and the CR) and complete mitochondrial sequences. The taxa compared were: *R. rupicapra rupicapra* of the clade mtE (rup); *R. pyrenaica pyrenaica* of the clade mtW (rup); *R. pyrenaica ornata* (orn) and *R. rupicapra cartusiana* (cart), both of the clade mtC.

Caption to Figure 2

Bayesian phylogenetic tree of complete mtDNA genomes. The calibration points are indicated with an arrow. Node values are divergence times and node bars are the 95% bounds of the highest posterior density (95% HPD). Accession numbers to GenBank are given as Supplementary Material.

References

Avise, J.C., 2004. Molecular markers. Natural history and evolution (2nd Ed.), Sinauer, Sunderland, MA.

Bibi, F., 2013. A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. BMC Evol. Biol. 13, 166.

Corlatti, L., Lorenzini, R., Lovari, S., 2011. The conservation of the chamois Rupicapra spp. Mammal Rev. 41, 163-174.

Crestanello, B., Pecchioli, E., Vernesi, C., Mona, S., Martínková, N., Janiga, M., Hauffe, H.C., Bertorelle, G., 2009. The Genetic Impact of Translocations and Habitat Fragmentation in Chamois (*Rupicapra*) spp. J.Hered. 100, 691-708.

Currat, M., Ruedi, M., Petit, R.J., Excoffier, L., 2008. The hidden side of invasions: massive introgression by local genes. Evolution. 62, 1908-1920.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Felsenstein, J., 1985. Confidence-limits on phylogenies - an approach using the bootstrap. Evolution. 39, 783-791.

Groves, C., Grubb, P., 2011. Ungulate taxonomy. Johns Hopkins University Press, Baltimore.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysi program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41, 95-98.

Hassanin, A., Ropiquet, A., Couloux, A., Cruaud, C., 2009. Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). J. Mol. Evol. 68, 293-310.

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., Jansen van Vuuren, B., Matthee, C., Ruiz-Garcia, M., Catzeflis, F., Areskoug, V., Nguyen, T.T., Couloux, A., 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. C.R. Biol. 335, 32-50.

Ho, S.Y., Phillips, M.J., Cooper, A., Drummond, A.J., 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol 22, 1561-1568.

Ho, S.Y., Phillips, M.J., 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst. Biol. 58, 367-380.

Lalueza-Fox, C., Castresana, J., Sampietro, L., Marques-Bonet, T., Alcover, J.A., Bertranpetit, J., 2005. Molecular dating of caprines using ancient DNA sequences of Myotragus balearicus, an extinct endemic Balearic mammal. BMC Evol.utionary Biol. 5, 70.

Lovari, S., Scala, C., 1980. Revision of Rupicapra Genus. 1. A statistical re-evaluation of Couturier`s data on the morphometry of six chamois subspecies. B. Zool. 47, 113-124.

Lovari, S., 1985. Behavioral repertoire of the Abruzzo Chamois (*Rupicapra pyrenaica ornata*). Säugetierkd Mitt. 32, 113-136.

Masini, F., Lovari, S., 1988. Systematics, phylogenetic-relationships, and dispersal of the chamois (*Rupicapra* spp). Quat. Res. 30, 339-349.

291 Melo-Ferreira, J., Farelo, L., Freitas, H., Suchentrunk, F., Boursot, P., Alves, P.C., 2014. Home-
292 loving boreal hare mitochondria survived several invasions in Iberia: the relative roles of
293 recurrent hybridisation and allele surfing. Heredity (Edinb).112, 265-273.
294 Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and
295 nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418-426.
296 Pérez, T., Albornoz, J., Domínguez, A., 2002. Phylogeography of chamois (*Rupicapra* spp.)
297 inferred from microsatellites. Mol. Phylogenet. Evol. 25, 524-534.
298 Pérez, T., Hammer, S.E., Albornoz, J., Domínguez, A., 2011. Y-chromosome phylogeny in the
299 evolutionary net of chamois (genus *Rupicapra*). BMC Evol. Biol. 11, 272.
300 Pérez, T., Essler, S., Palacios, B., Albornoz, J., Dominguez, A., 2013. Evolution of the
301 melanocortin-1 receptor gene (MC1R) in chamois (*Rupicapra* spp.). Mol. Phylogenet. Evol. 67,
302 621-625.
303 Petit, R.J., Excoffier, L., 2009. Gene flow and species delimitation. Trends Ecol. Evol. 24, 386-
304 393.
305 Rambaut, A., 2006. FigTree: Tree figure drawing tool, version 1.0. Institute of Evolutionary
306 Biology, University of Edinburgh. Avaliable at: http://tree.bio.ed.ac.uk/software/figtree/
307 Rambaut, A., Drummond, A. J., 2009. TRACER 1.5 Publisher: Institute of Evolutionary Biology,
308 University of Edinburgh. Available at: http://tree.bio.ed.ac.uk/software/tracer
309 Rodríguez, F., Hammer, S., Pérez, T., Suchentrunk, F., Lorenzini, R., Michallet, J., Martinkova,
310 N., Albornoz, J., Domínguez, A., 2009. Cytochrome b Phylogeography of Chamois (*Rupicapra*
311 spp.). Population Contractions, Expansions and Hybridizations Governed the Diversification of
312 the Genus. J. Hered. 100, 47-55.
313 Rodríguez, F., Perez, T., Hammer, S.E., Albornoz, J., Domínguez, A., 2010. Integrating
314 phylogeographic patterns of microsatellite and mtDNA divergence to infer the evolutionary
315 history of chamois (genus *Rupicapra*). BMC Evol. Biol. 10, 222.
316 Ropiquet, A., Hassanin, A., 2006. Hybrid origin of the Pliocene ancestor of wild goats. Mol.
317 Phylogenet. Evol. 41, 395-404.
318 Schaschl, H., Kaulfus, D., Hammer, S., Suchentrunk, F., 2003. Spatial patterns of mitochondrial
319 and nuclear gene pools in chamois (*Rupicapra r. rupicapra*) from the Eastern Alps. Heredity 91,
320 125-135.
321 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA 5:
322 Molecular Evolutionary Genetics Analysis Using Maximun Likelihood, Evolutionary Distance and
323 Maximun Parsimony Methods. Mol. Biol. Evol. 28, 2731-2739.
324 Valdez, R., 2011. Genus Rupicapra, in: Wilson, D.E., Mittermeier, R.A. (Eds.), Handbook of the
325 Mammals of the World, Hoofed Mammals, vol. 2. Lynx Editions, Barcelona, pp. 741-743.
326

Table 1

Table 1. Substitutions in protein coding sequences of *Rupicapra*. pS: number of synonymous substitutions per synonymous site; pN: number of non-synonymous substitutions per non-synonymous site; P(Z-test): significance of the difference pS-pN.

Gene	ATP6	ATP8	COX1	COX2	COX3	Cytb	ND1	ND2	ND3	ND4	ND4L	ND5	ND6
Codons	227	67	515	228	261	381	319	347	115	459	99	606	176
% variable aa	4.0	6.0	0.4	0.9	0.4	2.9	1.6	2.0	0.9	1.3	1.0	2.5	2.8
pS	0.0980	0.1270	0.1204	0.0546	0.0802	0.1075	0.1222	0.1027	0.0814	0.0887	0.0624	0.9600	0.1409
pN	0.0095	0.0150	0.0009	0.0019	0.0012	0.0074	0.0035	0.0049	0.0039	0.0033	0.0031	0.0060	0.0082
pN/pS	0.0969	0.1181	0.0075	0.0348	0.0150	0.0688	0.0286	0.0477	0.0479	0.0372	0.0497	0.0063	0.0582
P (Z-test)	0.0000	0.0043	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0006	0.0000	0.0056	0.0000	0.0000

Table 2

Nodes supported in the phylogenetic analysis of the different datasets. Support values are in order of NJ bootstrap/MP bootstrap /ML bootstrap/Bayesian PP

Dataset	Nº nucleotides	Node (pyr-mtC)	Alternative node supported
tRNAs	1590	-/0.41/0.44/0.56	(rup-mtC) 0.40/-/-/-
rRNAs	2565	0.93/0.93/0.96/1	
Coding	11,338	1/1/1/1	
CR	891	-/-/-/-	(pyr-rup) 0.70/0.95/0.89/0.96
Total mtDNA	16,384	1/1/1/1	

Figure 1

Click here to download high resolution image

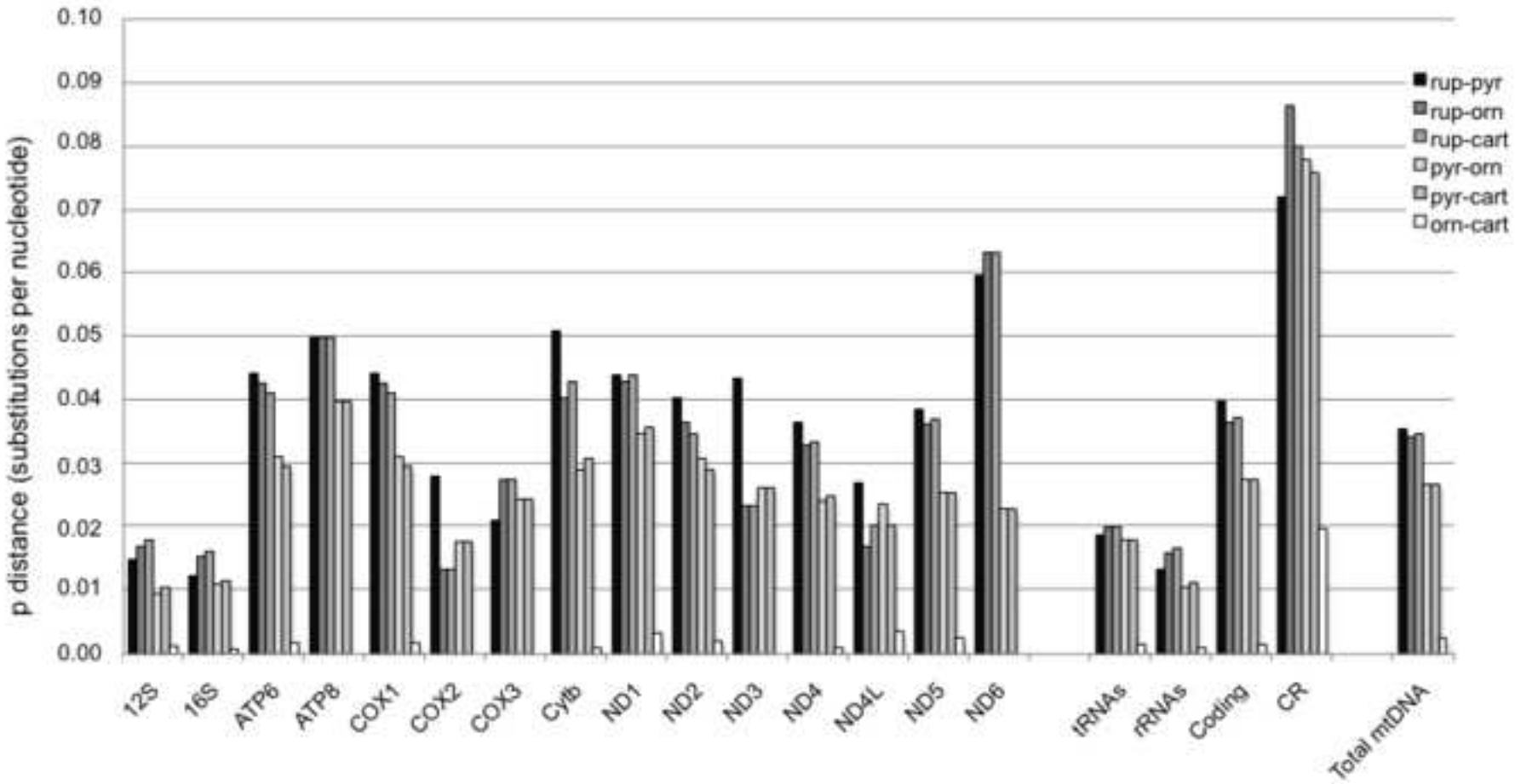
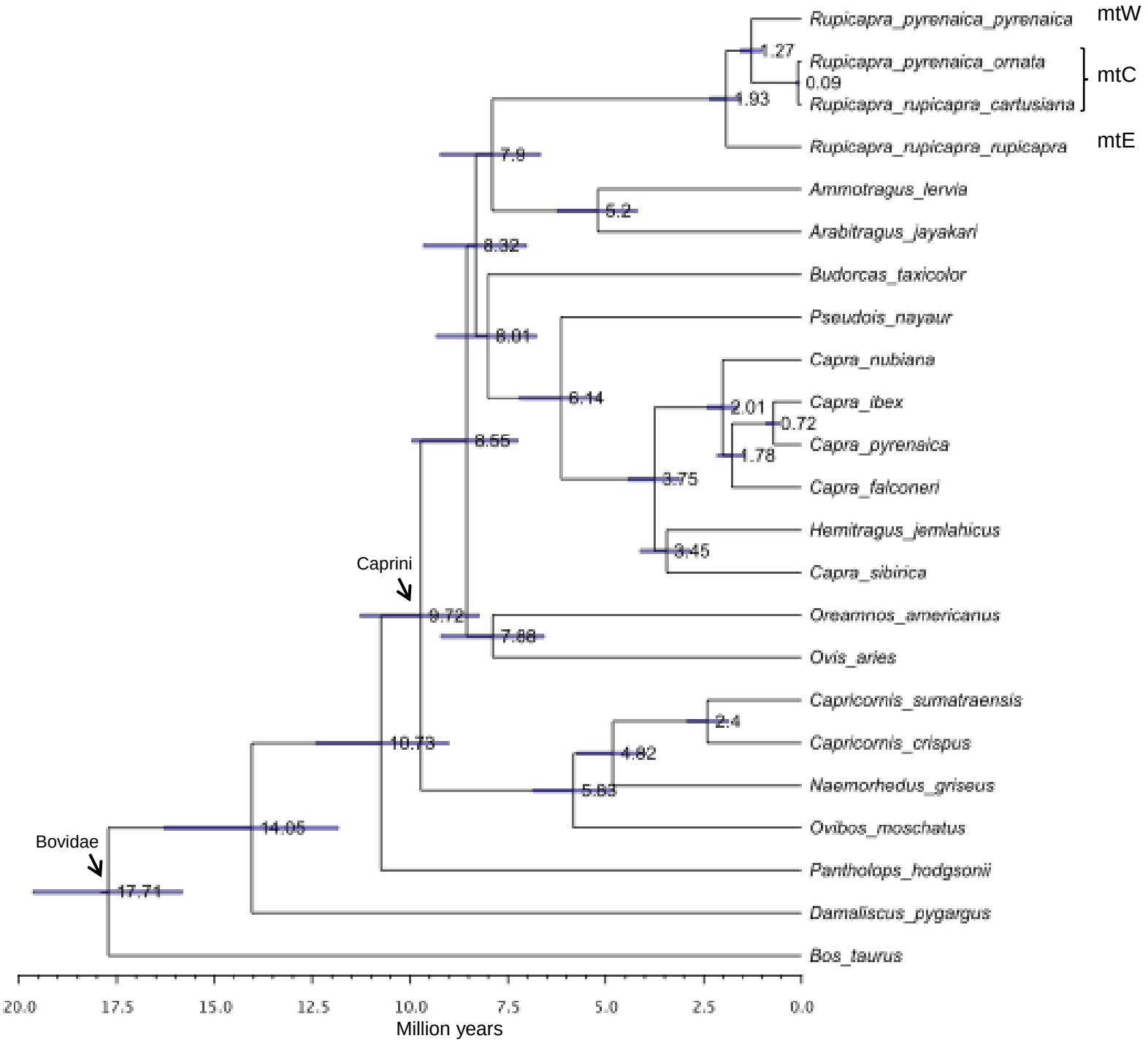


Figure 2


Rupicapra_pyrenaica_pyrenaica
mtW
1.27
Rupicapra_pyrenaica_ornata
0.09
Rupicapra_rupicapra_cartusiana
mtC
1.93
Rupicapra_rupicapra_rupicapra
mtE
7.9
Ammotragus_lervia
5.2
Arabitragus_jayakari
8.32
Budorcas_taxicolor
Pseudois_nayaur
8.01
Capra_nubiana
6.14
2.01
Capra_ibex
0.72
Capra_pyrenaica
8.55
1.78
3.75
Capra_falconeri
Hemitragus_jemlahicus
3.45
Capra_sibirica
Caprini
9.72
Oreamnos_americanus
7.88
Ovis_aries
Capricornis_sumatraensis
2.4
10.73
Capricornis_crispus
4.82
5.83
Naemorhedus_griseus
Ovibos_moschatus
14.05
Bovidae
Pantholops_hodgsonii
17.71
Damaliscus_pygargus
Bos_taurus
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Million years

Supplementary Material 1

Click here to download Supplementary Material: Supplementary material 1.doc

Supplementary Material 2

Click here to download Supplementary Material: Supp2MLtree.ppt