ANNUAL REVIEWS

Annual Review of Animal Biosciences

Phylogenomics and the Genetic Architecture of the Placental Mammal Radiation

William J. Murphy,[1] Nicole M. Foley,[1] Kevin R. Bredemeyer,[1] John Gatesy,[2] and Mark S. Springer[3]

[1]Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas 77843, USA; email: wmurphy@cvm.tamu.edu

[2]Division of Vertebrate Zoology, American Museum of Natural History, New York, NY 10024, USA

[3]Department of Evolution, Ecology and Organismal Biology, University of California, Riverside, California 92521, USA

Annu. Rev. Anim. Biosci. 2021. 9:29–53

First published as a Review in Advance on November 23, 2020

The *Annual Review of Animal Biosciences* is online at animal.annualreviews.org

https://doi.org/10.1146/annurev-animal-061220-023149

Keywords

mammals, phylogeny, genome, structural variation, coalescence, introgression

Abstract

The genomes of placental mammals are being sequenced at an unprecedented rate. Alignments of hundreds, and one day thousands, of genomes spanning the rich living and extinct diversity of species offer unparalleled power to resolve phylogenetic controversies, identify genomic innovations of adaptation, and dissect the genetic architecture of reproductive isolation. We highlight outstanding questions about the earliest phases of placental mammal diversification and the promise of newer methods, as well as remaining challenges, toward using whole genome data to resolve placental mammal phylogeny. The next phase of mammalian comparative genomics will see the completion and application of finished-quality, gapless genome assemblies from many ordinal lineages and closely related species. Interspecific comparisons between the most hypervariable genomic loci will likely reveal large, but heretofore mostly underappreciated, effects on population divergence, morphological innovation, and the origin of new species.

INTRODUCTION

Placental mammals are the crown clade of eutherian mammals and comprise the vast majority of living mammals, with more than 6,100 species estimated to exist and possibly thousands more cryptic lineages that await formal description as additional species (1). Placentals are remarkable in their sheer diversity of size and form: from 190-ton deep sea–diving whales to 2-g bats with powered flight, wooly mammoths to scaled anteaters, and blind subterranean moles to tarsiers with eyes as large as their brain. Scientists have long struggled to untangle the deep phylogenetic relationships among the morphologically diverse ordinal clades of mammals, as well as the relationships within recent adaptive radiations. In his landmark treatise on the classification of mammals, George Gaylord Simpson (2, p. 5) presciently argued, "The stream of heredity makes phylogeny; in a sense, it is phylogeny. Complete genetic analysis would provide the most priceless data for the mapping of this stream." Although he lamented at the time that the use of genetic characters was "an impossible goal," 75 years later Simpson would no doubt have reveled in the knowledge that hundreds of mammalian genomes have now been sequenced that sample every major taxonomic lineage (3). At the current pace at which new genomes are being generated, we will likely have (nearly) complete genetic blueprints for most living and numerous extinct species before the end of this decade (4).

This steep trajectory was fueled by two major revolutions that reshaped the course of scientific discovery. The first was the completion of the Human Genome Project (5), and with it the realization that comparisons between the genes and genomes of multiple divergent species would enlighten the function and evolution of the human genetic code. The National Institutes of Health's prioritization of sequencing the genomes of dozens of placental mammals in the early twenty-first century was informed by, and in turn stimulated further advances in, our understanding of placental mammal phylogeny (6–8). The second revolution was the precipitous drop in the cost per genome afforded by the advent of massively parallel DNA sequencing. The expertise required to accurately assemble ~90% of a mammalian genome now moved from a few genome centers to individual laboratories across the world. This change shifted the emphasis from sequencing whole genomes motivated by economic or human and animal health benefits to discovering the genetic underpinnings for adaptation, speciation, and trait evolution (9–12).

The emergence of mammalian whole genome alignments with dozens to hundreds of taxa (3) offers unparalleled power to finally resolve myriad phylogenetic controversies that linger in the scientific literature (13). These questions remain not so much because of a lack of data but because of variation in genome quality and gene annotation, accurate ascertainment of orthologs, and other systematic and analytical errors that produce gene tree errors (14). The longstanding hope of the molecular systematics community was that "if you build it, they will come": More specifically, with whole genomes, we would maximize phylogenetic signal relative to background noise, and the "true branches of the tree" would readily emerge from the data. However, recent phylogenomic studies based on whole genome sequence alignments have tempered this notion by demonstrating a far greater complexity in both the extent of reticulation and the distribution of phylogenetic signal within the genomes of recently diverged species with histories of gene flow (15–18).

The pursuit of a resolved mammalian phylogeny at the species level is also motivated by a desire to understand the genomic innovations underpinning myriad adaptations that enable mammals to occupy the most diverse habitats across the world, as well as the precise genomic changes that lead to reproductive isolation of species (19). Indeed, mammals show striking examples of ecomorphological convergence, where phenotypically similar species evolved independently in distantly related clades. For example, two Old World clades (pangolins and aardvarks) and one New World clade (anteaters) exhibit extreme morphological specializations for ant and termite eating that

were acquired independently in these groups (20, 21). These cases of parallel adaptive evolution present an unprecedented opportunity to study the genetic underpinnings of convergent phenotypes (22). From a health perspective, phylomedicine (23) and more recently PhyloOncology (24), fields that operate at the intersection of phylogenomics and medicine, benefit from assessments of evolutionary constraint across the mammalian tree (25). The calculation of single-nucleotide conservation scores from genomic sequence alignments (26) of divergent mammalian taxa relies on accurate phylogenies and branch lengths and can be used to help differentiate between benign and deleterious mutations in clinical samples. Well-resolved phylogenies and biodiversity discovery are also essential tools that can help trace the origins of zoonotic disease outbreaks (27).

Here we discuss the challenges in inferring accurate phylogenies from whole genome data. We highlight outstanding questions about early placental mammal relationships and emphasize the promise of newer methods and approaches to resolve existing controversies across the placental mammal family tree with statistical certainty. Finally, we discuss how higher-quality genome assemblies will enhance the study of diversification and speciation within recent mammalian radiations. Interspecific comparisons between the most variable gene families and repetitive sequences may reveal previously underappreciated effects of these iterated sequences on population divergence and the origin of new species.

UNRESOLVED QUESTIONS IN EARLY PLACENTAL MAMMAL EVOLUTION

The development and widespread application of molecular genetic markers, first from the mitochondrion (28) and later from the nuclear genome (29, 30), sparked a revolution in our understanding of the mammalian tree of life. Prior to this molecular revolution, the precise hierarchy among many mammalian species, as well as the deeper ordinal branching history, was largely incorrect because systematists had to rely on morphological characters (31). Within mammals, morphological characters are often limited in their application because of pervasive ecomorphological convergence and correlated character evolution across deep spans of time (32, 33). As such, early morphological character-based phylogenies included spurious clades of morphologically similar species that are not each other's closest relatives [e.g., Volitantia, a clade composed of colugos and bats (31)]. A key innovation that arose from the first molecular analyses with broad taxonomic and genetic sampling was the consistent identification of four superordinal clades: Afrotheria (e.g., elephants, aardvarks, and tenrecs) (34, 35), Xenarthra (e.g., armadillos, anteaters, and sloths), Laurasiatheria (e.g., carnivorans, bats, pangolins, perissodactyls, and cetartiodactyls), and Euarchontoglires (e.g., primates, treeshrews, rodents, and rabbits) (36–38) (**Figure 1**). These relationships have remained stable and well supported in nearly all subsequent studies based on DNA sequence data. Collectively, the nodes in the mammalian tree that have eluded resolution are products of rapid radiations that are characterized by closely spaced branching events. Below, we list several of the most challenging questions, resolution of which is important for understanding the early biogeographic history of Placentalia, deciphering the effects of the Cretaceous–Paleogene (KPg) mass extinction on ordinal diversification, and quantifying differential rates of character evolution.

The Root of Placentalia and Biogeography

Twenty years have passed since the redrawing and quartering of the placental mammal phylogeny, yet the resolution of the relationships at the base of Placentalia has remained elusive, even in the postgenomic era. Three hypotheses exist for the branching pattern at the root of living placental

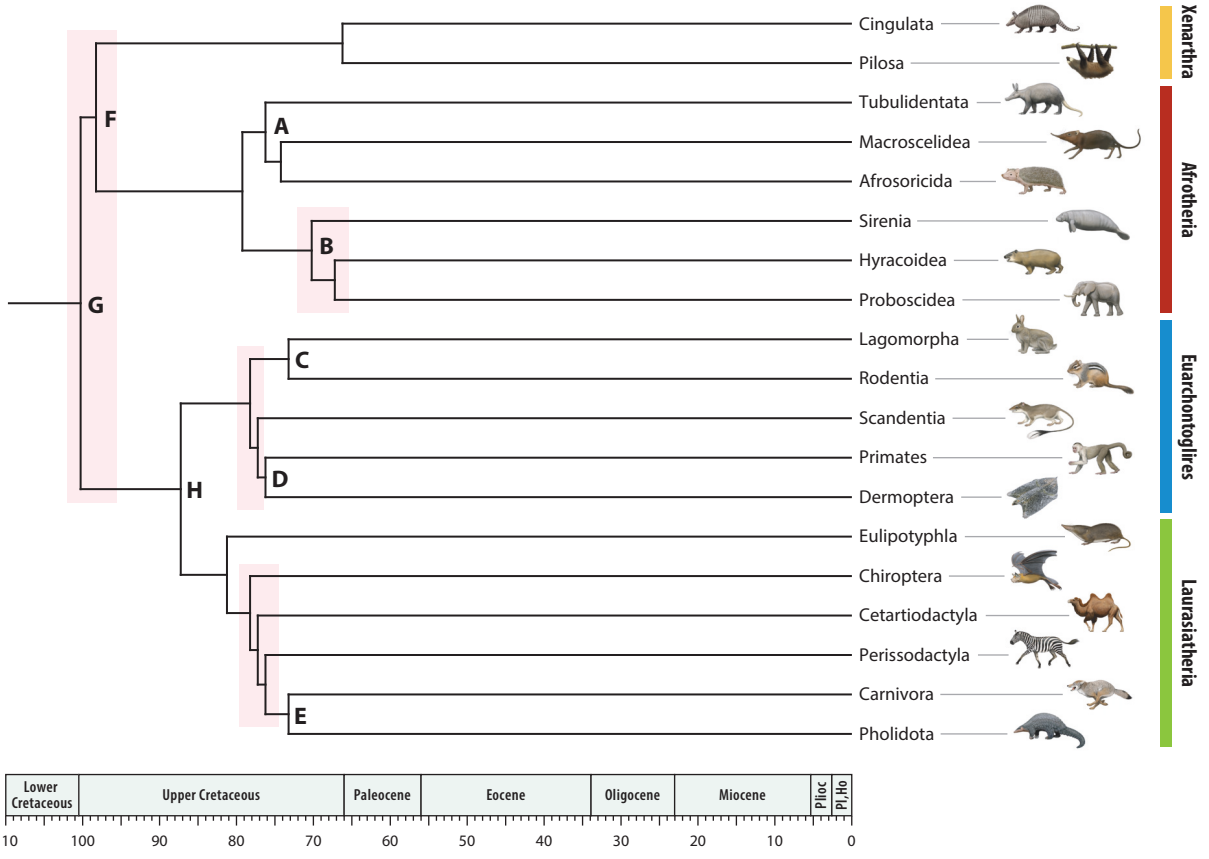

Figure 1

Phylogenetic relationships and timescale among placental mammal orders based on a consensus of published studies. The timescale is derived from Foley et al. (13), with minor modifications to be consistent with the shown topology. Nodes spanned by pink bars indicate relationships that remain controversial (see **Figure 2**). The four principal clades (defined in 33–35), shown to the right, are uncontroversial. Select superordinal clades highlighted in the text and used in **Figure 2** are as follows: (*A*) Afroinsectiphila, (*B*) Paenungulata, (*C*) Glires, (*D*) Primatomorpha, (*E*) Ferae, (*F*) Atlantogenata, and (*G*) Placentalia. Paintings are by C. Buell.

mammals: (*a*) Xenarthra versus Epitheria [i.e., Afrotheria + (Laurasiatheria + Euarchontoglires)], (*b*) Afrotheria versus Exafroplacentalia [i.e., Xenarthra + (Laurasiatheria + Euarchontoglires)], and (*c*) Atlantogenata (Afrotheria + Xenarthra) versus Boreoeutheria (Laurasiatheria + Euarchontoglires) (**Figure 2*a***). Elucidating these branching relationships, and their divergences times, is critical to understanding the biogeographic context of early placental mammal diversification. For example, some divergence time estimates for the origin of afrotherian lineages correlate well with a scenario in which they evolved in isolation in Africa following continental breakup but before land masses reconnected with other continents during the Cenozoic (34, 35, 39). Accurately recovering the original branching pattern, and determining if its timing matches plate tectonic models, would clarify what, if any, role continental breakup has played in the ordinal diversification of mammals, versus other biotic and abiotic factors (38, 40, 41).

Although each of the three hypotheses for the root of Placentalia found variable support in early molecular studies that sampled small numbers of nuclear and mitochondrial DNA loci, more recent studies that exploited information from draft whole genome sequences from all four major

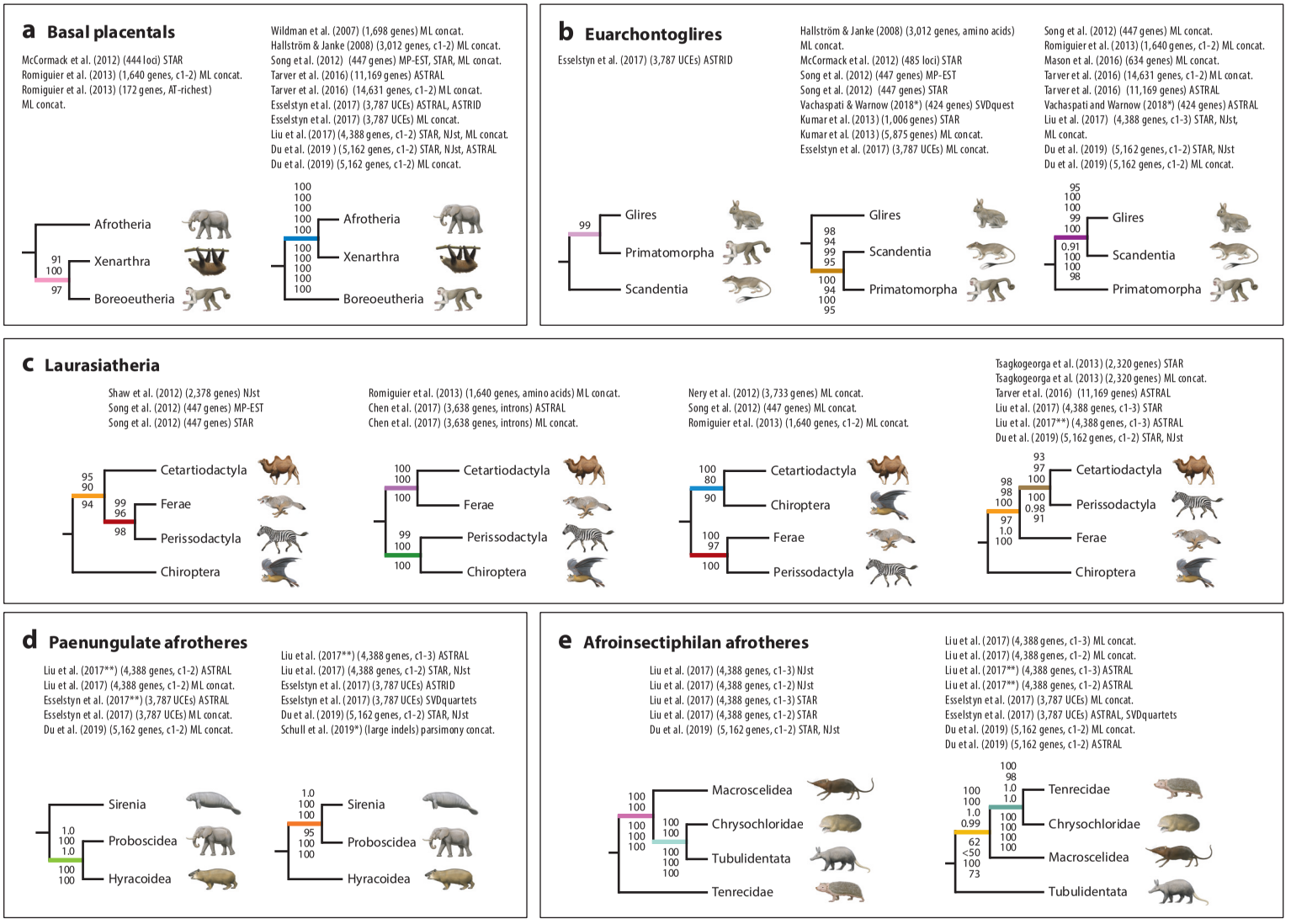

Figure 2

Controversial phylogenetic relationships between placental mammal ordinal/superordinal lineages. Studies that robustly support one or more relationships (depending on the type of analysis) are listed above each schematic tree. Support values (bootstrap or posterior probabilities) from each study are listed for each relevant sister-taxon relationship, in the order of the corresponding studies. Paintings are by C. Buell.

placental clades have typically favored either Atlantogenata or Exafroplacentalia (42–45). The only large-scale analysis to find support for Epitheria was a combined analysis of morphological and molecular data sets (46). The Atlantogenata hypothesis finds broad support from studies that sampled large numbers of taxa (19, 47) and genome-wide sets of protein-coding loci analyzed with concatenation- and coalescence-based methods (44, 50–52). GC nucleotide–rich regions of the genome are associated with higher gene-tree conflict than AT-rich regions, possibly because of incomplete lineage sorting (ILS) and GC-biased gene conversion (48, 49). The latter is thought to increase homoplasy (when the same nucleotide is gained, or lost, in independent evolutionary lineages) and lead to incorrect models of sequence evolution (model mis-specification) (51). It is therefore noteworthy that genome-wide phylogenies derived from AT-rich genes, as well as Ultra Conserved Element (UCE) data that are most commonly derived from AT-rich genomic regions, sometimes support the Exafroplacentalia hypothesis (45, 51, 53). However, even within the same studies, different treatments of the same data led to alternative support for different hypotheses. For example, Scornavacca & Galtier (52) recovered either the Atlantogenata or Exafroplacentalia hypothesis in their supertree analysis of protein-coding genes, depending on how branches with low support were treated in the input trees. Similarly, coalescence-based analysis of a large UCE data set supported Atlantogenata and Exafroplacentalia, depending on the underlying coalescence-based program used to conduct the analysis (53). By contrast, an analysis of retroposon insertions provided almost even support for the three competing hypotheses and slightly favored the Epitheria tree (54). In summary, the degree of conflict that has been demonstrated within and between phylogenomic data sets is a testament to the ongoing uncertainty surrounding the root of Placentalia.

Paenungulata and Other Afrotherian Affinities

The afrotherian superordinal clade Paenungulata, first recognized by Simpson (2), is composed of a morphologically diverse group of three orders that bear little living semblance to one another: manatees and dugongs (Sirenia), elephants (Proboscidea), and hyraxes (Hyracoidea) (**Figure 2**). There is strong molecular and morphological support for the monophyly of this clade (55). There is also strong morphological consensus on relationships between paenungulate orders, specifically the presence of numerous anatomical synapomorphies uniting Proboscidea and Sirenia in the clade Tethytheria (46, 56). Despite this consensus, clear support for paenungulate interrelationships has yet to emerge from molecular tree–building studies, which have variously supported all three possible topologies relating the clades to one another, in what has been described as a "seemingly unresolvable trichotomy" (57; see also 19, 47, 49, 53, 55) (**Figure 2**). Rare genomic changes have so far proved inconclusive: A limited-scale screen of retroposon insertions and derived chromosome rearrangements failed to identify characters unambiguously supporting either of the three topologies (58, 59). This is likely due to the rapid series of divergence events (<1–2 million years) that gave rise to the three extant lineages. Thus, it is surprising that so many morphological characters could have emerged in such a narrow span of time that would unite sirenians and proboscideans. If this sister-group relationship indeed proves to be true, it would suggest that the genomic changes underpinning these phenotypes might have arisen via structural variation (e.g., deletions, duplications, and more complex intra- and interchromosomal rearrangements) rather than via accumulation of point mutations. Whole genome sequence alignments and rare genomic changes may hold the key to identifying characters that can finally resolve the phylogeny of Paenungulata and the relationships between the remaining afrotherian orders (58, 60, 61). However, until very recently, the vast majority of sequenced mammalian genomes were biased toward species within Boreoeutheria, notably primates and domestic and companion animals.

Highly contiguous draft genome assemblies are not yet available for a similar diversity of living xenarthran or afrotherian species, and this should remain a high priority for the future.

Interordinal Relationships within Laurasiatheria

Laurasiatheria is one of two superordinal clades that comprise the more inclusive clade Boreoeutheria, lineages that most likely originated and rapidly diversified ~80–90 Mya throughout the northern supercontinent of Laurasia (47). This superorder is composed of six orders: Eulipotyphla (hedgehogs, moles, shrews, and solenodons), Carnivora (cats and dogs), Cetartiodactyla (camelids, pigs, ruminants, and whales), Perissodactyla (horses, tapirs, and rhinos), Pholidota (pangolins), and Chiroptera (bats) (**Figure 2**). Only two components of the branching pattern of this clade are consistently resolved in the literature: the basal position of Eulipotyphla and the sister-taxon relationship of Carnivora and Pholidota (19, 36, 38, 44, 45, 47–49, 51–53). The relationships among the remaining branches of the laurasiatherian tree are notoriously variable, with almost all possible arrangements observed among published data sets. Nonetheless, several studies support a basal position for Chiroptera relative to Fereuungulata [(Pholidota + Carnivora) + Cetartiodactyla + Perissodactyla] (37, 38), including analyses of retroposon insertions (62, 63), UCEs (45, 53), and a subset of studies analyzing large collections of protein-coding genes (49, 52, 64), as well as a phylogenomic analysis of both protein-coding genes and non-coding sequences (65).

The Phylogenetic Placement of Treeshrews

The position of the treeshrews (Scandentia) within Euarchontoglires remains challenging. Morphological analyses place treeshrews sister to colugos (Dermoptera) in the grandorder Euarchonta, along with Primates (46). However, a sister-group relationship between colugos and primates is strongly supported by protein-coding gene–based phylogenies, rare genomic changes inside protein-coding genes, and UCEs (47, 53, 66, 67) (**Figure 2**). Broad molecular support for Euarchonta has also been recovered from large molecular supermatrices, indels in protein-coding genes, genome-wide analyses of protein-coding regions, and UCEs (19, 38, 44, 45, 48, 53, 66, 68). However, some large molecular data sets have also recovered support for an alternative arrangement in which Scandentia is sister to Glires (rodents and lagomorphs) (47, 49, 51, 52).

KPg Extinction and Ordinal Diversification

Perhaps the single most contentious issue surrounding placental mammal phylogeny is the timing of both interordinal and intraordinal diversification relative to the KPg boundary, and whether the bolide impact and its catastrophic effects had any discernible role in triggering the diversification. At present, five models exist to describe different scenarios for the timing of the placental radiation with respect to the KPg boundary (reviewed in 33): (*a*) the Explosive model, which suggests that all extant lineages (superordinal and below) evolved in response to the niche space vacated by nonavian dinosaurs after the KPg boundary; (*b*) the Soft Explosive model, which depicts limited interordinal divergences before the KPg boundary followed by the bulk of interordinal and all intraordinal diversification after the KPg boundary; (*c*) the Long Fuse model, which posits that the initial interordinal diversification of mammals began when there was an increased abundance of insects and flowering plants during the Cretaceous Terrestrial Revolution, whereas the majority of intraordinal lineages emerged afterward; (*d*) the Short Fuse model, which posits a pulse of interordinal and some intraordinal diversification that preceded the KPg boundary, a slowdown in diversification rates across the KPg boundary, and an upsurge in diversification rates after the KPg that was catalyzed by the Early Eocene

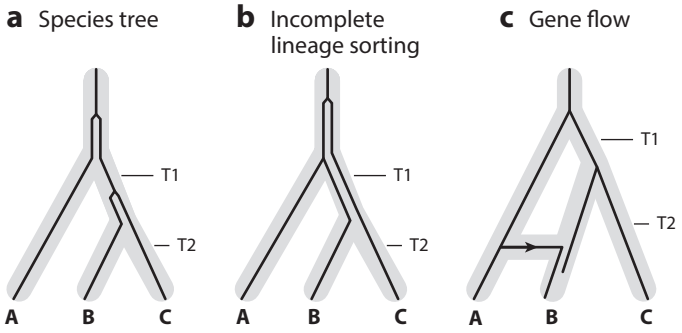

Figure 3

Major biological sources of phylogenomic discordance. (*a*) A phylogeny showing species trees for three taxa, coalescence of a gene tree in the most recent common ancestor of taxa B and C, and the ancestor of B + C and taxon A. (*b*) A phylogeny showing the same species tree for three taxa and deep coalescence of a gene for which the gene tree does not match the species tree. (*c*) A phylogeny showing the same species tree for three taxa and introgression from taxon A into taxon B, resulting in a gene tree that is inconsistent with the species tree.

Climatic Optimum; and similarly, (*e*) the Trans-KPg model, which also minimizes the role of the KPg boundary in mammalian diversification, suggesting that intraordinal and interordinal diversification were uninterrupted by the mass extinction event. A resolved phylogeny with accurate branch lengths, and by extension an accurate timescale, remains central to testing these hypotheses. Future progress in this area will likely be driven by several factors: (*a*) more contiguous and accurate genome alignments that improve upon detection of orthologous sequences (69), (*b*) an improved understanding of which genomic regions yield more accurate branching relationships and hence more precise estimates of divergence time (see below), (*c*) improvements in the calibration of nodes with fossils, and (*d*) improvements in relaxed clock methodologies.

PHYLOGENOMIC CONFLICT AND APPROACHES FOR RESOLVING COMPLEX EVOLUTIONARY RELATIONSHIPS

Sources of Phylogenomic Conflict

Difficulties in resolving phylogenetic relationships with multilocus data can arise from different factors, both biological and systematic/statistical (**Figure 3**). Assuming that sequence alignments include properly aligned orthologous sequences [a surprisingly untested assumption in many phylogenomic studies (14, 70)], variation in individual gene trees along a chromosome may arise from several sources. Homoplasy occurs when the same nucleotide occurs in distantly related species in a phylogeny owing to independent mutations or reversals, the probability of which increases with divergence time. A second source of gene tree discordance is ILS, the failure of lineages within a population to coalesce within their most recent common ancestor (71). ILS can occur within both ancient and recently diverged radiations of species, and hence similar multispecies coalescent (MSC) approaches can be applied at different taxonomic levels. Hybridization between distinct lineages/species, and subsequent introgression of alleles via backcrossing to one of the parental species, can also produce gene trees that are discordant with the species tree (i.e., the original cladogenic events). Distinguishing between the latter two can be accomplished via several statistical methods (72). One of the most commonly applied methods is the ABBA/BABA test (also referred to as Patterson's *D* statistic), originally developed by Green and colleagues (73) to detect human–Neanderthal gene flow. Newer variations on this same theme can accommodate more than

four taxa (74), can localize introgression along chromosomes (75), and can be performed without an outgroup on unrooted quartet trees (63).

Incomplete Lineage Sorting and the Concatenation versus Coalescence Debate

During the 1990s and 2000s, most efforts to elucidate phylogenetic relationships among placental mammals were based on analyses of concatenated data matrices that were compiled through the application of PCR (polymerase chain reaction) and Sanger sequencing to one, several, or tens of loci (7, 76). These analyses resulted in a mostly well-resolved tree for the orders and families of placental mammals. The more recent emergence of phylogenomic data sets with hundreds or thousands of loci, along with the development of new coalescence-based methods for inferring species trees, has resulted in a series of new studies that claim to resolve one or more of the outstanding questions discussed above. Unlike concatenation methods, which ignore the effects of ILS on species tree estimation, coalescence-based methods explicitly address the problem of ILS. This is important because concatenation analyses may result in an incorrect species tree if the true species tree has consecutive short branch lengths (in coalescent units) that reside in the anomaly zone. [In the anomaly zone, the most common gene tree(s) under ILS differs from the species tree.] At the same time, coalescence methods make their own assumptions (see below) and are not guaranteed to perform better than concatenation methods if these assumptions are violated.

The main categories of coalescence-based approaches for species tree estimation from multilocus sequence data include (*a*) methods such as *BEAST (77) and StarBEAST2 (78) that coestimate gene trees and species trees, (*b*) summary coalescence methods such as ASTRAL and MP-EST that estimate species trees from gene trees, and (*c*) SNP methods such as SVDquartets (79) and SVDquest (80) that estimate species trees directly from nucleotide site pattern frequencies. Among these approaches, methods such as *BEAST that co-estimate gene trees and species trees are the most computationally burdensome and therefore the least tractable for large phylogenomic data sets.

Many coalescence-based methods for species tree estimation are statistically consistent under the MSC, which includes the following key assumptions when species trees are inferred from sequence-based gene trees (81, 82):

1. There is free recombination between coalescence genes (c-genes) but no intralocus recombination within c-genes.
2. All gene tree heterogeneity results from ILS.
3. There is no selection, and sequences evolve neutrally.
4. Mating is random in each population (panmixia).
5. There is no interspecific gene flow (hybridization).

However, the statistical consistency of coalescence methods is not guaranteed if one or more of these critical assumptions are violated. Below, we briefly examine how violations of each of these assumptions can negatively impact species tree estimation with summary coalescence methods and sequence-based gene trees. Many of these problems (e.g., homology errors, gene flow) also impact the accuracy of concatenation methods.

Free recombination between c-genes but no intralocus recombination within c-genes. C-genes are the fundamental unit of analysis for summary coalescence methods that rely on sequence-based gene trees. Individual c-genes have their own unique histories, which are delineated by recombination breakpoints. If adjacent c-genes are concatenated into a single segment for gene tree estimation, then multiple genealogical histories are blended together into a single

mixed history. This "concatalescence" approach (83) is most problematic when complete protein-coding sequences are treated as individual c-genes because individual exons can be more than a megabase apart. Moreover, c-genes become smaller as more taxa are added to a data set because of the recombination ratchet and may shrink to a single nucleotide in length if enough taxa are added to a phylogenomic data set. In practice, however, very few coalescence studies have queried the underlying sequence data for recombination breakpoints. We recommend that this should be standard practice for the application of summary coalescence methods to segments of genomic DNA. Some authors have argued that species tree inference under the MSC is robust to intralocus recombination (84). However, this conclusion is paradoxical because in the most extreme case all c-genes would be merged into one enormous pseudo c-gene as in concatenation and fail in the anomaly zone (85).

All gene tree heterogeneity results from incomplete lineage sorting. Summary coalescence methods assume that all gene tree heterogeneity results from ILS, but this conclusion is not supported by empirical studies of placental mammal phylogeny (52). Instead, most gene tree heterogeneity in published coalescence studies of placental mammal phylogeny may be unrelated to ILS and instead result from other causes, such as long-branch misplacement, model mis-specification, missing data, arbitrary resolution of polytomies, and various issues with homology, including paralogous sequences, alignments of different exons to each other, and alignment of exons to introns (52, 86). Some studies have reported that ILS accounts for the vast majority of topological variation in gene trees based on the results of simulations, but these analyses are inherently circular and conflate ILS with other sources of gene tree heterogeneity (86).

Problems with long-branch misplacement have a long history in mammalian phylogenetics and are often intertwined with model mis-specification when the underlying sequence composition changes over time (nonstationarity) (87). These problems are especially acute with poor taxon sampling, but even with improved taxon sampling they cannot be completely avoided because of the deep split between placentals and their closest living outgroups (marsupials), extraordinary lineage-specific rate variation among living placental mammals, and the impossibility of subdividing long branches for placental orders that are monotypic for living species (i.e., Tubulidentata) or contain only a handful of living species that diverged relatively recently (e.g., Dermoptera). Misrooting problems in empirical studies of mammalian phylogeny have resulted in deep coalescence times for particular gene trees that exceed 100 million years, which is as deep as or deeper than most estimates for the crown age of Placentalia (86). Similarly, problems with homology can create artifactual branches on gene trees that exceed the credulity of ILS-related gene tree heterogeneity. Unfortunately, homology issues are prevalent in phylogenomic studies of placental mammal phylogeny that have employed complete protein-coding sequences (14, 70, 86). In part, these problems have resulted from incomplete assemblies and unreliable annotation pipelines. We recommend that orthologous protein-coding sequence alignments be extracted directly from multispecies chromosome alignments (where conserved synteny is a further guide) to avoid annotation issues that have negatively impacted the quality of orthologs when protein-coding sequences are instead predicted directly from assembled, unaligned genomes. Finally, some authors have recommended against manual inspection of alignments and gene trees, but quality control of phylogenomic data remains important for detecting problems that can invalidate the major conclusions of phylogenomic studies (88, 89).

No selection and panmixia. No selection and panmixia (random mating) are key assumptions of summary coalescence methods but are generally ignored in empirical studies. The assumption

of no selection (i.e., neutral evolution) is important because different kinds of selection can increase or decrease expected coalescence times relative to neutral evolution. Both positive selection and purifying selection have the effect of reducing ILS and therefore diminishing coalescence times. Balancing selection, in turn, can increase coalescence times. In both cases, selection can alter gene tree stoichiometry (the relative proportions of different gene tree topologies) and affect the results of summary coalescence analyses, including both topology and branch lengths. Unfortunately, most summary coalescence analyses of mammalian phylogeny have relied on protein-coding sequences and/or UCEs that are often under some type of strong selection: Protein-coding sequences are generally under purifying or sometimes positive selection, whereas UCEs have low rates of substitution because of strong purifying selection. Intronic sequences come closer to fitting the expectations of neutral evolution but become progressively more difficult to align with divergence time, especially between orders of placental mammals.

Under ILS with neutral sequence evolution, panmixia, and no interspecific hybridization, the rooted gene tree that agrees with the species tree for three species will have the highest frequency, and the two discordant trees will be equally frequent. By contrast, violation of the neutral evolution assumption can cause deviations in the frequency of the three gene trees if there are also differences in effective population size for different loci owing to different patterns of selection acting on subsets of genes (82). In extreme cases, one of the discordant trees can become the most frequent tree and cause summary coalescence methods (and concatenation) to fail even when there is no anomaly zone (82). This is a sobering conclusion for efforts to resolve deep divergences in placental mammal phylogeny that are associated with short internal branches.

No interspecific gene flow. Summary coalescence methods assume that there is no interspecific gene flow (introgression), but mounting evidence implicates introgression in diverse mammalian clades, including mysticetes (90), primates (91–93), proboscideans (94), and carnivorans (16, 95–98). Standard methods for phylogeny reconstruction, including both concatenation and coalescence approaches, are inadequate in these cases because they allow for cladogenesis but not reticulation. In the context of a genome-wide analysis, introgression can distort branch lengths as well as topological relationships when a set of genome-wide loci is analyzed with coalescence or concatenation. Here, phylogenetic network methods such as PhyloNet (99) and SNaQ (100) may be more appropriate, as they allow for both cladogenesis and reticulation. The fundamental challenges for network methods, aside from computational tractability, are determining the correct number of reticulation events without overfitting, as well as accounting for reticulation with extinct lineages, which greatly increases the complexity of the network approach to phylogenetic analysis.

Conclusions on the application of summary coalescence methods with sequence-based gene trees. Advocates of summary coalescence methods have rightly called attention to the problem of ILS for inferring species trees with concatenation. Some of these authors have also concluded that the incongruence introduced by concatenation methods is a major cause of the long-standing uncertainty in placental mammal phylogeny. However, claims for the inferiority of concatenation to summary coalescence analyses with sequence-based gene trees have proven false. As illustrated in **Figure 2**, competing coalescence studies have yielded different results for the same phylogenetic problem, and in some cases different summary coalescence methods have yielded contradictory results for the exact same data set. We conclude that the inconsistent results of concatenation for resolving difficult nodes in placental phylogeny have been equaled or even

surpassed by the incongruent results of summary coalescence analyses with sequence-based gene trees.

Coalescence analyses with retroelements. Retroelements are copy-and-paste transposons, such as short and long interspersed nuclear elements (SINEs and LINEs, respectively), that comprise significant fractions of mammalian genomes, e.g., 42.2% of the human genome (101). Retroelement insertions have emerged as powerful markers for resolving phylogenetic relationships but have traditionally been analyzed using approaches that do not accommodate ILS, such as maximum parsimony and distance methods. Recently, ILS-aware methods (ASTRAL_BP, SDPquartets) have been developed for species inference with retroelement insertions (63). Importantly, retroelement insertions satisfy assumptions of the MSC much better than sequence-based gene trees or SNPs. First, retroelement insertions are presence/absence events that are not subject to intralocus recombination. This property exempts retroelement insertions from problems with the recombination ratchet that plague sequence-based analyses. Second, sequence-based gene trees are susceptible to topological heterogeneity that is unrelated to ILS. By contrast, retroelement insertions occur predominantly at unique genomic locations and only rarely undergo precise excision. Conflicting retroelement insertions, when they occur, are almost always the product of ILS or introgression/gene flow rather than homoplasy (but see 102), as is often the case for nucleotide substitutions. Finally, retroelement insertions are regarded as largely neutral markers as assumed by the MSC. Indeed, the majority of retroelement insertions occur in safe-haven regions of the genome where they have no known functional or selective significance. For these reasons, retroelement insertions are promising alternatives to both sequence-based gene trees and SNPs for species tree inference in a coalescence-based framework. Other markers, such as nuclear DNA sequences of mitochondrial origin (or NUMTs), are also candidates for analysis with ILS-aware methods that were developed for retroelements.

Recombination Rate and Phylogenomic Signal

Meiotic recombination plays a critical role in shaping the genetic architecture of reproductive isolation and the distribution of phylogenomic signals. In eukaryotic genomes, recombination occurs nonrandomly along chromosomes, with rates being lower near the center of chromosomes, particularly larger chromosomes with long relative arm lengths (103). Smaller chromosomes have higher-than-average recombination rates compared with longer chromosomes, with a more uniform distribution of rates. Centromeres tend to suppress recombination but alone do not sufficiently explain the observed reduction near the center of larger chromosomes (103).

Recombination rate interacts with the effects of natural selection along chromosomes such that gene trees generated from regions with the highest recombination rates are more likely to possess signatures of hybridization, because the introgressed or foreign alleles become more effectively unlinked from targets of natural selection (104). By contrast, regions of low recombination are generally depleted in signatures of hybridization owing to stronger background selection and are typically enriched for the species tree. Recent phylogenomic studies of *Heliconius* butterflies (15, 17) and the cat family Felidae (16) showed a strong positive correlation between regional recombination rate and the frequency of gene trees that result from gene flow and conflict with the species tree (**Figure 4**). These findings indicate that recombination rate may be among the most reliable predictors of which gene trees best represent the species tree. In both butterflies and cats the species tree is notably enriched within the low-recombining Z and X chromosomes, respectively, suggesting that it is essential to partition chromosomes in phylogenomic analyses to identify local variation in topology and its interaction with recombination rate. Although these examples

Figure 4

The genomic architecture of phylogenomic signal in three divergent animal lineages. Karyotypes are shown for mosquitos [*Anopheles* (115)], butterflies [*Heliconius* (15)], and mammals [*Lynx* (16)]. Colors represent the densities of species trees (*orange*) or alternative gene trees (*green* or *blue*) along each chromosome and are meant to represent the general pattern along chromosomes; see each article for more detailed descriptions of gene tree variation. In *Anopheles*, the blue region on chromosome 2L indicates an inversion with a high density of a unique gene tree. Three conserved patterns are apparent across these three lineages. The first is that the number of loci that support the species tree (the density along each chromosome is shown in orange) is often a minority signal across the whole genome, which is frequently dominated by gene trees that reflect ancient and more recent bouts of introgressive hybridization. The second pattern is that the species tree is enriched on the X and Z chromosomes, which is consistent with the disproportionally large role that loci on these chromosomes play in reproductive isolation relative to autosomes (commonly referred to as the large X-effect) (122, 168). The third pattern is that the species tree is notably enriched in regions of low recombination across the genome. For example, in cat species, the species tree is most commonly observed in the vicinity of centromeres and other regions with locally suppressed recombination (169). *Heliconius* photo provided by N. Edelman.

come from relatively young species radiations, the same principles should apply to deeper evolutionary divergences but will require discovery of blocks of conserved synteny with historically low recombination rates across a phylogeny.

Both the butterfly and cat studies benefitted from recombination maps, resources that were at one point logistically impractical to generate for most species without large pedigrees. However, linkage disequilibrium methods (105, 106) applied to whole genome sequence data from many individuals have provided an alternative to infer broader-scale recombination maps, and with comparable accuracy (107). Newer linkage disequilibrium methods based on machine-learning approaches can infer recombination rates using genome sequences from just a few unrelated individuals (108). These recent advances suggest that recombination maps will soon be forthcoming

for numerous species that span the mammal phylogeny, allowing for a much broader application of recombination-aware phylogenomic methods. The identification of conserved patterns of recombination may be critical for resolving the placental mammal tree of life.

PHYLOGENOMICS AT THE SPECIES LEVEL: WHY FINISHED GENOMES MATTER

Genomic Architecture of Reticulation in Recent Adaptive Radiations

Whereas much of the mammalian systematics literature has focused on resolving early branching events in the placental mammal tree, relationships below the family level, and particularly within speciose clades, remain poorly resolved (19, 109). Recent studies that have applied large phylogenomic data sets to resolve species-level radiations have shown that these sections of the tree are just as troublesome to resolve as interordinal radiations and remain a major obstacle to precisely tracking the evolutionary history of genes and genomes. As mentioned earlier, one of the most important findings from this past decade is the pervasiveness of ancient and contemporary gene flow between what were otherwise thought to be reproductively isolated species (110). This awakening first occurred when the field of molecular systematics transitioned from mitochondrial to nuclear markers in the early 2000s. During this time, a tremendous amount of cytonuclear discordance became apparent (111, 112). Much of this conflict was initially attributed to ILS or systematic error, but the analysis of whole genome comparisons and new statistical methods that could distinguish between ILS and gene flow (72, 73, 113) has resulted in a sea change toward reconsidering the magnitude of the impact of interspecific gene flow on phylogenomic discordance and its role in the evolution of adaptive phenotypes (12, 114–118).

Divergence with gene flow confounds traditional methods for phylogenomic inference when majority-rule or democratic-vote decision-making processes (71) are employed. This applies to both coalescence and concatenation approaches. If the number of loci that act as barriers to gene flow are small or even moderate in number, then their signal, even if matching the most likely set of branching events, will be swamped out by the remainder of the genome where interspecific genetic exchange may have a minimal impact on fitness. Therefore, cases of lineage splitting that may have occurred early in the history of a clade (and could, for example, provide evidence of historical biogeographic processes) could be overwritten across most of the genome by more recent bouts of gene flow (e.g., owing to dispersal). As a result, those now rarer genetic signatures of earlier branching events are often attributed to the other systematic or biological sources of gene tree discordance discussed above. If one were to identify which gene trees across the genome were not the result of post-speciation gene flow, conceivably these could provide greater clarity into the genomic architecture of reproductive isolation (i.e., identify speciation genes).

Several predictable patterns indicate which genomic regions harbor the most likely species tree (see 119 for a nice discussion). One of these patterns is that introgressed segments are usually depleted on the Z and X sex chromosomes relative to autosomes (15, 16, 115, 120, 121) (**Figure 4**). Z- and X-chromosome gene trees are therefore more likely to retain the original branching patterns, as predicted by the large X-effect (122), and are also strongly correlated with local recombination rate (15–17). In placental mammals, recombination rates are lower on the X chromosome relative to autosomes, but significant rate variation is nonrandomly distributed along the chromosome (16). The pseudoautosomal region, which allows for pairing between the X and Y chromosomes during meiosis, has one of the highest recombination rates within the genome and is enriched with signals of introgression (16). By contrast, divergent boreoeutherian mammals share at least three multi-megabase recombination cold spots, each flanked by recombination hot spots (16) (**Figure 5**). The largest cold spot is upward of 40 Mb,

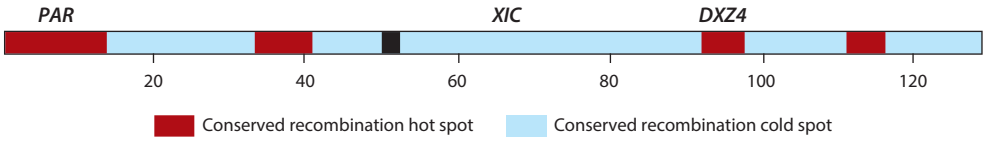

Figure 5

Conserved patterns of recombination rate across placental mammal X chromosomes. Alignments of recombination maps from placental mammals spanning three boreoeutherian orders (Primates: human; Carnivora: cat and dog; and Cetartiodactyla: pig) revealed conserved linkage of recombination cold and hot spots (16). Several functional features correspond to these regions, shown by their placement along the domestic cat X chromosome. The pseudoautosomal region (PAR) is responsible for pairing of the X and Y chromosomes during meiosis, and it possesses uniformly high rates of recombination. Three additional multi-megabase hot spots exist that flank the largest recombination cold spots. One of these contains the large tandem repeat macrosatellite, *DXZ4*, that is involved in formation of the bipartite structure during female X-chromosome inactivation (170). The X-inactivation center (*XIC*) encodes the noncoding RNA *XIST* and other loci that are required for initiation and maintenance of the inactive X chromosome in female somatic cells. The *XIC* is found centrally located within the largest recombination rate cold spot. This largest cold spot is delineated on either end by genes *JADE3* and *CHRDL1* in the cat genome. These conserved features likely existed in the ancestor of placental mammals.

lies within the center of the X chromosome, and cannot be explained by a centromere effect. The combined characteristics of very low historical recombination and high gene density (~500 genes) throughout this region resemble those of a supergene in many respects (123). The mega-cold spot harbors numerous targets for selection that would influence both female reproductive fitness, including a high density of *cis*-interacting genetic elements involved in X-chromosome inactivation as well as enrichment for testis-specific gene clusters. In addition, different evolutionary signatures are found in different clades and contexts: The same region is enriched for some of the highest genome-wide levels of genetic differentiation between closely related subspecies or species (16, 124, 125), whereas in other lineages it has been subject to selective sweeps and adaptive introgression (97, 126, 127). Future interrogation of phylogenomic signatures across placental mammal X chromosomes will likely be critical in resolving contentious phylogenetic relationships and understanding the genomic architecture of speciation. However, identifying species-specific genomic innovations and the targets of natural selection will, in many cases, require near-gapless chromosome assemblies that capture the most complex, rapidly evolving sequences that are currently absent in all but a handful of mammalian assemblies.

Repetitive DNA, Copy Number Variation, and Speciation

Approximately 40–60% of a placental mammalian genome is composed of repetitive elements, including retrotransposons and satellite repeat arrays, historically referred to as junk DNA (128). This DNA is not functionally inert but is embedded with numerous genes and gene families that are transcribed across many tissues. More than 99% of the 200+ mammalian genome assemblies in public databases are missing at least 5–10% of the euchromatic sequence that is enriched in long stretches of complex, highly repetitive DNA. The assembly of these long, high-identity, highly repetitive sequences [referred to as segmental duplications (129)] has been recalcitrant to all but the most cutting-edge sequencing approaches and is notably enriched on mammalian sex chromosomes. These regions have thus far been excluded from phylogenomic and comparative analyses owing to difficulties in assembly and alignment. Thus, this trove of genetic variation still remains dark matter in terms of its roles in normal mammalian biology as modulator of phenotypes, evolutionary novelty, phylogenomic signal, and barriers to gene flow between reproductively isolated species (130–132).

One central problem with capturing complex repetitive sequence content is diploidy. Because eukaryotic genome sequences are almost always displayed as a pseudohaploid representation

of the two parental haplotypes, genome assemblers fragment the sequence at sites of allelic variation between haplotypes. This results in numerous assembly gaps, and regions flanking the gaps are often artifactual representations of the original parental haplotypes. As a result, the most polymorphic and structurally divergent genomic regions, typically enriched in biological processes of sensory perception, immunity, and reproduction, are not available for annotation or comparative analyses. Furthermore, these gene classes have often been implicated in episodes of adaptive introgression (133).

Trio-binning is a powerful new assembly method that can be applied to any F1 hybrid to yield highly contiguous, single-haplotype genome assemblies of both parental genomes (134, 135). The beauty of this approach is that it removes the confounding issues of diploidy, opening the door to gapless genome assemblies given long enough sequence reads with high accuracy. Application of trio-binning to an F1 hybrid between a cow and a yak, two species that diverged ~5 Mya, yielded the two most continuous genome assemblies of any animal species to date, with almost one-third of the chromosomes lacking any gaps (135). This same approach produced similarly ultracontinuous genome assemblies for the domestic cat and the Asian leopard cat by sequencing an F1 Bengal cat hybrid (136). These bovid and cat assemblies resolved many complex repetitive sequences not present in earlier long-read draft assemblies, including the highly polymorphic major histocompatibility complex locus. Similar levels of genome continuity have been achieved recently for the human genome using the effectively haploid CHM1 cell line (137), including the first telomere-to-telomere chromosome assemblies (138, 139). As high-quality, gapless genomes become more widely available, they will provide pristine comparative genomic resolution, enabling researchers to discover the true extent to which species' genomes differ from one another (139, 140).

What impact does copy number variation of repetitive DNA have on population divergence and accumulation of barriers to gene flow? How does the absence of this sequence impact our understanding of speciation and lineage-specific innovation, as well as phylogenomic signal? Numerous studies in *Drosophila* have carefully documented the important role repetitive DNA plays in speciation (141–143). In mammals, much less is known about the evolution of the most repetitive genes within recent species radiations. Comparative studies in humans and great apes have provided the most comprehensive insights into the role of repetitive DNA in species-specific traits, because their genomes are always at the forefront of the newest genomic sequencing technologies (144–146). The earliest draft ape genome assemblies, based on short-read Illumina data or lower-coverage Sanger reads, were compared with the finished human reference sequence seeking to identify adaptive, species-specific changes that differentiated humans from other apes. However, many regulatory and copy number–mediated changes, as well as the discovery of novel human-specific genes (147–149), were only later identified with long-read sequencing and were altogether missed in these early studies because the first ape draft assemblies lacked complex structural architecture. For example, some of the most pronounced neural gene expression differences between human and chimpanzee are associated with inversions and other complex genomic loci that are prone to recurrent rearrangement (146). Long-read sequencing also allowed for novel insights into adaptive introgression during human evolution, including the first evidence of transfer of large and complex copy number variants from Neanderthal and Denisovan humans into Melanesian populations (150). These events are marked by selective sweeps, accelerated amino acid evolution, and the origin of new gene duplicates nearly absent from most human populations. These findings from just one small primate clade demonstrate that to fully appreciate and understand the genetic basis of mammalian phenotypic diversity and adaptive traits, we must fully resolve the most complex loci within the genomes of living species.

The Challenges of Sex Chromosome Assembly Limit Our Understanding of Speciation

Earlier, we mentioned that sex chromosomes are enriched with phylogenomic and population genomic signals consistent with barriers to gene flow, mirroring genetic mapping studies that have demonstrated a disproportionate density of X-linked major effect loci that are involved in hybrid dysfunction (17, 151–153). However, sex chromosome assemblies have the poorest contiguity and representation of all chromosomes owing to their enrichment with large, complex repeats (138, 154, 155). For example, X-linked mammalian hybrid sterility loci have been shown to broadly colocalize with repetitive sequences (156, 157); however, the specific genetic mechanisms connecting repeats to phenotypes are poorly described and understood. The inability to resolve and study this complexity has hampered the identification of the specific loci that underpin speciation (135). Given the well-documented importance of sex chromosomes to lineage divergence and speciation, high-quality, gapless interspecific chromosome alignments and complete gene annotations are essential to understanding the extent of functional and structural divergence in these processes (138, 156, 158, 159). In mouse and human, multi-megabase duplicated regions of high nucleotide identity (termed amplicons) were spanned on the sex chromosomes only by sequencing of BAC-tiling paths from single haplotypes, which is not practical for the majority of mammals (156, 158, 160, 161). More than 200 lineage-specific ampliconic genes were found to have emerged since the ancestors of human and mouse diverged nearly 85 Mya and are among the most divergent sequence classes between mammalian species (156). Extrapolating from this comparison, a minimum of several hundred genes spanning megabases of DNA sequence are most certainly missing from all but the few ultracontinuous mammalian draft assemblies. Therefore, it stands to reason that when characterizing and quantifying the rate of genomic divergence between closely and distantly related mammals, the resolution and inclusion of these functional sequences is of critical importance.

Sex chromosomes are by their nature genomic antagonists, and selfish genetic elements often arise and are maintained there. Sex linkage of selfish elements may distort sex ratios, leading to striking effects on population dynamics and genome evolution (162). An excellent example is provided by the *Slx*, *Sly* ampliconic gene families of mouse that span tens of megabases on the X and Y chromosomes, respectively (163). The massive expansion of these genes is unique to the laboratory mouse and its close relatives, having occurred during the past 3 million years (163). Both *Slx* (~100 gene family members in the mouse genome reference) and *Sly* (~80 gene family members) produce proteins that regulate gene expression in late stages of spermatogenesis (163). *Sly* expression represses XY gene expression during post-meiotic stages of spermatogenesis, whereas *Slx* increases XY expression. Intragenomic conflict manifests through rapid copy number divergence, with the relative amounts of gene product from each locus influencing proper sperm development, favoring transmission of either the X- or Y-containing sperm (164, 165). This antagonistic relationship suggests that a delicate balance in copy number between these gene families is required to maintain optimal sperm counts and organismal fitness. *Slx* and *Sly* are just two of many ampliconic gene families associated with the mouse sex chromosomes. Do similar complex gene families, and possibly other meiotic drive systems, exist in other mammalian species? Emerging evidence from companion animals suggests that sex chromosomes are hot spots of gene novelty and genomic innovation (166, 167). Resolution of these complex repetitive regions (genic and nongenic) across diverse branches of the mammalian tree will be essential for the identification and mechanistic understanding of loci that promote lineage divergence and reproductive isolation. They may also be the key to fully deciphering placental mammal phylogeny.

CONCLUSIONS

During the past 20 years, tremendous progress has been made toward reconstructing the complete phylogeny and evolutionary timescale of living mammals. Owing to the advances in high-throughput genome sequencing, catalyzed by the Human Genome Project, hundreds of mammalian genomes (of various qualities) are available, and we should assume that the majority of living species will be sequenced within the next 10 years. Whole genome data are not a panacea for phylogenomics, however, as emerging phylogenomic studies have demonstrated that majority-rule approaches may be misleading and that recovery of ancient branching events might be restricted to a minority of the genome. Understanding the roles of recombination and other aspects of chromosome architecture, as they relate to phylogenetic signal, will be critical if we are to fully interpret and utilize the information within whole genome sequences. In addition, we can expect a shift from gene-centric phylogenomic studies, which interrogate only a small fraction of the genome, to those that include the various classes of noncoding DNA and repetitive sequences. Finally, highly continuous genome assemblies will now enable researchers to explore the full spectrum of genetic diversity within the most complex and highly variable genomic regions, leading to a more mature understanding of the relationship between genomic divergence and speciation.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

W.J.M. is supported by National Science Foundation award DEB-1753760 and Morris Animal Foundation award D19FE-004. M.S.S. and J.G. are supported by National Science Foundation award DEB-1457735.

LITERATURE CITED

1. Burgin CJ, Colella JP, Kahn PL, Upham NS. 2018. How many species of mammals are there? *J. Mammal.* 99(1):1–14
2. Simpson GG. 1945. *The Principles of Classification and a Classification of Mammals*, Vol. 85. Bull. Am. Mus. Nat. Hist. New York: Am. Mus. Nat. Hist.
3. Zoonomia Consort. 2020. A comparative genomics multitool for scientific discovery and conservation. *Nature* 587:240–45
4. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, et al. 2018. Earth BioGenome Project: sequencing life for the future of life. *PNAS* 115(17):4325–33
5. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921
6. O'Brien SJ, Eizirik E, Murphy WJ. 2001. Genomics. On choosing mammalian genomes for sequencing. *Science* 292(5525):2264–66
7. Springer MS, Stanhope MJ, Madsen O, de Jong WW. 2004. Molecules consolidate the placental mammal tree. *Trends Ecol. Evol.* 19(8):430–38
8. Margulies EH, Vinson JP, NISC Comp. Seq. Program, Miller W, Jaffe DB, et al. 2005. An initial strategy for the systematic identification of functional elements in the human genome by low-redundancy comparative sequencing. *PNAS* 102(13):4795–800
9. Genome 10K Community Sci. 2009. Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. *J. Hered.* 100(6):659–74

10. Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346(6215):1320–31
11. McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol. Phylogenet. Evol.* 66(2):526–38
12. Ellegren H, Smeds L, Burri R, Olason PI, Backström N, et al. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 491(7426):756–60
13. Foley NM, Springer MS, Teeling EC. 2016. Mammal madness: Is the mammal tree of life not yet resolved? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371(1699):20150140
14. Springer MS, Gatesy J. 2018. On the importance of homology in the age of phylogenomics. *Syst. Biodivers.* 16(3):210–28
15. Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science* 366(6465):594–99
16. Li G, Figueiró HV, Eizirik E, Murphy WJ. 2019. Recombination-aware phylogenomics reveals the structured genomic landscape of hybridizing cat species. *Mol. Biol. Evol.* 36(110):2111–26
17. Martin SH, Davey JW, Salazar C, Jiggins CD. 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLOS Biol.* 17(2):e2006288
18. Bravo GA, Antonelli A, Bacon CD, Bartoszek K, Blom MPK, et al. 2019. Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. *PeerJ* 7:e6399
19. Upham NS, Esselstyn JA, Jetz W. 2019. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLOS Biol.* 17(12):e3000494
20. Springer MS, Meredith RW, Teeling EC, Murphy WJ. 2013. Technical comment on "The placental mammal ancestor and the post–K-Pg radiation of placentals." *Science* 341:613
21. Delsuc F, Metcalf JL, Wegener Parfrey L, Song SJ, González A, et al. 2014. Convergence of gut microbiomes in myrmecophagous mammals. *Mol. Ecol.* 23(6):1301–17
22. Emerling CA, Widjaja AD, Nguyen NN, Springer MS. 2017. Their loss is our gain: regressive evolution in vertebrates provides genomic models for uncovering human disease loci. *J. Med. Genet.* 54(12):787–94
23. Kumar S, Dudley JT, Filipski A, Liu L. 2011. Phylomedicine: an evolutionary telescope to explore and diagnose the universe of disease mutations. *Trends Genet.* 27(9):377–86
24. Somarelli JA, Ware KE, Kostadinov R, Robinson JM, Amri H, et al. 2017. PhyloOncology: understanding cancer through phylogenetic analysis. *Biochim. Biophys. Acta Rev. Cancer* 1867(2):101–8
25. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478(7370):476–82
26. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 20(1):110–21
27. Lam TT-Y, Jia N, Zhang Y-W, Shum MH-H, Jiang J-F, et al. 2020. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* 583:282–85
28. Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, et al. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *PNAS* 86:6196–200
29. Lyons LA, Laughlin TF, Copeland NG, Jenkins NA, Womack JE, O'Brien SJ. 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nat. Genet.* 15(1):47–56
30. Murphy WJ, O'Brien SJ. 2007. Designing and optimizing comparative anchor primers for comparative gene mapping and phylogenetic inference. *Nat. Protoc.* 2(11):3022–30
31. Novacek MJ. 1992. Mammalian phylogeny: shaking the tree. *Nature* 356(6365):121–25
32. Springer MS, Burk-Herrick A, Meredith R, Eizirik E, Teeling E, et al. 2007. The adequacy of morphology for reconstructing the early history of placental mammals. *Syst. Biol.* 56(4):673–84
33. Springer MS, Foley NM, Brady PL, Gatesy J, Murphy WJ. 2019. Evolutionary models for the diversification of placental mammals across the KPg boundary. *Front. Genet.* 10:1241
34. Springer MS, Cleven GC, Madsen O, de Jong WW, Waddell VG, et al. 1997. Endemic African mammals shake the phylogenetic tree. *Nature* 388(6637):61–64
35. Stanhope MJ, Waddell VG, Madsen O, de Jong WW, Hedges SB, et al. 1998. Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *PNAS* 95:9967–72

36. Madsen O, Scally M, Douady CJ, Kao DJ, DeBry RW, et al. 2001. Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409(6820):610–14
37. Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. 2001. Molecular phylogenetics and the origins of placental mammals. *Nature* 409(6820):614–18
38. Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, et al. 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294(5550):2348–51
39. Hedges SB. 2001. Afrotheria: Plate tectonics meets genomics. *PNAS* 98(1):1–2
40. Hedges SB, Parker PH, Sibley CG, Kumar S. 1996. Continental breakup and the ordinal diversification of birds and mammals. *Nature* 381(6579):226–29
41. Eizirik E, Murphy WJ, O'Brien SJ. 2001. Molecular dating and biogeography of the early placental mammal radiation. *J. Hered.* 92(2):212–19
42. Murphy WJ, Pringle TH, Crider TA, Springer MS, Miller W. 2007. Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res.* 17(4):413–21
43. Wildman DE, Uddin M, Opazo JC, Liu G, Lefort V, et al. 2007. Genomics, biogeography, and the diversification of placental mammals. *PNAS* 104(36):14395–400
44. Hallström BM, Janke A. 2008. Resolution among major placental mammal interordinal relationships with genome data imply that speciation influenced their earliest radiations. *BMC Evol. Biol.* 8:162
45. McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, Glenn TC. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res.* 22(4):746–54
46. O'Leary MA, Bloch JI, Flynn JJ, Gaudin TJ, Giallombardo A, et al. 2013. The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* 339(6120):662–67
47. Meredith RW, Janečka JE, Gatesy J, Ryder OA, Fisher CA, et al. 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* 334(6055):521–24
48. dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PCJ, Yang Z. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. Biol. Sci.* 279(1742):3491–500
49. Liu L, Zhang J, Rheindt FE, Lei F, Qu Y, et al. 2017. Genomic evidence reveals a radiation of placental mammals uninterrupted by the KPg boundary. *PNAS* 114(35):E7282–90
50. Tarver JE, dos Reis M, Mirarab S, Moran RJ, Parker S, et al. 2016. The interrelationships of placental mammals and the limits of phylogenetic inference. *Genome Biol. Evol.* 8(2):330–44
51. Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJP. 2013. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. *Mol. Biol. Evol.* 30(9):2134–44
52. Scornavacca C, Galtier N. 2017. Incomplete lineage sorting in mammalian phylogenomics. *Syst. Biol.* 66(1):112–20
53. Esselstyn JA, Oliveros CH, Swanson MT, Faircloth BC. 2017. Investigating difficult nodes in the placental mammal tree with expanded taxon sampling and thousands of ultraconserved elements. *Genome Biol. Evol.* 9(9):2308–21
54. Nishihara H, Maruyama S, Okada N. 2009. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *PNAS* 106(13):5235–40
55. Seiffert ER. 2002. The reality of afrotherian monophyly, and some of its implications for the evolution and conservation of Afro-Arabia's endemic placental mammals. *Afrotherian Conserv.* 1:3–6
56. McKenna MC, Bell SK. 1997. *Classification of Mammals: Above the Species Level*. New York: Columbia Univ. Press
57. Seiffert ER. 2007. A new estimate of afrotherian phylogeny based on simultaneous analysis of genomic, morphological, and fossil evidence. *BMC Evol. Biol.* 7:224
58. Nishihara H, Satta Y, Nikaido M, Thewissen JGM, Stanhope MJ, Okada N. 2005. A retroposon analysis of afrotherian phylogeny. *Mol. Biol. Evol.* 22(9):1823–33
59. Pardini AT, O'Brien PCM, Fu B, Bonde RK, Elder FFB, et al. 2007. Chromosome painting among Proboscidea, Hyracoidea and Sirenia: support for Paenungulata (Afrotheria, Mammalia) but not Tethytheria. *Proc. Biol. Sci.* 274(1615):1333–40

60. Robinson TJ, Fu B, Ferguson-Smith MA, Yang F. 2004. Cross-species chromosome painting in the golden mole and elephant-shrew: support for the mammalian clades Afrotheria and Afroinsectiphillia but not Afroinsectivora. *Proc. Biol. Sci.* 271(1547):1477–84
61. Schull JK, Turakhia Y, Dally WJ, Bejarano G. 2019. Champagne: Whole-genome phylogenomic character matrix method places Myomorpha basal in Rodentia. bioRxiv 803957. **https://doi.org/10.1101/803957**
62. Doronina L, Churakov G, Kuritzin A, Shi J, Baertsch R, et al. 2017. Speciation network in Laurasiatheria: retrophylogenomic signals. *Genome Res.* 27(6):997–1003
63. Springer MS, Molloy EK, Sloan DB, Simmons MP, Gatesy J. 2019. ILS-aware analysis of low-homoplasy retroelement insertions: inference of species trees and introgression using quartets. *J. Hered.* 111(2):147–68
64. Chen M-Y, Liang D, Zhang P. 2017. Phylogenomic resolution of the phylogeny of Laurasiatherian mammals: exploring phylogenetic signals within coding and noncoding sequences. *Genome Biol. Evol.* 9(8):1998–2012
65. Jebb D, Huang Z, Pippel M, Hughes GM, Lavrichenko K, et al. 2020. Six reference-quality genomes reveal evolution of bat adaptations. *Nature* 583(7817):578–84
66. Janečka JE, Miller W, Pringle TH, Wiens F, Zitzmann A, et al. 2007. Molecular and genomic data identify the closest living relative of primates. *Science* 318(5851):792–94
67. Mason VC, Li G, Minx P, Schmitz J, Churakov G, et al. 2016. Genomic analysis reveals hidden biodiversity within colugos, the sister group to primates. *Sci. Adv.* 2(8):e1600633
68. Kumar V, Hallström BM, Janke A. 2013. Coalescent-based genome analyses resolve the early branches of the Euarchontoglires. *PLOS ONE* 8(4):e60019
69. Armstrong J, Hickey G, Diekhans M, Fiddles IT, Novak AM, et al. 2020. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* 587(7833):246–51
70. Gatesy J, Springer MS. 2017. Phylogenomic red flags: homology errors and zombie lineages in the evolutionary diversification of placental mammals. *PNAS* 114(45):E9431–32
71. Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24(6):332–40
72. Joly S, McLenachan PA, Lockhart PJ. 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. *Am. Nat.* 174(2):E54–70
73. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328(5979):710–22
74. Pease JB, Hahn MW. 2015. Detection and polarization of introgression in a five-taxon phylogeny. *Syst. Biol.* 64(4):651–62
75. Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* 32(1):244–57
76. Murphy WJ, Pevzner PA, O'Brien SJ. 2004. Mammalian phylogenomics comes of age. *Trends Genet.* 20(12):631–39
77. Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* 27(3):570–80
78. Ogilvie HA, Bouckaert RR, Drummond AJ. 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Mol. Biol. Evol.* 34(8):2101–14
79. Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30(23):3317–24
80. Vachaspati P, Warnow T. 2018. SVDquest: improving SVDquartets species tree estimation using exact optimization within a constrained search space. *Mol. Phylogenet. Evol.* 124:122–36
81. Liu L, Yu L, Pearl DK, Edwards SV. 2009. Estimating species phylogenies using coalescence times among sequences. *Syst. Biol.* 58(5):468–77
82. He C, Liang D, Zhang P. 2020. Asymmetric distribution of gene trees can arise under purifying selection if differences in population size exist. *Mol. Biol. Evol.* 37(3):881–92
83. Gatesy J, Springer MS. 2013. Concatenation versus coalescence versus "concatalescence." *PNAS* 110(13):E1179

84. Lanier HC, Knowles LL. 2012. Is recombination a problem for species-tree analyses? *Syst. Biol.* 61(4):691–701
85. Springer MS, Gatesy J. 2018. Delimiting coalescence genes (c-genes) in phylogenomic data sets. *Genes* 9(3):123
86. Springer MS, Gatesy J. 2016. The gene tree delusion. *Mol. Phylogenet. Evol.* 94:1–33
87. Bergsten J. 2005. A review of long-branch attraction. *Cladistics* 21(2):163–93
88. Bromham L. 2019. Six impossible things before breakfast: assumptions, models, and belief in molecular dating. *Trends Ecol. Evol.* 34(5):474–86
89. Ali RH, Bogusz M, Whelan S. 2019. Identifying clusters of high confidence homologies in multiple sequence alignments. *Mol. Biol. Evol.* 36(10):2340–51
90. Árnason Ú, Lammers F, Kumar V, Nilsson MA, Janke A. 2018. Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Sci. Adv.* 4(4):eaap9873
91. Svardal H, Jasinska AJ, Apetrei C, Coppola G, Huang Y, et al. 2017. Ancient hybridization and strong adaptation to viruses across African vervet monkey populations. *Nat. Genet.* 49(12):1705–13
92. Fan Z, Zhou A, Osada N, Yu J, Jiang J, et al. 2018. Ancient hybridization and admixture in macaques (genus *Macaca*) inferred from whole genome sequences. *Mol. Phylogenet. Evol.* 127:376–86
93. Kuhlwilm M, Han S, Sousa VC, Excoffier L, Marques-Bonet T. 2019. Ancient admixture from an extinct ape lineage into bonobos. *Nat. Ecol. Evol.* 3(6):957–65
94. Palkopoulou E, Lipson M, Mallick S, Nielsen S, Rohland N, et al. 2018. A comprehensive genomic history of extinct and living elephants. *PNAS* 115(11):E2566–74
95. Cahill JA, Stirling I, Kistler L, Salamzade R, Ersmark E, et al. 2015. Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Mol. Ecol.* 24(6):1205–17
96. Li G, Davis BW, Eizirik E, Murphy WJ. 2016. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Res.* 26(1):1–11
97. Figueiró HV, Li G, Trindade FJ, Assis J, Pais F, et al. 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Sci. Adv.* 3(7):e1700299
98. vonHoldt BM, Aardema ML. 2020. Updating the bibliography of interbreeding among *Canis* in North America. *J. Hered.* 111(3):249–62
99. Wen D, Yu Y, Zhu J, Nakhleh L. 2018. Inferring phylogenetic networks using PhyloNet. *Syst. Biol.* 67(4):735–40
100. Solís-Lemus C, Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLOS Genet.* 12(3):e1005896
101. Bannert N, Kurth R. 2004. Retroelements and the human genome: new perspectives on an old relation. *PNAS* 101(Suppl. 2):14572–79
102. Doronina L, Reising O, Clawson H, Ray DA, Schmitz J. 2019. True homoplasy of retrotransposon insertions in primates. *Syst. Biol.* 68(3):482–93
103. Haenel Q, Laurentino TG, Roesti M, Berner D. 2018. Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Mol. Ecol.* 27(11):2477–97
104. Schumer M, Xu C, Powell D, Durvasula A, Skov L, et al. 2018. Natural selection interacts with the local recombination rate to shape the evolution of hybrid genomes. *Science* 360(6389):656–60
105. Chan AH, Jenkins PA, Song YS. 2012. Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLOS Genet.* 8(12):e1003090
106. McVean G, Awadalla P, Fearnhead P. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160(3):1231–41
107. Auton A, Fledel-Alon A, Pfeifer S, Venn O, Ségurel L, et al. 2012. A fine-scale chimpanzee genetic map from population sequencing. *Science* 336(6078):193–98
108. Adrion JR, Galloway JG, Kern AD. 2020. Predicting the landscape of recombination using deep learning. *Mol. Biol. Evol.* 37(6):1790–808
109. Faurby S, Svenning J-C. 2015. A species-level phylogeny of all extant and late Quaternary extinct mammals using a novel heuristic-hierarchical Bayesian approach. *Mol. Phylogenet. Evol.* 84:14–26
110. Pennisi E. 2016. Shaking up the Tree of Life. *Science* 354(6314):817–21
111. Roca AL, Georgiadis N, O'Brien SJ. 2005. Cytonuclear genomic dissociation in African elephant species. *Nat. Genet.* 37(1):96–100

112. Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* 21(16):3907–30
113. Schaefer NK, Shapiro B, Green RE. 2016. Detecting hybridization using ancient DNA. *Mol. Ecol.* 25(11):2398–412
114. Jones MR, Mills LS, Alves PC, Callahan CM, Alves JM, et al. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* 360(6395):1355–58
115. Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, et al. 2015. Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347(6217):1258524
116. *Heliconius* Genome Consort. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487(7405):94–98
117. Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, et al. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* 507(7492):354–57
118. Irisarri I, Singh P, Koblmüller S, Torres-Dowdall J, Henning F, et al. 2018. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nat. Commun.* 9:3159
119. Martin SH, Jiggins CD. 2017. Interpreting the genomic landscape of introgression. *Curr. Opin. Genet. Dev.* 47:69–74
120. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, et al. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.* 23(11):1817–28
121. Storchová R, Reif J, Nachman MW. 2010. Female heterogamety and speciation: reduced introgression of the Z chromosome between two species of nightingales. *Evolution* 64(2):456–71
122. Presgraves DC. 2018. Evaluating genomic signatures of "the large X-effect" during complex speciation. *Mol. Ecol.* 27(19):3822–30
123. Thompson MJ, Jiggins CD. 2014. Supergenes and their role in evolution. *Heredity* 113:1–8
124. Carneiro M, Blanco-Aguiar JA, Villafuerte R, Ferrand N, Nachman MW. 2010. Speciation in the European rabbit (*Oryctolagus cuniculus*): islands of differentiation on the X chromosome and autosomes. *Evolution* 64(12):3443–60
125. Carneiro M, Albert FW, Afonso S, Pereira RJ, Burbano H, et al. 2014. The genomic architecture of population divergence between subspecies of the European rabbit. *PLOS Genet.* 10(8):e1003519
126. Nam K, Munch K, Hobolth A, Dutheil JY, Veeramah KR, et al. 2015. Extreme selective sweeps independently targeted the X chromosomes of the great apes. *PNAS* 112(20):6413–18
127. Ai H, Fang X, Yang B, Huang Z, Chen H, et al. 2015. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* 47(3):217–25
128. Ludwig MZ. 2016. Noncoding DNA evolution: junk DNA revisited. In *The Encyclopedia of Evolutionary Biology*, Vol. 1, ed. R Kliman, pp. 124–29. Amsterdam: Elsevier
129. Eichler EE. 2001. Segmental duplications: What's missing, misassigned, and misassembled—and should we care? *Genome Res.* 11(5):653–56
130. Lee H, Schatz MC. 2012. Genomic dark matter: the reliability of short read mapping illustrated by the genome mappability score. *Bioinformatics* 28(16):2097–105
131. Sedlazeck FJ, Lee H, Darby CA, Schatz MC. 2018. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat. Rev. Genet.* 19(6):329–46
132. Dumbovic G, Forcales S-V, Perucho M. 2017. Emerging roles of macrosatellite repeats in genome organization and disease development. *Epigenetics* 12(7):515–26
133. Gouy A, Excoffier L. 2020. Polygenic patterns of adaptive introgression in modern humans are mainly shaped by response to pathogens. *Mol. Biol. Evol.* 37(5):1420–33
134. Koren S, Rhie A, Walenz BP, Dilthey AT, Bickhart DM, et al. 2018. *De novo* assembly of haplotype-resolved genomes with trio binning. *Nat. Biotechnol.* 36:1174–82
135. Rice ES, Koren S, Rhie A, Heaton MP, Kalbfleisch TS, et al. 2020. Continuous chromosome-scale haplotypes assembled from a single interspecies F1 hybrid of yak and cattle. *GigaScience* 9(4):giaa029
136. Bredemeyer KR, Harris AJ, Li G, Foley NM, Roelke-Parker M, et al. 2021. Ultracontinuous single haplotype genome assemblies for the domestic cat (*Felis catus*) and Asian leopard cat (*Prionailurus bengalensis*). *J. Hered.* In press

137. Vollger MR, Logsdon GA, Audano PA, Sulovari A, Porubsky D, et al. 2020. Improved assembly and variant detection of a haploid human genome using single-molecule, high-fidelity long reads. *Ann. Hum. Genet.* 84(2):125–40
138. Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, et al. 2020. Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 585(7823):79–84
139. Logsdon GA, Vollger MR, Hsieh P, Mao Y, Liskovykh MA, et al. 2020. The structure, function, and evolution of a complete human chromosome 8. bioRxiv. **https://doi.org/10.1101/2020.09.08.285395**
140. Low WY, Tearle R, Koren S, Rhie S, Bickhart DM, et al. 2020. Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle. *Nat. Commun.* 11:2071
141. Henikoff S, Ahmad K, Malik HS. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293(5532):1098–102
142. Bayes JJ, Malik HS. 2009. Altered heterochromatin binding by a hybrid sterility protein in *Drosophila* sibling species. *Science* 326(5959):1538–41
143. Ferree PM, Barbash DA. 2009. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLOS Biol.* 7(10):e1000234
144. Marques-Bonet T, Ryder OA, Eichler EE. 2009. Sequencing primate genomes: What have we learned? *Annu. Rev. Genom. Hum. Genet.* 10:355–86
145. Gordon D, Huddleston J, Chaisson MJP, Hill CM, Kronenberg ZN, et al. 2016. Long-read sequence assembly of the gorilla genome. *Science* 352(6281):aae0344
146. Kronenberg ZN, Fiddes IT, Gordon D, Murali S, Cantsilieris S, et al. 2018. High-resolution comparative analysis of great ape genomes. *Science* 360(6393):eaar6343
147. Dougherty ML, Nuttle X, Penn O, Nelson BJ, Huddleston J, et al. 2017. The birth of a human-specific neural gene by incomplete duplication and gene fusion. *Genome Biol.* 18:49
148. Nuttle X, Giannuzzi G, Duyzend MH, Schraiber JG, Narvaiza I, et al. 2016. Emergence of a *Homo sapiens*-specific gene family and chromosome 16p11.2 CNV susceptibility. *Nature* 536(7615):205–9
149. Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, et al. 2001. Positive selection of a gene family during the emergence of humans and African apes. *Nature* 413(6855):514–19
150. Hsieh P, Vollger MR, Dang V, Porubsky D, Baker C, et al. 2019. Adaptive archaic introgression of copy number variants and the discovery of previously unknown human genes. *Science* 366(6463):eaax2083
151. Meiklejohn CD, Landeen EL, Gordon KE, Rzatkiewicz T, Kingan SB, et al. 2018. Gene flow mediates the role of sex chromosome meiotic drive during complex speciation. *eLife* 7:e35468
152. Presgraves DC. 2008. Sex chromosomes and speciation in *Drosophila*. *Trends Genet.* 24(7):336–43
153. Presgraves DC. 2010. The molecular evolutionary basis of species formation. *Nat. Rev. Genet.* 11(3):175–80
154. Warburton PE, Hasson D, Guillem F, Lescale C, Jin X, Abrusan G. 2008. Analysis of the largest tandemly repeated DNA families in the human genome. *BMC Genom.* 9:533
155. Bellott DW, Skaletsky H, Pyntikova T, Mardis ER, Graves T, et al. 2010. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* 466(7306):612–16
156. Mueller JL, Skaletsky H, Brown LG, Zaghlul S, Rock S, et al. 2013. Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat. Genet.* 45(9):1083–87
157. Davis BW, Seabury CM, Brashear WA, Li G, Roelke-Parker M, Murphy WJ. 2015. Mechanisms underlying mammalian hybrid sterility in two feline interspecies models. *Mol. Biol. Evol.* 32(10):2534–46
158. Soh YQS, Alföldi J, Pyntikova T, Brown LG, Graves T, et al. 2014. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* 159(4):800–13
159. Church DM, Goodstadt L, Hillier LW, Zody MC, Goldstein S, et al. 2009. Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLOS Biol.* 7(5):e1000112
160. Bellott DW, Cho T-J, Hughes JF, Skaletsky H, Page DC. 2018. Cost-effective high-throughput single-haplotype iterative mapping and sequencing for complex genomic structures. *Nat. Protoc.* 13(4):787–809
161. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423(6942):825–37
162. Meiklejohn CD, Tao Y. 2010. Genetic conflict and sex chromosome evolution. *Trends Ecol. Evol.* 25(4):215–23

163. Good JM. 2012. The conflict within and the escalating war between the sex chromosomes. *PLOS Genet*. 8(9):e1002955
164. Cocquet J, Ellis PJI, Mahadevaiah SK, Affara NA, Vaiman D, Burgoyne PS. 2012. A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. *PLOS Genet*. 8(9):e1002900
165. Kruger AN, Brogley MA, Huizinga JL, Kidd JM, de Rooij DG, et al. 2019. A neofunctionalized X-linked ampliconic gene family is essential for male fertility and equal sex ratio in mice. *Curr. Biol.* 29(21):3699–706.e5
166. Janečka JE, Davis BW, Ghosh S, Paria N, Das PJ, et al. 2018. Horse Y chromosome assembly displays unique evolutionary features and putative stallion fertility genes. *Nat. Commun.* 9:2945
167. Brashear WA, Raudsepp T, Murphy WJ. 2018. Evolutionary conservation of Y chromosome ampliconic gene families despite extensive structural variation. *Genome Res*. 28(12):1841–51
168. Coyne JA, Orr HA. 1989. Two rules of speciation. In *Speciation and Its Consequences*, ed. D Otte, JA Endler, pp. 180–207. Sunderland, MA: Sinauer Assoc.
169. Li G, Hillier LW, Grahn RA, Zimin AV, David VA, et al. 2016. A high-resolution SNP array-based linkage map anchors a new domestic cat draft genome assembly and provides detailed patterns of recombination. *G3* 6(6):1607–16
170. Deng X, Ma W, Ramani V, Hill A, Yang F, et al. 2015. Bipartate structure of the inactive mouse X chromosome. *Genome Biol*. 16:152
171. Du Y, Wu S, Edwards SV, Liu L. 2019. The effect of alignment uncertainty, substitution models and priors in building and dating the mammal tree of life. *BMC Evol. Biol.* 19:203
172. Nery MF, González DJ, Hoffmann FG, Opazo JC. 2012. Resolution of laurasiatherian phylogeny: evidence from genomic data. *Mol. Phylogenet. Evol.* 64(3):685–89
173. Shaw TI, Srivastava A, Chou W-C, Liu L, Hawkinson A, et al. 2012. Transcriptome sequencing and annotation for the Jamaican fruit bat (*Artibeus jamaicensis*). *PLOS ONE* 7(11):e48472
174. Song S, Liu L, Edwards SV, Wu S. 2012. Resolving conflicts in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc. Natl. Acad. Sci. USA* 109(37):14942–47
175. Tsagkogeorga G, Parker J, Stupka E, Cotton JA, Rossiter SJ. 2013. Phylogenomic analyses elucidate the evolutionary relationships of bats. *Curr. Biol.* 23(22):2262–67

Annual Review of
Animal Biosciences

Volume 9, 2021

Contents

Errata

An online log of corrections to *Annual Review of Animal Biosciences* articles may be found at http://www.annualreviews.org/errata/animal