

No Predictable Relationship Between Presacral Vertebral Number and Hes7 Intron Number Across Mammals

Patrick Arnold (✉ arnold.patrick.zoo@gmail.com)

Universität Potsdam <https://orcid.org/0000-0002-6310-5876>

Marie Gurke

Universität Potsdam

Short report

Keywords: Segmentation clock, clock and wavefront model, Hes7, somitogenesis, vertebral spine, mammals

Posted Date: July 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-47722/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

In mammals, the number of vertebrae and the somites they derive from is highly limited. Nevertheless, there are some lineages that have an increased number of presacral vertebrae and thus an elongated trunk. This suggests that somitogenesis, the process of somite formation in early development, is altered in these lineages. According to the 'clock and wavefront' model of somitogenesis, temporal information of somite boundary formation is generated by a traveling wave of cyclic expression of oscillator genes. *Hes7* has been suggested to be a key oscillator gene of this molecular segmentation clock. A previous study showed that reducing the number of introns within the *Hes7* gene results in a more rapid tempo of *Hes7* oscillation and an increased number of presacral vertebrae. Variation in *Hes7* intron number could therefore be a potential evolutionary mechanism for varying vertebral number across mammals. In order to test this hypothesis, *Hes7* intron number is here compared to presacral vertebral number across a variety of mammals.

No significant relationship between both metrics could be detected as their variation across the mammalian phylogeny is fundamentally different. Integrating our data in the previously published mathematical model of *Hes7* oscillation confirms the finding that variation in intron number does not predict variation in presacral vertebrae, rendering a direct causal relationship unlikely. However, our data support the previous suggestion that at least two introns are required for *Hes7* pace making function of the segmentation clock.

Introduction

The number of vertebrae in mammals is highly limited. In the neck, it is constrained to seven in all of the ~ 6400 species except sloths and manatees [1, 2]. Similarly, the number of dorsal vertebrae (sum of thoracic and lumbar vertebrae) is limited to 19 or 20 in the majority of mammals [2, 3]. Thus, nearly all mammals have 26 or 27 presacral vertebrae (PSV, sum of cervical, thoracic and lumbar number). As vertebrae derive from somites, the number of (trunk) somites is accordingly conservative across mammals, too. This contrasts with birds and reptiles in which PSV number as well as the number within the individual vertebral regions is highly variable (e.g., [4]). It has been suggested that a combination of developmental and biomechanical innovations set the meristic constraint very early in mammalian evolution [3, 5]. Nevertheless, there are few lineages that broke these constraints and evolved an increased number of dorsal (and thus presacral) vertebrae: elephants (23–24 dorsal vertebrae), hyraxes (28–31), sea cows (24), golden moles (22–24, these four lineages are part of the Afrotheria), horses and rhinos (perissodactyls, 22–24), and Hero shrews (22–25; [6, 7]). As already the increase of one or two vertebrae is extremely rare among most mammals, these taxa represent significant trunk elongations among mammals.

Somites (as the developmental precursors of vertebrae) are formed in early development via the process of somitogenesis. During somitogenesis, a pair of somites buds off from the paraxial, presomitic mesoderm every two hours in mouse embryos, suggesting that somite segmentation is controlled by a biological clock with a two-hour cycle [8, 9]. The process of somitogenesis has been formalized in a theoretical model – the ‘clock and wavefront’ model [10]. In this model, temporal and spatial information are integrated and determine somite boundaries. Spatial information is provided by the wavefront of Fgf expression which continuously regresses in the posterior direction according to the posterior elongation of the body axis [11, 12]. Temporal information is generated by a traveling wave of cyclic activation/expression of oscillator genes from posterior to anterior [11, 12]. When oscillator expression and wavefront meet, somites are periodically generated. Thus, the period for the formation of one somite and the size of each somite are defined by the period of the oscillator and the distance that the wavefront moves during one cycle of the oscillation, respectively.

The basic helix-loop-helix factor Hes7 is a key effector of Notch signaling during somitogenesis. Its expression follows a two-hour oscillatory cycle controlled by negative feedback and each cycle of Hes7 expression coincides with the generation of each pair of somites [13]. This Hes7 oscillation is proposed to be the molecular basis for the somite segmentation clock [13–15]. An important aspect of the negative feedback control is the transcriptional delay between the pre-mRNA and the final protein due to the transcription, splicing, translation, and transport of the mRNA [13, 15–20]. The transcriptional delay of a gene is suggested to be affected by its number of introns that have to be removed during splicing. Variation in intron number would lead to variation in mRNA maturation time and, thus, in transcriptional delay. Using transgenic mice, Harima and colleagues [21] showed that reducing the number of introns from three (wild-type condition) to two or one within the Hes7 gene actually shortens the delay and results in the acceleration of both Hes7 oscillation and somite segmentation. This eventually led to an increase in the number of somites and vertebrae in the cervical and upper thoracic region, thus increasing total PSV number. Their results suggested that the number of introns is important for the appropriate tempo of oscillatory expression and that Hes7 is a key regulator of the pace of the segmentation clock [16, 21]. Variation in Hes7 intron number could therefore be a potential evolutionary mechanism for varying PSV number across mammals. In order to test this hypothesis, we inferred Hes7 intron number from published genomes across mammals with varying PSV number.

Results And Discussion

Due to the reduction of the sacrum and hindlimbs, PSV number cannot reliably be determined in fully aquatic mammals (cetaceans and sirenians). For this reason, we used data for all mammals (aquatic + non-aquatic) to examine variation in Hes7 intron number but data for non-aquatic mammals only for the comparison with presacral vertebral number. We retrieved information on Hes7 intron number from the genomes of 55 mammals (nine aquatic + 46 non-aquatic) available on NCBI GenBank (see Table S1 in Additional File 1). Although 32 of 55 mammals have three introns like the mouse, there is considerable variation in the remaining species. Variation ranges from two introns in the horse (*Equus caballus*) to six in the Weddell seal (*Leptonychotes weddellii*) (Fig. 1). In addition, Hes7 splicing variants have been

predicted in 13 out of 55 mammalian genomes, resulting in variable number of introns, splicings, and thus mRNA maturation time. For instance, there are variants requiring the splicing of two, three or four introns in the gray short-tailed opossum (*Monodelphis domestica*) whereas there two variants with five plus one with three introns in the sperm whale (*Physeter macrocephalus*). Variation in Hes7 intron number shows no significant phylogenetic signal no matter of minimum, mean or maximum intron number is tested (all mammals: $K = 0.28$ $p = 0.12$, $K = 0.25$ $p = 0.18$, $K = 0.21$ $p = 0.35$, respectively; non-aquatic mammals: $K = 0.30$ $p = 0.13$, $K = 0.26$ $p = 0.24$, $K = 0.20$ $p = 0.49$, respectively). A Blomberg's $K = 1$ prompts that a trait follows a Brownian motion model of evolution [22], the low and non-significant K values observed here suggest that phylogenetic relatedness cannot explain variation in Hes7 intron number. It further suggests that intron number might be quite different in closely related lineages (Fig. 1).

PSV number was collected for the same 46 non-aquatic mammalian species from the literature [2, 3, 6]. 35 species show the conserved number of 26 or 27 PSV but in the remaining ones it varies between 22 in the nine-banded armadillo (*Dasypus novemcinctus*) to 31 in the horse (*Equus caballus*) and the Cape golden mole (*Chrysochloris asiatica*) (Fig. 1). There is a strong and significant phylogenetic signal in PSV ($K = 0.84$ $p = 0.001$), indicating that phylogenetic relatedness is a good predictor of variation in PSV. The signal is even stronger when thoracic ($K = 0.92$ $p = 0.001$) and lumbar number ($K = 0.92$ $p = 0.001$) are tested separately. These findings are largely driven by the shared increase in vertebral number among afrotheres (elephants, tenrecs and relatives) and the shared decrease in apes (humans, chimpanzees and relatives) (Fig. 1). The relationship between Hes7 intron number and PSV number was tested using phylogenetic informed generalized least square analysis (PGLS) to account for phylogenetic independences of variables (particularly PSV number). There is no significant relationship detectable between Hes7 intron number and PSV across mammals (Fig. 2A) no matter if minimum ($t = -0.66$ $p = 0.52$), mean ($t = -0.30$ $p = 0.77$) or maximum intron number ($t = -0.01$ $p = 0.99$) is tested (see PGLS details in Additional File 1).

To further explore the impact of our findings, we implemented the mathematical model of Hes7 oscillating expression presented by Hirata and colleagues [15]. As periodic peaks in protein concentration are suggested to produce somite boundaries when they meet the wavefront [13], the number of peaks in a given time (i.e., the oscillation period) is a direct representation of resulting somites in this model. Among others, the model also includes the parameter of transcriptional delay (T_m). $T_m = 29$ min has been shown to result in a pattern of somite formation closely resembling the condition in wild type mice with three introns [21]. Reducing T_m by five minutes ($T_m = 24$ min) decreases the oscillation period of the Hes7 protein and would result in an additional somite in a given time (but with decreased amplitudes in protein concentration). With this parameter set, the model mimics the condition in transgenic mouse lines with two introns [21]. Further decreasing T_m , however, dampens the oscillation until there is no somite formation expected any more at $T_m = 10$ min [16, 21]. Input values were taken from [21] but T_m was varied according the range of intron number in our empirical data set. Whereas two introns result in an additional peak (and expected somite/vertebra), Hes7 protein peak and expected somite number decreases with increasing intron number by one peak/somite per intron (Fig. 2B). According the model,

the Weddell seal (*Leptonychotes weddellii*) is expected to lack four somites (and thus vertebrae) given six introns in the *Hes7* gene. However, it owns 27 PSV instead of 23. Contrary, the African elephant (*Loxodonta africana*) with four introns is expected to have 25 PSV but owns 29. Accordingly, it is not possible to predict PSV number from *Hes7* intron number, particularly in mammals deviating from 26/27 PSV. The horse (*Equus caballus*) might be the only exception as its reduction to only two introns matches its increased PSV count. A deeper look into a possible correlation of *Hes7* intron/PSV number among odd-toed ungulates (perissodactyl: horses and donkeys, tapirs, rhinos) might therefore be interesting in future studies. The pilot whale (*Globicephala melas*) is the only other mammal examined that has only splicing variants with two introns (all other species with minimum number of two introns in Fig. 1 have at least one other variant with more introns). Although its number of PSV cannot exactly be determined, it has only a medium total number of vertebrae among cetaceans [23].

Conclusions

Our findings suggest that changes in the number of *Hes7* introns are not generally causative for the evolutionary variation of PSV across mammals, at least in the experimental and theoretical framework presented by Harima and colleagues [21]. They further highlight that outcomes from evo-devo studies on mice might not always be transferable to evolutionary patterns in mammals in general. However, this does not deny the evolutionary impact of Harima and colleagues' [21] results as they provide crucial insights into the developmental mechanics of axial patterning through the segmentation clock. Their transgenic mouse lines show that simple changes in the transcriptional delay of key oscillator genes are sufficient to modify the number of produced somites without resulting in deleterious pleiotropic effects due to complete gene loss. Our results further support the authors' hypothesis that at least two introns are required for *Hes7* pace making function of the segmentation clock. Transgenic mouse lacking two or all introns show severe segmentation defects [16, 21]. According to their mathematical model, this is based on the dampened oscillation of *Hey7* protein concentration that vanishes after a few circles (Fig. 2B). In accordance with this, no species among the investigated mammals has less than two introns. Other genetic mechanisms resulting in varying body axis length in vertebrates have previously been discussed. For instance, they propose the alteration of the basic body patterning into head, trunk, and tail [24], disturbance of the Notch signaling pathway [25], or elongated growth of presomitic mesoderm [26]. The genetic basis underlying the elongated trunk in afrotheres and perissodactyls, however, remains to be determined in future investigations.

Methods

In order to identify orthologs among mammals, mouse *Hes7* protein (BAB39526.1) was blasted against all protein sequences of annotated mammalian genomes available on NCBI GenBank (blastp). Blast hits with less than 65% query coverage were discarded to avoid including proteins of other members of the *Hes* gene family based on the conserved helix-loop-helix only. Remaining sequences were realigned using MAFFT v7.310 [27]. A protein tree was subsequently computed in RAxML 8.2.11 [28] using the

PROTGAMMABLOSUM62 substitution model. A group of proteins of phylogenetically diverse species closely clustered together but split with an unusually long branch from other protein sequences. We therefore assumed that these sequences still represent other members of the Hes gene family that have been misannotated as Hes7 and discarded them from the data set. Nevertheless, there always was another and putative Hes7 protein for all species for which sequences have been discarded in this step. Remaining sequences were more accurately realigned in MAFFT using the E-INS-I algorithm for local similarities but dissimilarities in-between. A final protein tree was recomputed in RAxML, too (see Fig. S1 in Additional File 1). Genomic sequences of Hes7 genes were subsequently identified based on protein IDs in the annotation files. Numbers of predicted exons for each splicing variant were collected. Number of introns (i.e., required splicings between collected exons) were inferred based on these exon counts for every splicing variant in every species.

The phylogenetic species tree and divergence times were gathered from the TimeTree database [29]. All phylogenetic comparative analyses were conducted in R 3.6.0 [30] using the packages ape [31], geiger [32], nlme [33], and phytools [34]. The phylogenetic signal was calculated using Blomberg's K [22] and its significance estimated in 1000 simulations. PGLS was conducted by applying generalized least square regression with a covariation structure following Brownian motion. Model parameters were estimated using maximum likelihood. The mathematical model of Hes7 was implemented in Matlab R2020 (MathWorks, Inc.) using the equations presented by [15, 21]. Proceeding from $T_m = 29$ min being equivalent of the transcriptional delay resulting from splicing three introns, changes in intron number were implemented by changing T_m by five minutes per intron (following [21]).

Declarations

Ethnic approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated and/or analyzed during the current study are included in this published article and its additional information files.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors contribution

PA designed the study, did statistical analyses, implemented the model, prepared the figures and drafted the manuscript. MG collected and analyzed the sequences to retrieve intron numbers. All authors approved the final version of the manuscript.

Acknowledgements

We thank Stefanie Hartman and Achim Klittich for help with sequence retrieval and implementation of the mathematical model, respectively.

References

1. Galis F. Why do almost all mammals have seven cervical vertebrae? Developmental constraints, Hox genes, and cancer. *J Exp Zool.* 1999;285:19–26.
2. Narita Y, Kuratani S. Evolution of the vertebral formulae in mammals: a perspective on developmental constraints. *Journal of Experimental Zoology Part B: Molecular Developmental Evolution.* 2005;304:91–106.
3. Williams SA, Spear JK, Petrullo L, Goldstein DM, Lee AB, Peterson AL, Miano DA, Kaczmarek EB, Shattuck MR. Increased variation in numbers of presacral vertebrae in suspensory mammals. *Nature Ecology Evolution.* 2019;3:949–56.
4. Müller J, Scheyer TM, Head JJ, Barrett PM, Werneburg I, Ericson PGP, Pol D, Sánchez-Villagra MR: Homeotic effects, somitogenesis and the evolution of vertebral numbers in recent and fossil amniotes. *Proceedings of the National Academy of Sciences* 2010, 107:2118–2123.
5. Galis F, Carrier DR, Van Alphen J, Van Der Mije SD, Van Dooren TJM, Metz JAJ, ten Broek CMA: Fast running restricts evolutionary change of the vertebral column in mammals. *Proceedings of the National Academy of Sciences* 2014, 111:11401–11406.
6. Sánchez-Villagra MR, Narita Y, Kuratani S. Thoracolumbar vertebral number: the first skeletal synapomorphy for afrotherian mammals. *Syst Biodivers.* 2007;5:1–7.
7. Stanley WT, Robbins LW, Malekani JM, Mbalitini SG, Migurimu DA, Mukinzi JC, Hulselmans J, Prévot V, Verheyen E, Hutterer R. A new hero emerges: another exceptional mammalian spine and its potential adaptive significance. *Biol Let.* 2013;9:20130486.
8. Saga Y, Takeda H. The making of the somite: molecular events in vertebrate segmentation. *Nat Rev Genet.* 2001;2:835–45.

9. Pourquié O. Vertebrate somitogenesis. *Annu Rev Cell Dev Biol.* 2001;17:311–50.
10. Cooke J, Zeeman EC. A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *J Theor Biol.* 1976;58:455–76.
11. Oates AC, Morelli LG, Ares S. Patterning embryos with oscillations: structure, function and dynamics of the vertebrate segmentation clock. *Development.* 2012;139:625–39.
12. Yabe T, Takada S. Molecular mechanism for cyclic generation of somites: lessons from mice and zebrafish. *Dev Growth Differ.* 2016;58:31–42.
13. Bessho Y, Hirata H, Masamizu Y, Kageyama R. Periodic repression by the bHLH factor Hes7 is an essential mechanism for the somite segmentation clock. *Genes Dev.* 2003;17:1451–6.
14. Bessho Y, Sakata R, Komatsu S, Shiota K, Yamada S, Kageyama R. Dynamic expression and essential functions of Hes7 in somite segmentation. *Genes Dev.* 2001;15:2642–7.
15. Hirata H, Bessho Y, Kokubu H, Masamizu Y, Yamada S, Lewis J, Kageyama R. Instability of Hes7 protein is crucial for the somite segmentation clock. *Nat Genet.* 2004;36:750–4.
16. Takashima Y, Ohtsuka T, González A, Miyachi H, Kageyama R: Intronic delay is essential for oscillatory expression in the segmentation clock. *Proceedings of the National Academy of Sciences* 2011, 108:3300–3305.
17. Zeiser S, Rivera O, Kuttler C, Hense B, Lasser R, Winkler G. Oscillations of Hes7 caused by negative autoregulation and ubiquitination. *Comput Biol Chem.* 2008;32:48–52.
18. Lewis J. Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. *Curr Biol.* 2003;13:1398–408.
19. Monk NAM. Oscillatory expression of Hes1, p53, and NF- κ B driven by transcriptional time delays. *Curr Biol.* 2003;13:1409–13.
20. Jensen MH, Sneppen K, Tiana G. Sustained oscillations and time delays in gene expression of protein Hes1. *Febs Letters.* 2003;541:176–7.
21. Harima Y, Takashima Y, Ueda Y, Ohtsuka T, Kageyama R. Accelerating the tempo of the segmentation clock by reducing the number of introns in the Hes7 gene. *Cell Reports.* 2013;3:1–7.
22. Blomberg SP, Garland T Jr, Ives AR, Crespi B. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution.* 2003;57:717–45.
23. Tinker SW: *Whales of the world.* Brill Archive; 1988.
24. Aires R, Jurberg AD, Leal F, Nóvoa A, Cohn MJ, Mallo M. Oct4 is a key regulator of vertebrate trunk length diversity. *Dev Cell.* 2016;38:262–74.
25. Mikawa S, Sato S, Nii M, Morozumi T, Yoshioka G, Imaeda N, Yamaguchi T, Hayashi T, Awata T. Identification of a second gene associated with variation in vertebral number in domestic pigs. *BMC Genet.* 2011;12:5.
26. Gomez C, Özbudak EM, Wunderlich J, Baumann D, Lewis J, Pourquié O. Control of segment number in vertebrate embryos. *Nature.* 2008;454:335–9.

27. Katoh K, Misawa K, Kuma Ki, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30:3059–66.
28. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312–3.
29. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol.* 2017;34:1812–9.
30. R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
31. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* 2019;35:526–8.
32. Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, Alfaro ME, Harmon LJ. geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics.* 2014;30:2216–8.
33. Pinheiro J, Bates D, DebRoy S, Sarkar D. R Core Team. 2019. nlme: linear and nonlinear mixed effects models. R package version 3.1–141.
34. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol.* 2012;3:217–23.
35. Additional. files.
36. Additional. file 1.
37. File name. additional_file1.pdf.
38. File format. PDF-Document (.pdf).
39. Title of data. Table S1, Detailed results of PGLS analyses, Figure S1.
40. Description of data. List of species, accession numbers, inferred introns and vertebral numbers. Detailed results of PGLS analyses. Maximum likelihood tree of Hes7 proteins that past filtering.

Figures

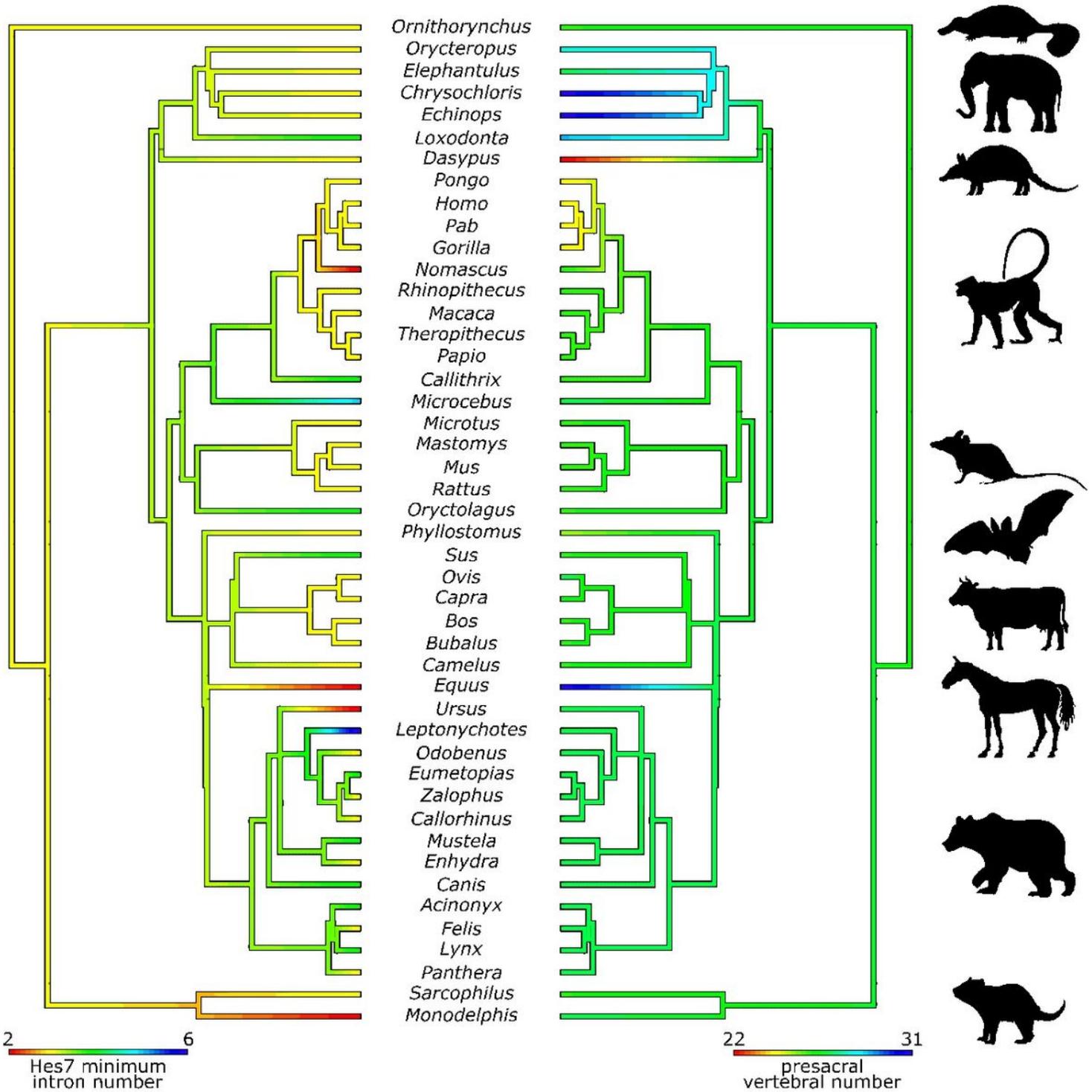


Figure 1

A) Hes7 minimum intron number and B) presacral vertebral number mapped to the phylogeny of 46 non-aquatic mammals.

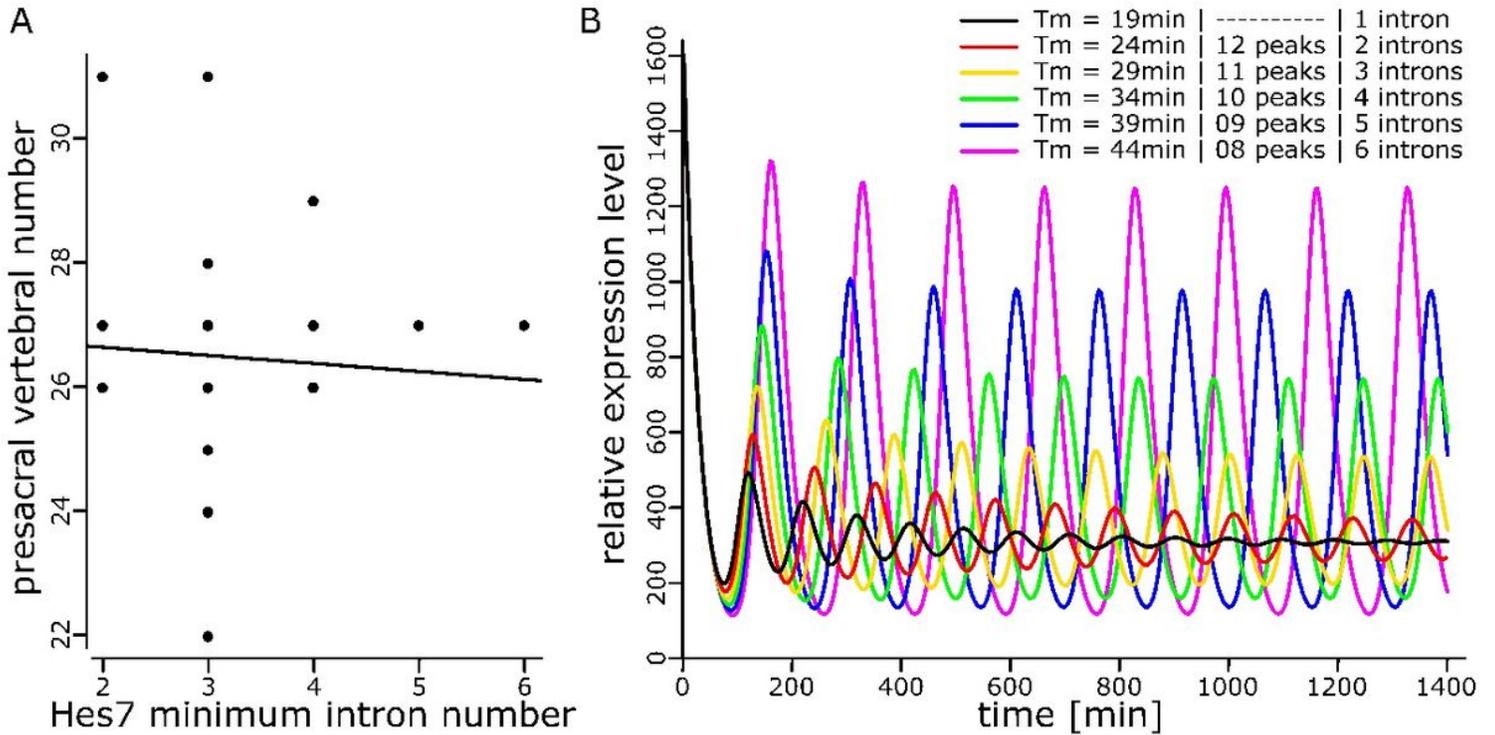


Figure 2

A) The phylogenetic informed generalized least square analysis (PGLS) shows no significant relationship between Hes7 minimum intron and presacral vertebral number across mammals. B) Mathematical model of Hes7 cyclic expression [15, 21]. Increasing transcriptional delay (T_m), representing increasing intron number, results in decreasing number of expression peaks within a certain time (i.e., increasing period, decreasing frequency) that would meet the wavefront to form somites (i.e., decreasing number of somites). T_m corresponding to only one intron results in oscillation that is quickly dampened.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [additionalfile1.pdf](#)