Gene expression phylogenies and ancestral transcriptome reconstruction resolves major transitions in the origins of pregnancy

Katie Mika[1,2], Camilla M. Whittington[3], Bronwyn M. McAllan[3,4], and Vincent J. Lynch[5,*]

[1] Department of Human Genetics, The University of Chicago, 920 East 58th Street, CLSC 319C, Chicago, IL 60637, USA.

[2] Current address: Department of Organismal Biology and Anatomy, University of Chicago, 1025 E 57th Street, Chicago, IL, 60637, USA

[3] School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia

[4] Charles Perkins Centre, University of Sydney, Sydney, NSW 2006, Australia

[5] Department of Biological Sciences, University at Buffalo, SUNY, 551 Cooke Hall, Buffalo, NY, 14260, USA.

[*]Correspondence: vjlynch@buffalo.edu

Abstract

Structural and physiological changes in the female reproductive system underlie the origins of pregnancy in multiple vertebrate lineages. In mammals, for example, the glandular portion of the lower reproductive tract has transformed into a structure specialized for supporting fetal development. These specializations range from relatively simple maternal provisioning in egg-laying monotremes to an elaborate suite of traits that support intimate maternal-fetal interactions in Eutherians. Among these traits are the maternal decidua and fetal component of the placenta, but there is considerable uncertainty about how these structures evolved. We identified the origins of pregnancy utilizing ancestral transcriptome reconstruction to infer functional evolution of the maternal-fetal interface. Remarkably, we found that maternal gene expression profiles are correlated with degree of placental invasion. These results indicate that an epitheliochorial-like placenta evolved early in the mammalian stem-lineage and that the ancestor of Eutherians had a hemochorial placenta, and suggest maternal control of placental invasiveness. Collectively, these data resolve major transitions in the evolution of pregnancy and indicate that ancestral transcriptome reconstruction can be used to study the function of ancestral cell, tissue, and organ systems.

Introduction

Studies of the fossil record have revealed in fine detail major stages in the origin and diversification of evolutionary novelties (structures) and innovations (functions) such as the vertebrate limb and skull (Abzhanov, 2015; Hirasawa and Kuratani, 2015), the turtle shell (Lyson and Bever, 2020), feathers (Chen et al., 2015), and flowers (Chanderbali et al., 2016). Many features of soft tissues and macromolecular structures, however, are lost during the fossilization process and thus leave little to no trace in the fossil record. To reconstruct the history of these characters, including DNA and amino acid sequences, morphology, physiology, and behavior, among others, evolutionary studies have traditionally relied on comparative (phylogenetic) methods such as parsimony or model-based maximum likelihood or Bayesian inference (Lewis and Olmstead, 2001). These methods infer ancestral characters from the distribution of character states among extant species, the latter of which also allows increased model complexity including unequal character state transitions (Lewis and Olmstead, 2001; Pagel, 1999), variable rates among sites and branches (Galtier, 2001; Yang, 1994, 1993), and character dependent diversification (Maddison et al., 2007).

2

Extant mammals span crucial transitions in the origins of pregnancy (**Figure 1A**) and are an excellent system in which to explore the origins of evolutionary novelties. The platypus and echidna (monotremes), for example, are oviparous (egg-laying) but retain the egg in the glandular portion of uterus for about two weeks. During this period the developing embryo is nourished by uterine secretions (matrotrophy) delivered through a simple yolk sac placenta (Hughes and Hall, 1998; Renfree and Shaw, 2013). Viviparity (live-birth) evolved in the stem-lineage of therian mammals, but marsupials and eutherians have very different reproductive strategies, particularly in the ontogenetic origins of the definitive placenta and the arrangement of the maternal-fetal interface (Freyer and Renfree, 2009; Renfree and Shaw, 2013; Renfree, 2010). In most marsupials, the embryonic portion of the placenta is derived from the yolk sac, which in may come into direct contact but does not invade the maternal endometrium. While the yolk sac has essential functions during early pregnancy in eutherians, the definitive placenta is derived from chorion and allantois (chorioallantois) and varies in its degree of endometrial invasion (Swanson and Skinner, 2018)(**Figure 1B**). Thus, matrotrophy and a yolk-sac derived placenta were present in the mammalian stem-lineage and transitioned to a chorioallantoic placenta in eutherians.

Unfortunately, most characters related to pregnancy leave little to no trace in the fossil record. Thus, studies exploring the evolution of pregnancy have relied on comparing morphological differences between extant mammals to reconstruct the steps in the origins of pregnancy. These comparative analyses have used multiple methods to reconstruct the arrangement of the mammalian maternal-fetal interface and reached contradictory conclusions about the degree of placental invasiveness in the eutherian ancestor (**Table 1**). Here, we use comparative transcriptomics and maximum likelihood to infer ancestral gene expression states (ancestral transcriptome reconstruction) from 23 amniotes with different parity modes and degrees of placental invasion to identify the evolutionary history of the maternal-fetal interface. We found strong evidence that the last common ancestor of eutherian mammals had an invasive hemochorial placenta, as well as convergence in gene expression profiles during the independent evolution of non-invasive epitheliochorial placentas in marsupials and some eutherian lineages. These data indicate that the degree of placental invasion can be inferred from endometrial gene expression profiles, suggesting that placental invasiveness is regulated by gene expression profiles in the maternal endometrium rather than the fetal portion of the placenta.

3

Table 1. Inferences of placental invasiveness in the eutherian ancestor 1880-2021. Note that we follow Mossman (Mossman, 1937, p. 156) and define a placenta as "any intimate apposition or fusion of the fetal organs to the maternal (or paternal) tissues for physiological exchange."

Study	Data	Method(s)	State
(Haeckel 1883)	–	Ontogeny	"Non-Invasive"
(Turner, 1876)	–	Ontogeny	Epitheliochorial
(Wislocki, 1929)	–	Implicit Parsimony?	Hemochorial
(Hill, 1932)	–	Ontogeny	Epitheliochorial
(Mossman, 1937)		Ontogeny	Endotheliochorial
(Portmann, 1938)	–	Ontogeny	Hemochorial
(Martin, 1969)	–	Ontogeny / Implicit Parsimony	Endotheliochorial
(Kihlström, 1972)	–	Gestation length / Placenta-type Correlation	(Endothelio/hemo)chorial
(Luckett, 1976, 1975, 1974)	–	Ontogeny	Epitheliochorial
(Carter and Enders, 2004)	14 taxa	Ontogeny / Implicit Parsimony	Endotheliochorial
(Vogel, 2005)	22 taxa	Implicit Parsimony	Hemochorial
(Elliot and Crespi, 2006)	88 taxa	Maximum Likelihood	Hemochorial
(Mess and Carter, 2006)	36 taxa	Explicit Parsimony	(Endothelio/hemo)chorial
(Wildman et al., 2006)	44 taxa	Explicit Parsimony & Maximum Likelihood	Hemochorial
(Martin, 2008)	18 taxa	Parsimony	Endotheliochorial (ordered) (Endothelio/hemo)chorial (unordered)
(Elliot and Crespi, 2009)	334 taxa	Explicit Parsimony & Maximum Likelihood	(Endothelio/hemo)chorial (MP) Hemochorial (ML)
(McGowen et al., 2014)	66 taxa	Maximum Likelihood	(Superficial) Hemochorial

Results and Discussion

Endometrial Gene Expression Profiling

We previously assembled a collection of transcriptomes from the pregnant or gravid uterine endometrium of therian mammals with varying degrees of placental invasiveness, as well as a monotreme (platypus), a bird, as well as viviparous, oviparous, and reproductively bi-modal lizards (Marinić et al., 2021). The complete dataset includes expression information for 21,750 genes from 23 species (**Figure 2–source data 1**). Principal Component Analysis (PCA) based on gene expression levels in transcripts per million (TPMs) generally grouped species randomly (**Figure 2A**), consistent with noise in gene expression data overwhelming phylogenetic signal. Therefore, we transformed quantitative gene expression values in transcripts per million (TPM) into discrete character states – Genes with TPM ≥ 2.0 were coded as expressed (state=1), genes with TPM < 2.0 were coded as not expressed (state=0), and genes without data in specific species coded as missing (?) (Marinić et al., 2021; Mika et al., 2021). In contrast to PCA based on TPM values, PCA of the binary encoded endometrial transcriptome dataset grouped species by phylogenetic relatedness (**Figure 2B**), indicating significant noise reduction in the binary encoded dataset.

Phylogenetic analyses of endometrial transcriptomes

Next, we used IQ-TREE to infer the best fitting model of character evolution (GTR2+FO+R3), the maximum-likelihood (ML) phylogeny, and branch support metrics for the binary encoded endometrial gene expression dataset. The ML phylogeny (**Figure 3A**) generally followed taxonomic relationships (**Figure 3B**), however, four discordant relationships within therian mammals were inferred with high support: 1) Rather than grouping marsupials into a monophyletic sister-clade to the eutherians, opossum and wallaby were placed as sister-species within the Boreoeutheria; 2) armadillo groups within the Euarchontoglires rather than as sister to all other eutherians; 3) bat groups within the Euarchontoglires rather than within Laurasiatheria; and 4) dog groups within the Euarchontoglires rather than with Laurasiatheria (**Figure 3A/B**). Remarkably while these relationships are incorrect with respect to the species phylogeny, they are correct with respect to placenta-type, i.e., species form well-supported clades based on their degree of placental invasiveness. Wallaby and opossum, for example, have epitheliochorial placentas similar to ungulates, whereas dog has an invasive endotheliochorial placenta similar to the invasive hemochorial placenta of Euarchontoglires. We also used multiple non-parametric

topology tests to directly compare the inferred ML tree to alternative trees with the correct phylogenetic placement of discordant lineages, all of which rejected alternative trees in favor of the ML tree (**Figure 3C**).

These data show significant phylogenetic support for species uterine transcriptomes grouping by parity mode and degree of placental invasiveness rather than phylogenetic relationships. This discordance is particularly striking for opossum and wallaby, which are deeply nested within eutherians with epitheliochorial placentas based on uterine transcriptome data, because the eutherian and marsupial placenta is derived from different extra-embryonic tissues, the chorion and allantois in the former and the yolk-sac in the latter. Thus, there must be significant convergence in endometrial gene expression profiles between eutherians with chorioallantois-derived epitheliochorial placentas and marsupials with yolk-sac epitheliochorial placentas.

Fuzzy C-Means transcriptome clustering

We also used IQ-TREE and the species phylogeny (**Figure 3B**) to reconstruct ancestral gene expression states for each gene (ancestral transcriptome reconstruction). To explore the similarity of extant and reconstructed transcriptomes we used Fuzzy C-Means (FCM) clustering, a "soft" clustering method that allows each sample to have membership in multiple clusters and assigns samples to clusters based on their degree of cluster membership. FCM with two to four clusters (K=2-4) had a clear biological interpretation (**Figure 4**): 1) K=2 clustered eutherians and non-eutherians; 2) K=3 clustered most therians with non-invasive (epitheliochorial) placentas, eutherians with invasive (endotheliochorial or hemochorial) placentas, platypus and sauropsids (i.e. viviparous and oviparous lizards, and birds); 3) K=4 clustered eutherians with invasive placentas, eutherians with epitheliochorial placentas, opossum/wallaby, and sauropsids. A notable exception with K=3-4 is the cluster membership of dunnart, which is discussed in greater detail below. FCM clusters with K=5 and K=6 were similar to K=4, but divided Eutherians with hemochorial placentas into two clusters, and clustered dog, dunnart, and the viviparous skink *Chalcides ocellatus* (**Figure 4**); Beyond K=6 clusters had no clear biological interpretation.

Ancestral transcriptome reconstructions generally clustered with extant species having similar parity modes and degrees of placenta invasiveness (**Figure 4**). For example, FCM with K=2 grouped extant eutherians and their ancestral lineages as well as extant non-eutherians and their ancestral lineages. Similarly, FCM with K=4 grouped extant eutherians with invasive placentas and their ancestral lineages with extant therians with non-invasive placentas and their

ancestral lineages. FCM with K=3–5 clustered the ancestral eutherian (AncEutheria) transcriptome with extant species that have invasive hemochorial placentas and clustered the ancestral therian (AncTheria) and mammalian (AncMammalia) transcriptomes with extant mammals that have non-invasive epitheliochorial placentas. These data suggest that FCM clustering of extant and ancestral reconstructions of endometrial transcriptomes can predict ancestral placenta invasiveness, implying that an epitheliochorial placenta evolved early in the development of mammalian pregnancy and that a hemochorial placenta is ancestral for eutherians.

While FCM clustering generally groups extant and ancestral transcriptomes by phylogenetic relatedness and degree of placental invasiveness, a notable exception is the marsupial fat-tailed dunnart (*Sminthopsis crassicaudata*). FCM clusters with K=2–5 grouped dunnart with non-therians while FCM K=6 clustered dog, dunnart, and the skink, *C. ocellatus* (**Figure 4**). FCM cluster membership coefficients of all three species with K=2–5 were mixed, with significant membership across clusters. In contrast, dog, dunnart, and *C. ocellatus* formed a distinct cluster from all other species at K=6 with nearly 100 % FCM cluster membership in group 6 (**Figure 4**). Remarkably, both dog and dunnart have endotheliochorial placentas whereas the *Chalcides* maternal-fetal interface has been described as either endotheliochorial or epitheliochorial with extensive vascularization and interdigitating folds of hypertrophied uterine and chorioallantoic tissue (Blackburn, 1993; Blackburn and Callard, 1997; Corso et al., 2000). These data suggest that species with endotheliochorial placentas have a gene expression profile that is intermediate between epitheliochorial or hemochorial placentas, yet is also distinct, and that gene expression at the *Chalcides* maternal-fetal interface are converging on a therian-like endotheliochorial pattern.

Maternal control of placental invasion in mammals

Our observation that endometrial gene expression patterns are correlated with degree of placental invasiveness might seem surprising, however, the endometrium controls trophoblast invasion (Cui et al., 2012; Graham and Lala, 1991). For example, the trophoblast of mammals with hemochorial placentas, such as humans and rodents, is only permissive to invasion when the "window of implantation" is opened by the endometrium. Similarly, while the trophoblast of mammals with non-invasive endotheliochorial and epitheliochorial placentas, such as cats, dogs, horses, cows, pigs, and sheep cannot invade into the endometrium, they can invade into ectopic

sites (Reviewed in (Corpa, 2006)). These data indicate that trophoblast invasiveness is ancestral in eutherian mammals. While similar data on ectopic pregnancy is lacking for marsupials, some marsupial lineages, including dunnart, have evolved invasive placentation. In contrast, there is no invasion of maternal tissues during ectopic pregnancy in viviparous reptiles with epitheliochorial placentation such as *Pseudemoia entrecasteauxii* (Griffith et al., 2013). Thus, the invasive ability of trophoblasts most likely either evolved in the stem-lineage of therian mammals or multiple times in therians–in the stem-lineage of eutherians and some lineages of marsupials. Regardless, ancestral maternal control of placental invasion likely allows us to infer ancestral placental invasiveness from ancestral endometrial transcripomes. These data also suggest that maternal control of trophoblast invasion may be mechanistically related to resistance to metastasis in some eutherian lineages with non-invasive placentas (Boddy et al., 2020; Kshitiz et al., 2019; Wagner et al., 2020).

Caveats and limitations

A limitation of this study is that we have only sampled a small number of species. For example, we lack pregnant endometrial samples from most mammals, particularly those with endotheliochorial placentas, as well as a diversity of oviparous and viviparous squamates (there are at least 115 origins of viviparity in squamates (Blackburn, 2015; Blackburn and Brandley, 2015; Blackburn and Starck, 2015). Thus, our inferences from phylogenetic, ancestral reconstruction, and clustering analyses may be biased by small sample sizes and non-random sampling. We also assume that models of evolution designed for phylogenetic inference and ancestral reconstruction of morphological and molecular data are appropriate for gene expression data or binary encoded gene expression data, which may affect our results. Similarly, while Fuzzy C-Means clustering is conceptually similar to topic ("grade of membership") models used in population genetics, its underlying assumptions may be violated for gene expression and binary encoded gene expression data. More detailed studies will be necessary to determine if our results are robust to potential sources of error such as model mis-specification, small sample sizes, and non-random taxon sampling.

Conclusions

Previous critiques of statistical methods to infer ancestral states, particularly in the context of parity mode evolution in squamates, have suggested that ancestral state reconstructions of

morphological characters must be supported by additional kinds of biological support such as anatomical, physiological, and ecological evidence, to be persuasive (Griffith et al., 2015). Here we explored the evolution of parity mode and placental invasiveness in amniotes utilizing comparative gene expression data. While our study also relies on statistical methods to infer ancestral (gene expression) states, this approach is orthogonal to traditional methods that infer ancestral states from morphological characters among extant species. Indeed, gene expression ultimately underlies the development, evolution, and function of anatomical systems. Thus, by reconstructing the evolution of entire transcriptomes we may be able to infer function of ancestral cell, tissue, and organ systems. Our results resolve several evolutionary transformations during the origins of pregnancy, including the early evolution of an epitheliochorial-like placenta in the mammalian stem-lineage, a hemochorial placenta in the ancestor of eutherians, multiple reversions to non-invasive epitheliochorial placentas within some eutherian lineages, convergent evolution of gene expression profiles among species with different ontogenetic origins of epitheliochorial placentas, and maternal control of placental invasiveness.

Materials and Methods

Endometrial Gene Expression Profiling

We previously published a dataset of uterine endometrial transcriptomes, which is also used in this study. Interested readers are referred to Mika, Marinic, and Lynch (2021) for specific details. Briefly, we searched the NCBI BioSample, Short Read Archive (SRA), and Gene Expression Omnibus (GEO) databases using the search terms "uterus", "endometrium", "decidua", "oviduct", and "shell gland". These anatomical terms refer to the glandular portion of the female reproductive tract, which is specialized for maternal-fetal interactions or shell formation. We then manually curated transcriptomes and excluded those that did not indicate whether tissue samples were from pregnant or gravid tissues and datasets composed of pooled tissues. Gene expression data were analyzed with Kallisto (Bray *et al.*, 2016) version 0.42.4 to pseudo-align the raw RNA-Seq reads to reference transcriptomes and to generate transcript abundance estimates (see **Figure 3–source data 1** for accession numbers and reference genome assemblies); Kallisto was run using default parameters, bias correction, and 100 bootstrap replicates.

Gene expression phylogeny and ancestral transcriptome reconstruction

We used the binary encoded endometrial transcriptome dataset for phylogenetic analyses and to reconstruct ancestral gene expression states. Gene expression phylogenies were inferred with IQ-TREE (Nguyen et al., 2015) using the best-fitting model of character evolution determined by ModelFinder (Kalyaanamoorthy et al., 2017). The best fitting model was inferred to be the General Time Reversible model for binary data (GTR2), with character state frequencies optimized by maximum-likelihood (FO), and a FreeRate model of among site rate heterogeneity with three categories (R3) (Soubrier et al., 2012). Ancestral gene expression states for each gene were inferred using the empirical Bayesian method implemented in IQ-TREE, the GTR2+FO+R3 model of character evolution, and the species phylogeny as a constraint tree (**Figure 3B**).

Branch support was assessed using the standard (StdBoot) and ultrafast (UFBoot) bootstraps, which assess the effects of sampling bias on branch support (Hoang et al., 2018; Minh et al., 2020). We also used several single branch tests, including the SH-like aLRT and the parametric aLRT (Anisimova et al., 2006; Guindon et al., 2010), aBayes (Anisimova et al., 2011), and the local (LBoot) bootstrap tests (Minh et al., 2020); single branch tests assess whether a branch provides a significant likelihood improvement compared to a null hypothesis that collapses the branch to a polytomy but leaves the rest of the tree topology unaltered. We considered a clade to be highly-supported if its StdBoot support ≥ 80%, UFboot ≥ 95%, SH-aLRT ≥ 80%, aBayes ≥ 0.90, parametric aLRT ≥ 0.95, and LBoot ≥ 90% (Anisimova et al., 2011).

The bootstrap and single branch tests assess the robustness of individual branch bipartitions and cannot directly compare complex alternate tree topologies. Therefore we used non-parametric topology tests to directly compare the inferred ML tree to alternative trees with the correct phylogenetic placement of armadillo, dog, marsupials (opossum and wallaby), and bat, as well as the correct species phylogeny (**Figure 2B**); tests included the BP-RELL, KH-test (Kishino et al., 1990; Kishino and Hasegawa, 1989), SH-test (Anisimova et al., 2011; Guindon et al., 2010), c-ELW (Strimmer and Rambaut, 2002), weighted KH- and SH-tests, and the AU-test (Shimodaira, 2002). We note that the KH-test compares two *a priori* defined trees rather than the ML and alternative trees (Goldman et al., 2000) and does not correct for multiple hypothesis tests, it is included solely for comparison to other methods and previous studies. The SH-test can be used to compare the ML tree to multiple alternative trees selected *a priori* (i.e., is dataset independent) and corrects for multiple hypothesis tests, but is too conservative when many trees

are tested. The AU-test, in contrast, resolves the conservative nature of the SH-test and thus is the preferred test.

Clustering methods

We evaluated multiple methods to summarize and visualize the binary encoded extant and ancestral reconstructed transcriptomes, including: 1) Logistic Principal Component Analysis (LPCA), a version of principal component analysis for dimensionality reduction of binary data; 2) classical Multi-Dimensional Scaling (MDS); 3) Uniform Manifold Approximation and Projection (UMAP); 4) tSNE; and 5) Fuzzy C-Means (FCM) clustering. All clustering analyses were conducted in R after removing columns (genes) with missing data (coded as ?/NA) or that were invariant (all 0 or all 1). LPCA was performed using the `logisticPCA` R package (Landgraf and Lee, 2015), which implements three methods: exponential family PCA applied to Bernoulli data, logisitic PCA, and the convex relaxation of logistic PCA. For each of the methods, we fit the parameters assuming two-dimensional representation, returning four principal components (ks=4), and selecting the best m value to approximate the saturated model for cross validation. MDS was performed using the `vegan` R package (Oksanen et al., 2008) with four reduced dimensions. UMAP was performed using the `umap` R package. tSNE was performed using the `Rtsne` R package.

To explore the data in greater detail we focused on FCM clustering because the results were qualitatively similar to the other methods, and it has several desirable properties including providing a statistically sound way to identify clusters rather than an *ad hoc* approach that might be applied to the other methods. FCM also allows each sample to have membership in multiple clusters and is conceptually similar to topic ("grade of membership") models used in population genetics to visualize private and shared genetic structure across populations. FCM membership coefficients can thereby account for multiple sources of similarity including noise, phylogenetic signal, and convergence of gene expression. FCM was performed in R using the `cluster` R package, using Manhattan distances (cluster membership was not altered by using other distance metrics), and an estimated fuzzifier (m=1.034978). FCM clustering requires *a priori* knowledge of the number of clusters (K) to include, therefore we evaluated FCM with K=2–9 following the suggestions given in https://www.r-bloggers.com/2019/01/10-tips-for-choosing-the-optimal-number-of-clusters/. First, we used the "elbow" method, in which the sum of squares of each cluster number is calculated and graphed and the optimal number of clusters estimated by a

change of slope from steep to shallow (the elbow). We also assessed the optimal number of clusters using the clustree R package, which assess the optimal number of clusters by considering how samples change groupings as the number of clusters increases; clustree is useful for estimating which clusters are distinct and which are unstable but cannot determine the optimal number of clusters (K).

Acknowledgements

References

Abzhanov A. 2015. Evolution of the Animal Face: from Principles to Mechanisms. *Faseb J* **29**. doi:10.1096/fasebj.29.1_supplement.867.1

Anisimova M, Gascuel O, Sullivan J. 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biol* **55**:539–552. doi:10.1080/10635150600755453

Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic Biol* **60**:685–699. doi:10.1093/sysbio/syr041

Blackburn DG. 2015. Evolution of vertebrate viviparity and specializations for fetal nutrition: A quantitative and qualitative analysis. *J Morphol* **276**:961–990. doi:10.1002/jmor.20272

Blackburn DG. 1993. Histology of the late-stage placentae in the matrotrophic skink Chalcides chalcides (Lacertilia; Scincidae). *J Morphol* **216**:179–195. doi:10.1002/jmor.1052160206

Blackburn DG, Brandley M. 2015. Viviparous placentotrophy in reptiles and the parent-offspring conflict. *J Exp Zoology Part B Mol Dev Evol* **324**:532–548. doi:10.1002/jez.b.22624

Blackburn DG, Callard IP. 1997. Morphogenesis of placental membranes in the viviparous, placentotrophic lizard Chalcides chalcides (Squamata: Scincidae). *J Morphol* **232**:35–55. doi:10.1002/(sici)1097-4687(199704)232:1<35::aid-jmor2>3.0.co;2-n

Blackburn DG, Starck JM. 2015. Morphological specializations for fetal maintenance in viviparous vertebrates: An introduction and historical retrospective. *J Morphol* **276**:E1–E16. doi:10.1002/jmor.20410

Boddy AM, Abegglen LM, Pessier AP, Schiffman JD, Maley CC, Witte C. 2020. {Lifetime cancer prevalence and life history traits in mammals}. *Evol Medicine Public Heal*. doi:10.1093/emph/eoaa015

Carter AM, Enders AC. 2004. Comparative aspects of trophoplast development and placentation. *Reprod Biol Endocrin* **2**:46.

Chanderbali AS, Berger BA, Howarth DG, Soltis PS, Soltis DE. 2016. Evolving Ideas on the Origin and Evolution of Flowers: New Perspectives in the Genomic Era. *Genetics* **202**:1255–1265. doi:10.1534/genetics.115.182964

Chen C-F, Foley J, Tang P-C, Li A, Jiang TX, Wu P, Widelitz RB, Chuong CM. 2015. Development, Regeneration, and Evolution of Feathers. *Annu Rev Anim Biosci* **3**:1–27. doi:10.1146/annurev-animal-022513-114127

Corpa JM. 2006. Ectopic pregnancy in animals and humans. *Reproduction* **131**:631–640. doi:10.1530/rep.1.00606

Corso G, Delitala GM, Carcupino M. 2000. Uterine morphology during the annual cycle in Chalcides ocellatus tiligugu (Gmelin) (Squamata: Scincidae). *J Morphol* **243**:153–165. doi:10.1002/(sici)1097-4687(200002)243:2<153::aid-jmor4>3.0.co;2-q

Cui Y, Wang W, Dong N, Lou J, Srinivasan DK, Cheng W, Huang X, Liu M, Fang C, Peng J, Chen S, Wu S, Liu Z, Dong L, Zhou Y, Wu Q. 2012. Role of corin in trophoblast invasion and uterine spiral artery remodelling in pregnancy. *Nature* **484**:246–250. doi:10.1038/nature10897

Elliot MG, Crespi BJ. 2009. Phylogenetic Evidence for Early Hemochorial Placentation in Eutheria. *Placenta* **30**:949–967. doi:10.1016/j.placenta.2009.08.004

Elliot MG, Crespi BJ. 2006. Placental Invasiveness Mediates the Evolution of Hybrid Inviability in Mammals. *Am Nat* **168**:114–120. doi:10.1086/505162

Freyer C, Renfree MB. 2009. The mammalian yolk sac placenta. *J Exp Zoology Part B Mol Dev Evol* **312B**:545–554. doi:10.1002/jez.b.21239

Galtier N. 2001. Maximum-Likelihood Phylogenetic Analysis Under a Covarion-like Model. *Mol Biol Evol* **18**:866–873. doi:10.1093/oxfordjournals.molbev.a003868

Goldman N, Anderson JP, Rodrigo AG, Olmstead R. 2000. Likelihood-based tests of topologies in phylogenetics. *Systematic Biol* **49**:652–670. doi:10.1080/106351500750049752

Graham CH, Lala PK. 1991. Mechanism of control of trophoblast invasion in situ. *J Cell Physiol* **148**:228–234. doi:10.1002/jcp.1041480207

Griffith OW, Blackburn DG, Brandley MC, Dyke JUV, Whittington CM, Thompson MB. 2015. Ancestral state reconstructions require biological evidence to test evolutionary hypotheses: A case study examining the evolution of reproductive mode in squamate reptiles. *J Exp Zoology Part B Mol Dev Evol* **324**:493–503. doi:10.1002/jez.b.22614

Griffith OW, Dyke JUV, Thompson MB. 2013. No implantation in an extra-uterine pregnancy of a placentotrophic reptile. *Placenta* **34**:510–511. doi:10.1016/j.placenta.2013.03.002

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biol* **59**:307–321. doi:10.1093/sysbio/syq010

Hill JP. 1932. II. Croonian lecture. - the developmental history of the primates. *Philosophical Transactions Royal Soc Lond Ser B Contain Pap Biological Character* **221**:45–178. doi:10.1098/rstb.1932.0002

Hirasawa T, Kuratani S. 2015. Evolution of the vertebrate skeleton: morphology, embryology, and development. *Zoological Lett* **1**:2. doi:10.1186/s40851-014-0007-7

Hoang DT, Chernomor O, Haeseler A von, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**:518–522. doi:10.1093/molbev/msx281

Hughes RL, Hall LS. 1998. Early development and embryology of the platypus. *Philosophical Transactions Royal Soc Lond Ser B Biological Sci* **353**:1101–1114. doi:10.1098/rstb.1998.0269

Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A von, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**:587–589. doi:10.1038/nmeth.4285

Kihlström JE. 1972. Period of gestation and body weight in some placental mammals. *Comp Biochem Physiology Part Physiology* **43**:673–679. doi:10.1016/0300-9629(72)90254-x

Kishino H, Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol* **29**:170–179. doi:10.1007/bf02100115

Kishino H, Miyata T, Hasegawa M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J Mol Evol* **31**:151–160. doi:10.1007/bf02109483

Kshitiz, Afzal J, Maziarz JD, Hamidzadeh A, Liang C, Erkenbrack EM, Kim HN, Haeger J-D, Pfarrer C, Hoang T, Ott T, Spencer T, Pavličev M, Antczak DF, Levchenko A, Wagner GP. 2019. Evolution of placental invasion and cancer metastasis are causally linked. *Nat Ecol Evol* **3**:1743–1753. doi:10.1038/s41559-019-1046-4

Landgraf AJ, Lee Y. 2015. Dimensionality Reduction for Binary Data through the Projection of Natural Parameters. *arXiv.org* **stat.ML**.

Lewis PO, Olmstead R. 2001. A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. *Systematic Biol* **50**:913–925. doi:10.1080/106351501753462876

Luckett P. 1976. Cladistic Relationships among Primate Higher Categories: Evidence of the Fetal Membranes and Placenta. *Folia Primatol* **25**:245–276. doi:10.1159/000155719

Luckett WP. 1975. Phylogeny of the Primates, A Multidisciplinary Approach 157–182. doi:10.1007/978-1-4684-2166-8_7

Luckett WP. 1974. Comparative development and evolution of the placenta in primates. *Contrib Primatol* **3**:142–234.

Lyson TR, Bever GS. 2020. Origin and Evolution of the Turtle Body Plan. *Annu Rev Ecol Evol Syst* **51**:1–24. doi:10.1146/annurev-ecolsys-110218-024746

Maddison WP, Midford PE, Otto SP. 2007. Estimating a Binary Character's Effect on Speciation and Extinction. *Systematic Biol* **56**:701–710. doi:10.1080/10635150701607033

Marinić M, Mika K, Chigurupati S, Lynch VJ. 2021. Evolutionary transcriptomics implicates HAND2 in the origins of implantation and regulation of gestation length. *Elife* **10**:e61257. doi:10.7554/elife.61257

Martin RD. 2008. Evolution of Placentation in Primates: Implications of Mammalian Phylogeny. *Evol Biol* **35**:125–145. doi:10.1007/s11692-008-9016-9

Martin RD. 1969. The evolution of reproductive mechanisms in primates. *J Reprod Fertil Suppl* 49–66.

McGowen MR, Erez O, Romero R, Wildman DE. 2014. The evolution of embryo implantation. *Int J Dev Biol* **58**:155–161. doi:10.1387/ijdb.140020dw

Mess A, Carter AM. 2006. Evolutionary transformations of fetal membrane characters in Eutheria with special reference to Afrotheria. *J Exp Zoology Part B Mol Dev Evol* **306B**:140–163. doi:10.1002/jez.b.21079

Mika K, Marinić M, Singh M, Lynch VJ. 2021. Evolutionary transcriptomics implicates new genes and pathways in human pregnancy and adverse pregnancy outcomes. *Biorxiv* 2021.04.19.440514. doi:10.1101/2021.04.19.440514

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von HA. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**:1530–1534. doi:10.1093/molbev/msaa015

Mossman 1898- Harland W. (Harland Winfield). 1937. Comparative morphogenesis of the fetal membranes and accessory uterine structures, by Harland W. Mossman ... With twenty-four plates and twelve text-figures. Washington.

Nguyen L-T, Schmidt HA, Haeseler A von, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* **32**:268–274. doi:10.1093/molbev/msu300

Oksanen J, Kindt R, Legendre P, O'Hara B, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2008. vegan: Community Ecology Package.

Pagel M. 1999. The Maximum Likelihood Approach to Reconstructing Ancestral Character States of Discrete Characters on Phylogenies. *Systematic Biol* **48**:612–622. doi:10.1080/106351599260184

Portmann A. 1938. Die Ontogenese der Säugetiere als Evolutionsproblem. S. Karger.

Renfree M, Shaw G. 2013. eLS. doi:10.1038/npg.els.0001856

Renfree MB. 2010. Review: Marsupials: Placental Mammals with a Difference. *Placenta* **31**:S21–S26. doi:10.1016/j.placenta.2009.12.023

Shimodaira H. 2002. An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biol* **51**:492–508. doi:10.1080/10635150290069913

Soubrier J, Steel M, Lee MSY, Sarkissian CD, Guindon S, Ho SYW, Cooper A. 2012. The Influence of Rate Heterogeneity among Sites on the Time Dependence of Molecular Rates. *Mol Biol Evol* **29**:3345–3358. doi:10.1093/molbev/mss140

Strimmer K, Rambaut A. 2002. Inferring confidence sets of possibly misspecified gene trees. *Proc Royal Soc Lond Ser B Biological Sci* **269**:137–142. doi:10.1098/rspb.2001.1862

Swanson P, Skinner MK. 2018. Encyclopedia of Reproduction 1–2. doi:10.1016/b978-0-12-811899-3.20528-4

Turner W. 1876. Some general observations on the placenta, with special reference to the theory of evolution. *J Anat Physiol* 33–53.

Vogel P. 2005. The current molecular phylogeny of Eutherian mammals challenges previous interpretations of placental evolution. *Placenta* **26**:591–596. doi:10.1016/j.placenta.2004.11.005

Wagner GP, Kshitiz K, Levchenko A. 2020. Comments on Boddy et al. 2020: Available data suggest positive relationship between placental invasion and malignancy. *Evol Medicine Public Heal* **2020**:211–214. doi:10.1093/emph/eoaa024

Wildman DE, Chen C, Erez O, Grossman LI, Goodman M, Romero R. 2006. Evolution of the mammalian placenta revealed by phylogenetic analysis. *P Natl Acad Sci Usa* **103**:3203–3208. doi:10.1073/pnas.0511344103

Wislocki GB. 1929. On the placentation of primates, with a consideration of the phylogeny of the placenta. Carnegie Institution.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J Mol Evol* **39**:306–314. doi:10.1007/bf00160154

Yang Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol Biol Evol* **10**:1396–401. doi:10.1093/oxfordjournals.molbev.a040082
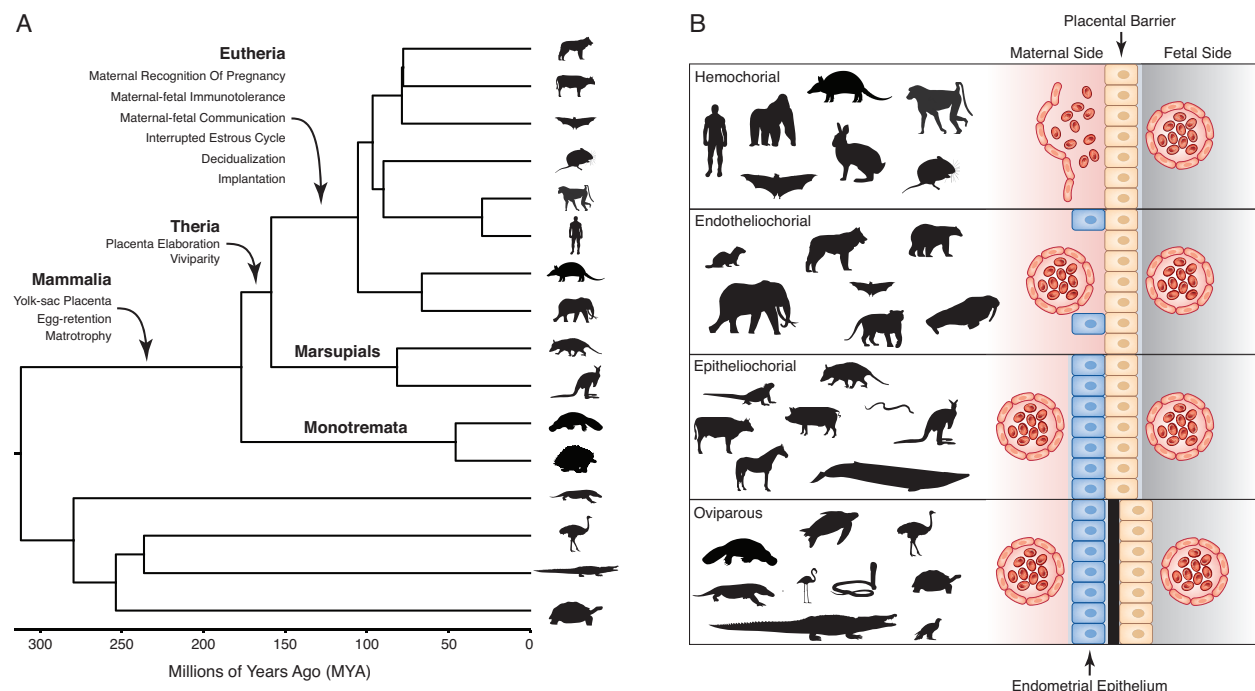
Figure 1. Evolution and arrangement of the maternal-fetal interface and degree of placental invasion in viviparous Amniotes.

(A) Phylogenetic relationships among amniote lineages, mammalian lineages labeled. Major evolutionary steps in the evolution of pregnancy are shown for Mammalia, Theria, and Eutheria.

(B) Placenta classification based on the arraignment of the maternal-fetal interface and degree of invasiveness. Epitheliochorial placenta, in which placental invasion does not occur and the barrier between maternal blood and the chorion (tan cells) consists of maternal vascular endothelium (red cells) and uterine epithelium (blue cells). Endotheliochorial placenta, in which the placenta invades through the uterine epithelium and the barrier between maternal blood and the chorion consists of the maternal vascular endothelium. Hemochorial placenta, in which the placenta invades through both the uterine epithelium and vascular endothelium, and maternal blood directly bathes the chorionic villi. Examples of species with each type of placenta are also shown. While the egg passes through the glandular uterine portion of the oviduct in oviparous species, the immunologically inert shell (black) prevents direct contact between maternal and fetal tissues.
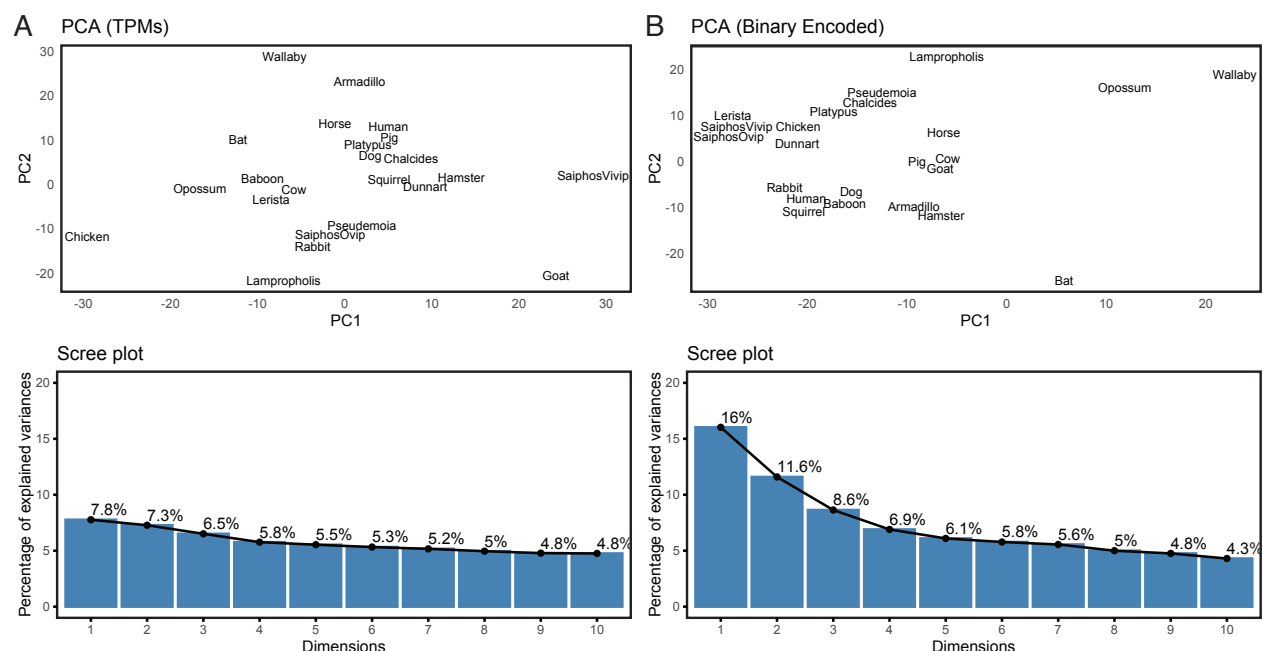
18

Figure 2. Binary encoding uncovers phylogenetic signal in transcriptome data.

(A) Upper, Principal Component Analysis (PCA) of gene expression levels (TPMs) grouped species randomly, consistent with significant noise. Lower, scree plot showing percent of variance explained by dimensions 1-10. Note that the flat curve of variance explained indicates there are no significant dimensions to the PCA.

(B) Upper, logistic PCA of the binary encoded endometrial transcriptome dataset groups species by phylogenetic relatedness, indicating significant noise reduction in the binary encoded dataset. Lower, scree plot showing percent of variance explained by dimensions 1-10. Note that the "elbow" of the graph is around dimension 4, suggesting that dimensions 1-4 of the PCA are significant.

Figure 2–Source data file 1. Species names (common and binomial), genome annotations used for RNA-Seq analysis and parity mode.

Figure 2–Source data file 2. Binary encoded endometrial transcriptome dataset.
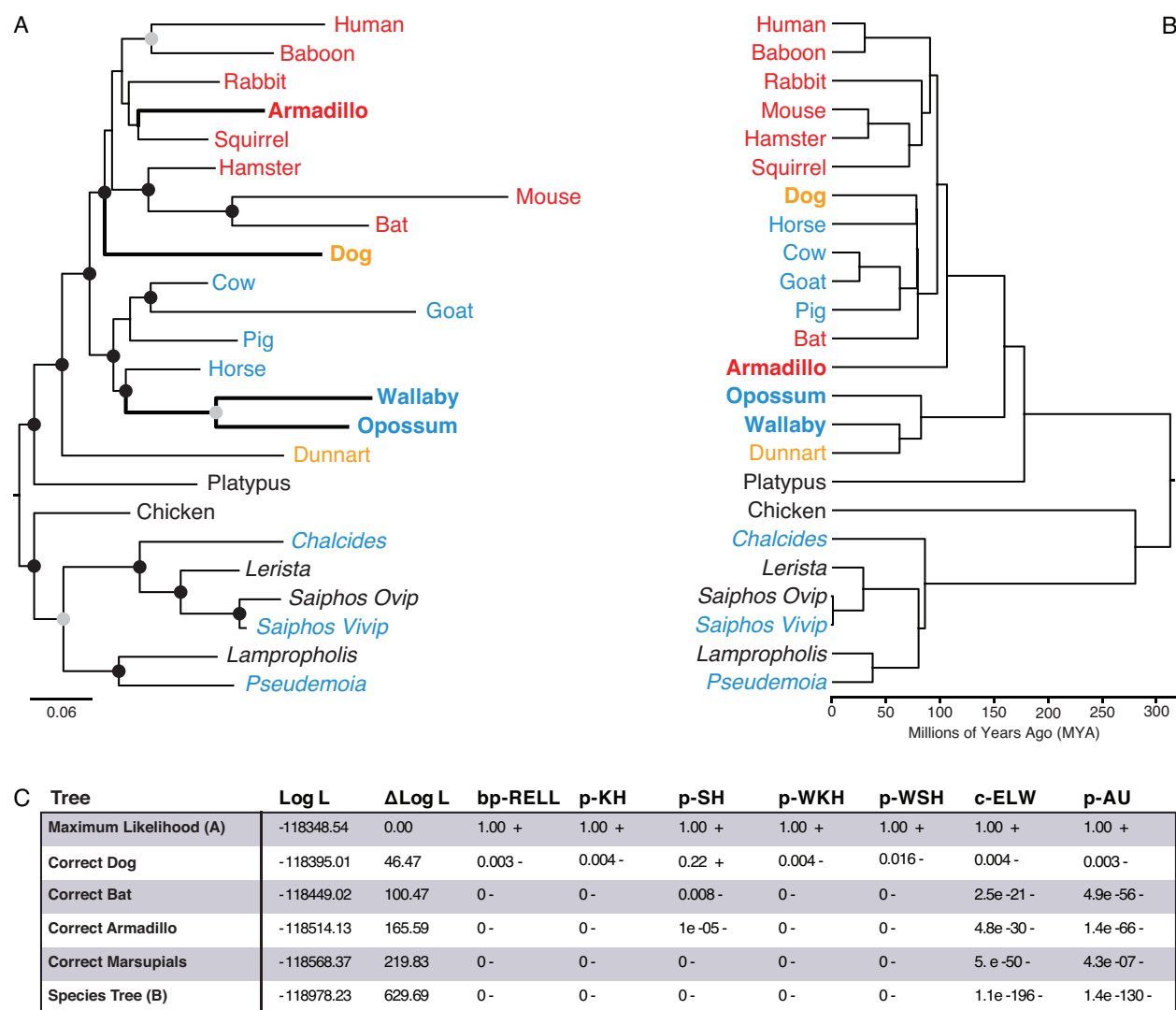
Tree	Log L	ΔLog L	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
Maximum Likelihood (A)	-118348.54	0.00	1.00 +	1.00 +	1.00 +	1.00 +	1.00 +	1.00 +	1.00 +
Correct Dog	-118395.01	46.47	0.003 -	0.004 -	0.22 +	0.004 -	0.016 -	0.004 -	0.003 -
Correct Bat	-118449.02	100.47	0 -	0 -	0.008 -	0 -	0 -	2.5e -21 -	4.9e -56 -
Correct Armadillo	-118514.13	165.59	0 -	0 -	1e -05 -	0 -	0 -	4.8e -30 -	1.4e -66 -
Correct Marsupials	-118568.37	219.83	0 -	0 -	0 -	0 -	0 -	5. e -50 -	4.3e -07 -
Species Tree (B)	-118978.23	629.69	0 -	0 -	0 -	0 -	0 -	1.1e -196 -	1.4e -130 -

Figure 3. Maximum-likelihood (ML) phylogeny of binary encoded endometrial transcriptome data.

(A) ML phylogeny of binary encoded endometrial transcriptome data inferred by IQ-TREE under the GTR2+FO+R3 model. Highly-supported branch splits, i.e., those with SH-aLRT ≥ 80%, LBoot ≥ 90%, aBayes ≥ 0.90, UFboot ≥ 95%, StdBoot support ≥ 80%, and parametric aLRT ≥ 0.95, are shown with black circles. Branch splits that are highly-supported by at least 4 of 6 methods are shown with gray circles. Particularly discordant phylogenetic relationships are shown in bold. Oviparous species are in black, viviparous species with epitheliochorial placentas in blue, endotheliochorial placentas in orange, and hemochorial placentas in red. Lizard species are shown with genus names in italics.

(B) Phylogenetic relationships of species in (A). Discordant phylogenetic relationships compared to the ML tree are shown in bold.

(C) Tree topology tests comparing the maximum likelihood tree shown in panel A to alternate trees that correct the phylogenetic placement of specific species ("Correct") or the species tree shown in panel B. Delta Log L: log L difference from the maximal Log L in the set (A). bp-RELL: bootstrap proportion using RELL method. p-KH: p-value of one sided Kishino-Hasegawa test (1989). p-SH: p-value of Shimodaira-Hasegawa test. p-WKH: p-value of weighted KH test. p-WSH: p-value of weighted SH test. c-ELW: Expected Likelihood Weight. p-AU: p-value of approximately unbiased (AU) test. Plus sign denotes the 95% confidence sets. Minus sign denotes significant exclusion. All tests performed 100000 resamplings using the RELL method.
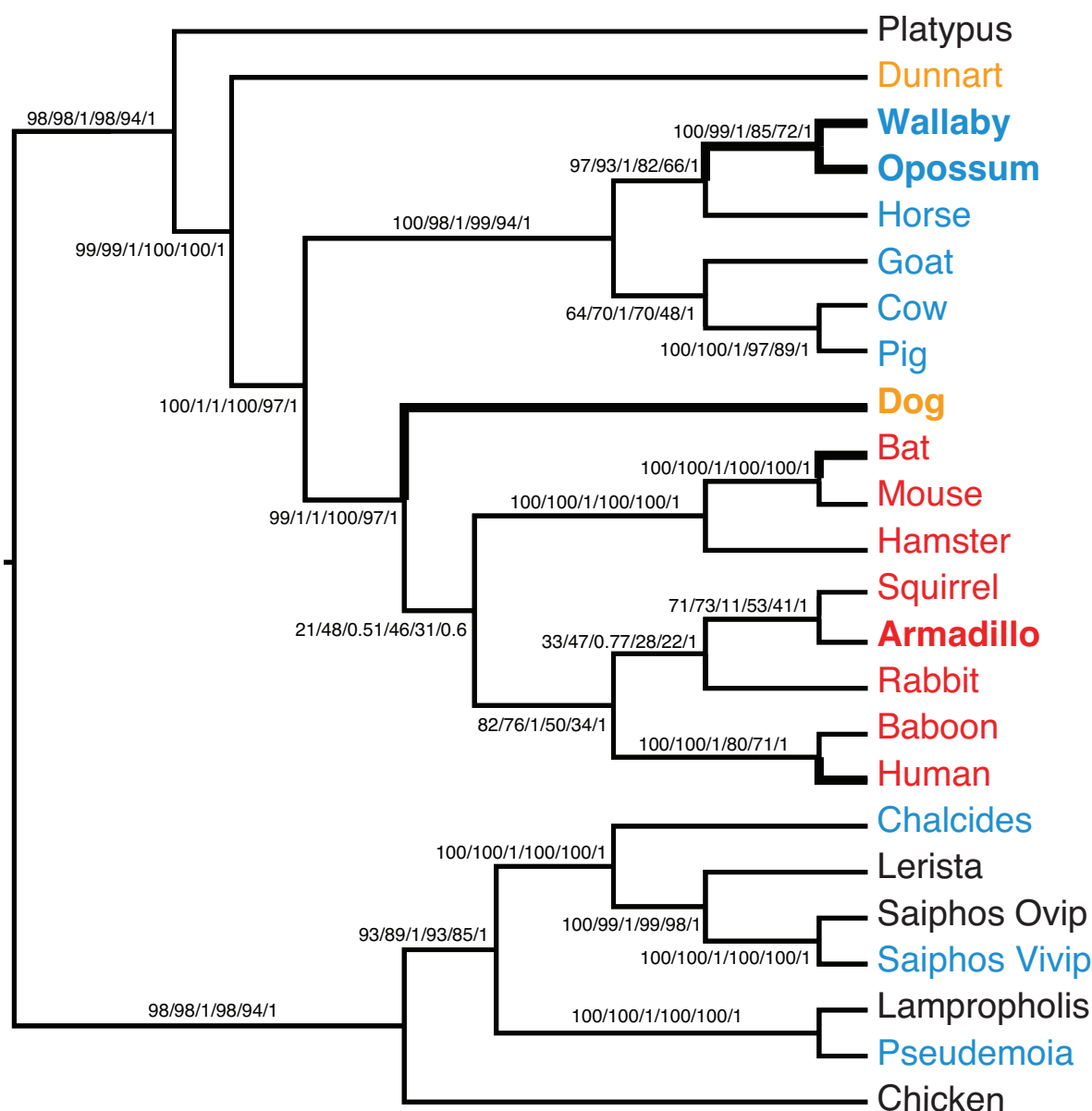
SH-aLRT (%) / LBoot (%) / aBayes / UFBoot (%) / StdBoot (%) / parametric aLRT

Figure 3–figure supplement 1. ML phylogeny of binary encoded endometrial transcriptome data inferred by IQ-TREE under the best-fitting GTR2+FO+R3 model. Branch supports are shown for SH-aLRT (%) / LBoot (%) / aBayes / UFBoot (%) / StdBoot (%) / parametric aLRT. Clades are considered highly-supported if its SH-aLRT ≥ 80%, LBoot ≥ 90%, aBayes ≥ 0.90, UFboot ≥ 95%, StdBoot support ≥ 80%, and parametric aLRT ≥ 0.95. Particularly discordant phylogenetic relationships are shown in bold. Oviparous species are in black, viviparous species with epitheliochorial placentas in blue, endotheliochorial placentas in orange, and hemochorial placentas in red. Lizard species are shown with genus names in italics.
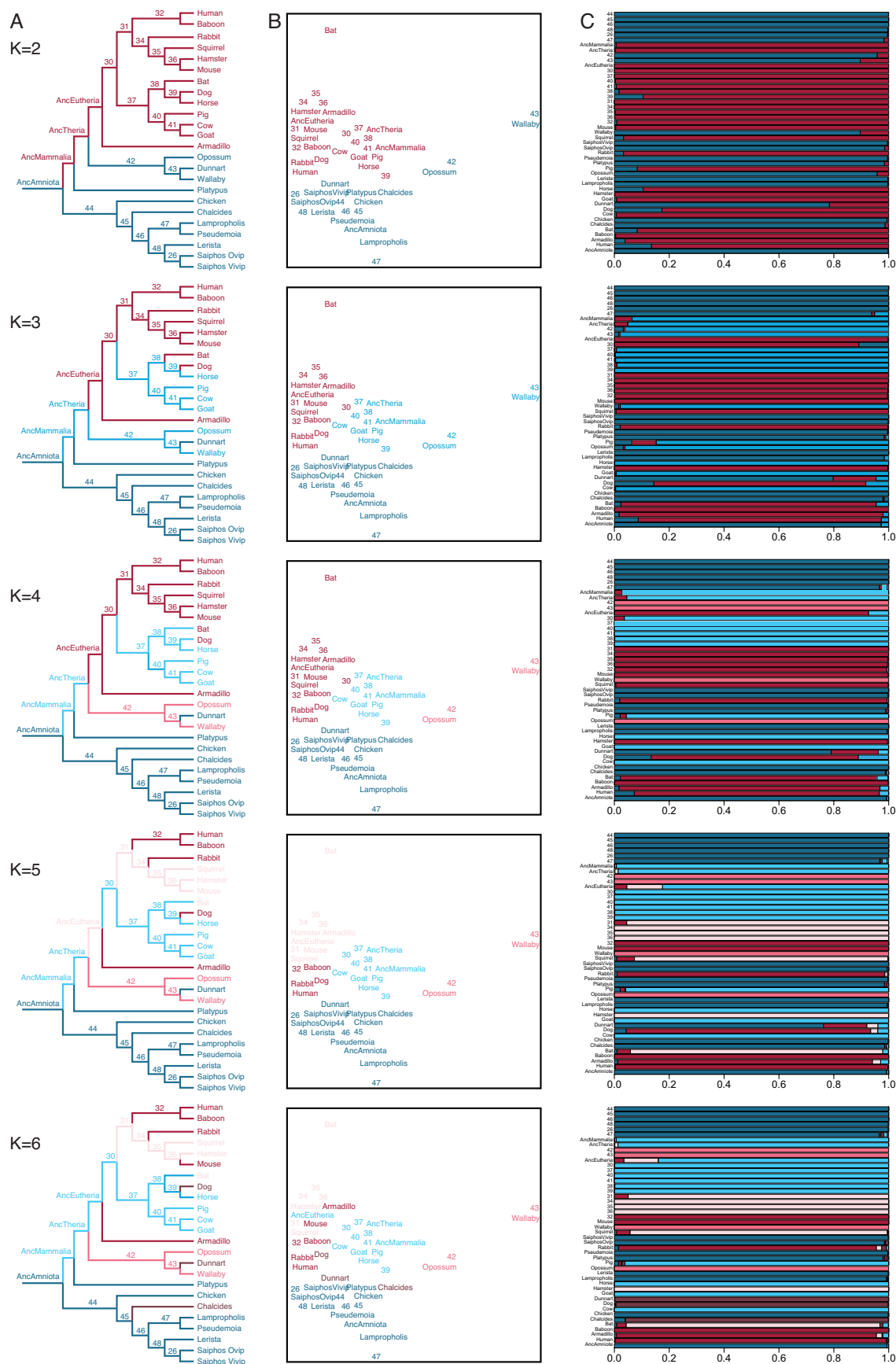
22

Figure 4. Fuzzy C-Means clustering of extant and ancestral transcriptomes.

(A) Species phylogeny with extant species and ancestral lineages colored according to Fuzzy C-Means cluster membership (K=2–6) shown in (B). Note that ancestral nodes are numbered. Lizard species are shown with genus names in italics.

(B) Fuzzy C-Means clustering of extant and ancestral transcriptomes (K=2–6). Extant species and ancestral lineages colored according to maximum degree of cluster membership shown in (C). Lizard species are shown with genus names in italics.

(C) Fuzzy C-Means cluster membership of extant and ancestral transcriptomes (K=2–6). Degree of cluster membership is shown as a 100% stacked bar and colored according to proportion of membership in each cluster. Lizard species are shown with genus names.
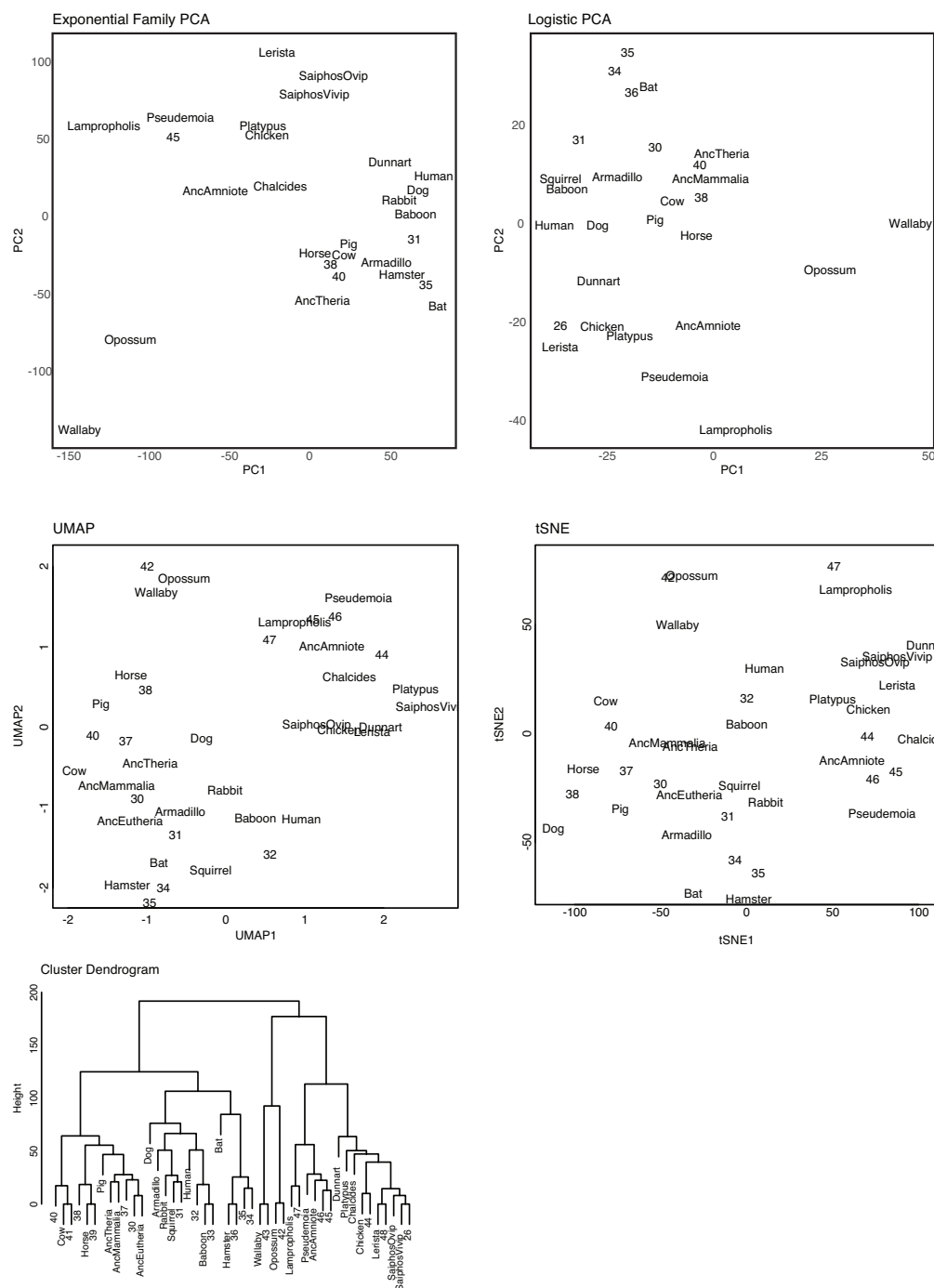
Figure 4–figure supplement 1. Exploratory clustering of extant and ancestral binary encoded transcriptomes, including: 1) Exponential Family Principal Component Analysis (PCA); 2) Logistic Principal Component Analysis (LPCA); 3) Uniform Manifold Approximation and Projection (UMAP); 4) t-distributed stochastic neighbor embedding (t-SNE); and 5) Hierarchical clustering (Manhattan distances).
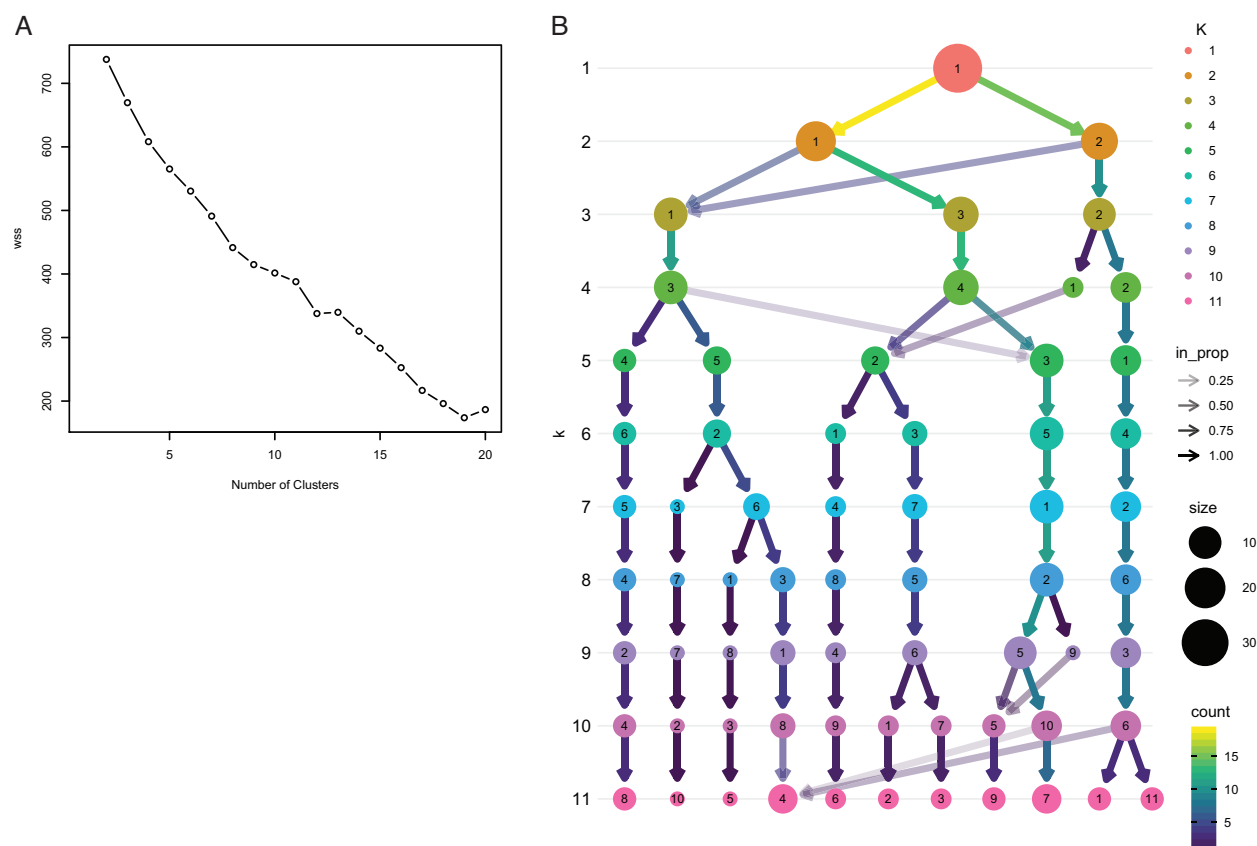
Figure 4–figure supplement 2. Exploratory analyses of Fuzzy C-Means cluster number.

(A) FCM cluster number estimate using the elbow plot method with K=1–20, which was uninformative in choosing the optimal number of clusters.

(B) Exploratory analyses of Fuzzy C-Means cluster number using the clustree method with K=1–11. In this figure the size of each node corresponds to the number of samples in each cluster, and the arrows colored according to the number of samples in each cluster. Transparent arrows (the incoming node proportion), show how samples from move between lower and higher cluster numbers and can be used as an indicator of cluster instability. In this graph samples switch clusters between K=1–5, are stable between K=6–9, and unstable with more than 9 clusters. These data indicate that 6 clusters is optimal for resolution, i.e., more than 6 clusters no longer provides informative cluster memberships, while more than K≥9 is over-clustered.

26