Collection: Methods for ecological and evolutionary data analysis

Paper type: Article

Landmark-based estimates of genomic disparity

Ashwini V. Mohan[1,2], Anjali Goswami[1], and Jeffrey W. Streicher[1*]

[1]Natural History Museum, London SW7 5BD, United Kingdom

[2]Institut de Biologie, Université de Neuchâtel, Neuchâtel, Switzerland

* Author for correspondence Jeffrey W. Streicher (ORCID ID: 0000-0002-3738-4162)

Email: j.streicher@nhm.ac.uk

Author Contributions: J.W.S., A.G., and A.V.M. designed research; performed research; conducted theoretical analysis; conducted empirical analysis; agree to be held responsible for all errors associated with the analyses (excluding genome assembly which was not carried out by the authors); and wrote the paper.

Inclusion and Ethics Statement: All three authors fulfill the Nature Portfolio criteria for authorship. This research is the result of the lead author's (A.V.M) postdoctoral fellowship that was co-hosted by the other two authors (J.W.S. and A.G) at the Natural History Museum, London. All authors edited and approved the manuscript before submission.

Competing Interest Statement: The authors declare no competing interest.

Data and Code Availability Statement: Genomes and individual chromosomes analyzed in this study were downloaded from the NCBI database. Study references and accession IDs for genome assemblies and individual chromosomes are listed in Table 1 and/or the SI Appendix. The code (R and shell scripts) used in the disparity pipeline are available via a GitHub repository (https://github.com/nhm-herpetology/genomic-disparity/blob/main/README.md) along with two detailed step-by-step tutorials.

Keywords: Multivariate analyses, genomic architecture, comparative genomics, genomic landmarks, syntenic disparity

This file includes:

Main Text
Figures 1 to 8
Tables 1 to 2

Collection: Methods for ecological and evolutionary data analysis

Paper type: Article

Abstract

Genomic architecture has played a key role in the evolution of biodiversity. Structural comparisons of genome sequences have informed the study of supergenes, sex chromosomes, and some of the earliest divergences in the tree of life. However, multi-species comparisons of whole genome sequences still have many computational and analytical limitations. Here we present a computationally non-intensive approach, based on geometric morphometrics, that generates disparity scores using conserved sequences as landmarks. This 'geno-metric' method captures structural differences in homologous chromosomal regions and can be applied at micro- and macroevolutionary scales. Using ultraconserved elements (UCEs) as landmarks, we provide empirical demonstrations with chromosome 5 from the *Drosophila virilis* group and 40 placental mammal genomes. These examples identify chromosomal rearrangements while also confirming that UCE landmark placement has strong parallels with morphometrics regarding chromosome size, landmark orientation, phylogenetic signal of landmarks, and landmark availability. Using simulations, we show that UCE-inferred disparity is correlated with overall levels of chromosome evolution which can be further studied using macroevolutionary models of continuous trait evolution. Landmark-based estimates of disparity are a valuable addition to the comparative genomic toolkit in that they offer an intuitive, rapid mechanism for detecting chromosomal expansions, contractions, and rearrangements. We also highlight the potential for this approach to identify translocations and genome duplications.

Significance Statement

The increased throughput of genome sequencing has outpaced the development of analytical tools to extract information from these big data, creating an urgent need for novel methodological perspectives in genomics. We describe and validate a morphometric approach based on conserved genomic landmarks that effectively captures multi-species variation of chromosome structure. Using the principles of morphological disparity, the throughput of comparative genomics can be greatly increased, and the syntenic organization of chromosomes placed in quantitative and macroevolutionary contexts.

Main Text

Introduction

We are now firmly in the age of genomics with fully sequenced genomes commonplace for most organismal groups. Because of this, multi-species genomic comparisons are increasingly used in evolutionary and ecological research. This is even the case for organisms with relatively large nuclear genomes (1–3). Excitingly, the comparison of whole genome sequences has offered valuable evolutionary insights into a range of topics including gene orthology (4), transposable element evolution (5), and adaptation to novel environments (6).

One important area of comparative genomics research is the study of genome architecture (7). The structural architecture of genomes and their constituent chromosomes is highly relevant to the evolution of biodiversity and disease (8). From supergenes (9) to sex chromosomes (10), evolutionary shifts in chromosome structure are known to have a substantial influence on organismal phenotype. Recent research utilizing forward in time simulations of chromosomal rearrangements has indicated their important role in sustaining adaptation (11). While chromosome-based methods like Hi-C scaffolding have substantially improved individual genome assemblies (12, 13), comparative methods for studying chromosome structure based on sequence data have notable limitations when applied to multiple species.

Most, if not all, existing comparative methods 'break up' chromosome assemblies into whole genome alignments (WGA;(14)). Multi-species comparisons of chromosomal structure then require post hoc comparisons of the WGA achieved using mapping files and (often) a reference genome to refer subsets of the WGA back to the original chromosome assemblies. For example, this approach is used to conduct analyses like protein prediction (15), and synteny assessment (16). To compare chromosome structure across multiple species, many pairwise comparisons are performed, and visual interpretation is achieved through complex network diagrams (e.g. (17–19)).

The limitations of current multi-species approaches are three-fold. First, when applied to large, divergent genomes WGA tends to be computationally intensive, making species-rich comparisons non-tractable for most researchers. Second, available methods for assessing structural differences among chromosome sequences (i.e. synteny) require users to make *a priori* decisions about the size of syntenic blocks, sliding windows, and gaps making results sensitive to user settings (20). Third, visualization of results typically occurs in a pairwise format (i.e. chromosome structure of species A to species B), making it difficult to interpret overall patterns across multi-species datasets simultaneously. Even though multi-species comparisons have been made possible more recently (e.g. (21)), the methods are limited to visualizing synteny across species without a quantitative measure of disparity between species. Given these limitations, a reference-free, computationally non-intensive method that allows researchers to assess structural variation across multiple genomes simultaneously is desirable.

Within the study of comparative or evolutionary morphology, structural variation across species is commonly termed disparity (22, 23). Measures of morphological disparity are routinely used to explore the evolutionary dynamics of complex phenotypes (e.g. (24, 25)). Here we describe and validate a disparity-based method that uses conserved landmark sequences to capture genomic variation across different species in a single analytical framework. Our 'geno-metric' method is based on a simple premise: genomic structure is morphology (i.e., a shape, form, or structure) and there are well-validated methods for capturing morphological disparity across multiple species (22, 23, 26). Furthermore, genomes are filled with homologous sequences (=landmarks) making them amenable for such analyses (e.g. (27)). Because they do not require whole genome alignment, landmark-based estimates of genomic disparity hold promise for use as a convenient, computationally non-intensive tool for exploring the organization and structure of chromosomes from high-quality genome assemblies across many species simultaneously.

A multi-species morphometric approach for comparing chromosomes

Identification of genomic landmarks. To develop a pipeline for genomic disparity analysis, we used ultraconserved elements (UCEs) as landmarks (28). We selected UCEs because they are highly conserved, single copy and lack repeats (29, 30), making them ideal landmarks. They have also been implicated as relevant to the structure of key developmental and transcriptomic pathways (31, 32), making variation in their relative chromosomal placements of biological interest.

Our data preparation pipeline includes multiple steps (**Fig. 1**). We begin by acquiring chromosome assemblies from NCBI and then reference mapping UCE landmarks to the assemblies using bwa (33). We tested other approaches for reference mapping including minimap2 (34) but found bwa resulted in the most mapped landmarks per chromosome (SI Appendix, **Fig. S1**). After orthologous landmarks are mapped, we either prepare landmarked chromosomes for downstream analysis (when chromosomal homology is known), or cluster the chromosomes based on the presence/absence of UCE landmarks via metric multi-dimensional scaling (when chromosomal homology is not known). The latter option allows for the identification

of chromosomes that share large numbers of landmarks, and thus likely large homologous genomic blocks. When preparing homologous, landmarked chromosomes for downstream analysis there are several steps including, removal of landmarks that are multi-copy or not shared across homologs, complementarity checks (5'-3' versus 3'-5'), and exportation of landmark position information into a spreadsheet. The spreadsheet has rows representing species or individuals (one for each species or individual included in the input assemblies) and columns are landmark positions on a homologous chromosome set. This spreadsheet is then used as input for the disparity inference stage. Unless otherwise indicated, all steps in the pipeline occur using base functions in the R statistical package (35). The pipeline is modular and can be modified for a variety of other uses. Two step-by-step tutorials along with custom R and shell scripts are available on GitHub (https://github.com/nhm-herpetology/genomic-disparity/blob/main/README.md). Additional methods details can be found in the SI Appendix.

Interpretation and inference of genomic disparity. As with other disparity analyses, we use ordination techniques to reduce the dimensionality of variation in landmark position and the spreadsheet of landmark positions is analyzed using principal component analysis (**Fig. 1**). As an illustrative example, we ran our pipeline using chromosome 5 from six individuals of the *Drosophila virilis* complex: *D. americana*, *D. flavomontana*, *D. montana*, *D. novamexicana*, and two individuals of *D. virilis*. We selected this chromosome and these species because multiple chromosomal inversions have occurred during their speciation (36). We used a set of dipteran UCEs (37) as landmarks. A step-by-step online tutorial outlining this example is available on GitHub (see SI Appendix). Based on 191 landmarks, we found that three PC axes explained nearly 100% of the variance in landmark placement (PC1, 94%; PC2 5.2%; PC3, 0.07%). When visualizing the distribution of datapoints across these axes, datapoints close to one another in geno-metric space have highly conserved, syntenic landmark placements. Notably, this is the case for the two individuals of *D. virilis* (**Fig. 2A**) and close relatives *D. americana* and *D. novamexicana* (**Fig. 2B**; 38). In contrast, separation in multivariate space corresponds to distinct inversion events. PC1 separates *D. flavomontana* from the other taxa and this appears to be explained by two blocks of UCE landmarks with substantially different placements on the chromosome (**Fig. 2C**). PC2 separates *D. montana* from the other taxa which is likely related to a pericentric chromosomal inversion (**Fig. 2D**). Importantly, when we compare the landmark orientations on *D. montana* and *D. flavomontana*, we see evidence (**Fig. 2E**) of the two blocks of UCE landmarks (PC1) and the pericentric inversion (PC2). PC3 maximizes separation between *D. virilis* and *D. americana* + *D. novamexicana* which is likely explained by a paracentric inversion that *D. virilis* has relative to both *D. americana* (**Fig. 2F**) and *D. novamexicana* (**Fig. 2G**). This example illustrates that UCE landmarks are capable of capturing variation associated with landmark re-arrangements caused by chromosomal inversions among closely related species. However, this analytical approach can also be used to study structural genome evolution across deep evolutionary time.

***De novo* homology discovery using whole genome assemblies.** Over long periods of evolutionary time, genome structure may change substantially eroding the homology of chromosomes. However, it is also apparent that large genomic regions often remain evolutionarily intact (e.g., 17) among which disparity measures can be calculated. The chromosome clustering module of our pipeline can be used to identify these large genomic regions. To demonstrate this, we obtained 40 genome assemblies of mammals (**Table 1**). We used a previously described dataset of tetrapod UCEs (30) as landmarks. Depending on the research question, other single-copy landmarks (e.g. BUSCO genes [39]) could be used as landmarks, however, for basic exploration we recommend UCEs because we found that the number of UCEs mapped onto each mammal chromosome was positively and significantly correlated with chromosome size ($r = 0.75$, $p < 0.05$; SI Appendix, **Fig. S2**). This indicates that UCEs are ideal landmarks for generalized disparity analysis as they are widely distributed across chromosomes.

During the MMDS clustering step we used score thresholding to identify two sets of chromosomes that (1) had highly similar in their UCE compositions and (2) contained only a single chromosome per species in the dataset (**Fig. 3**). Using protein annotations from the Ensembl database (release 110; 40), we generated Jaccard similarity coefficients to compare protein compositions on the two focal chromosome sets across a subset of species (N = 14). We found that protein content was significantly more similar within chromosome sets than between them (W = 6742, $p < 0.0001$; nonparametric Wilcoxon tests; SI Appendix, **Fig. S3**) indicating that focal chromosomes tended to also share functional (protein-coding) homologs in addition to UCE landmarks. Collectively, this suggests that the identified chromosomes share large homologous blocks of genomic sequence. This clustering approach will identify many sets of chromosomes in closely related taxa (including among macro- and micro-chromosomes) due to the conserved nature of their genomic sequences. In more distantly related taxa, only the first few axes may identify chromosome sets because extensive evolutionary changes (such as translocations or fissions) have likely eroded the extent to which whole chromosomes are homologous.

Landmarking of UCES on the two mammal chromosome sets resulted in 184 (set 1, 37 species) and 162 (set 2, 36 species) landmarks, respectively (**Fig. 4**, **A, B**). There was not complete taxonomic overlap in the two chromosome sets because a handful of species did not have the focal UCE landmarks on a single chromosome, suggesting that the homologous blocks of landmarks may not exist in these species due to chromosome evolution (**Table 1**). We found that most of the variation in both chromosome sets was explained by the first principal component (PC). In chromosome set 1, PC1 described 89.1% of the variation and in chromosome set 2, PC1 described 97.5% of the variation. We discovered that in both chromosome sets, PC1 was significantly correlated with chromosome size (set 1, rho = 0.70, $p < 0.000$, **Fig. 4C**; set 2, rho = 0.97, $p < 0.000$, **Fig 4D**). This was also the case for PC2, although with weaker correlative fits than PC1 (set 1, rho = 0.65, $p < 0.000$; set 2, rho = 0.41, $p = 0.0128$). While describing only a small percentage of the overall variation, PC2 and PC3 identified several disparate taxa. For example, in set 1 analysis of PC2 (8.8% variance) and PC3 (1.2% variance) identified the chromosomes of *Sus scrofa*, *Pan paniscus*, *Microcebus murinus,* and *Rhinopithecus roxellana* to be disparate outliers (**Fig. 5A**). Similarly, in chromosome set 2, PC2 (2.4% variance) and PC3 (0.3% variance) identified five taxa as exceptionally disparate, with the lowest PC2 scores belonging to four species of muroid rodents, *Mus pahari*, *Rattus norvegicus*, *Peromyscus maniculatus,* and *Cricetulus griseus* (**Fig. 5B**). Correcting for chromosome size reduced the amount of variance explained by PC1 and increased the amount of variance explained by PC2 in both chromosome sets (see SI Appendix). Despite this, clustering patterns were highly similar between the uncorrected and size-corrected analyses including for the disparate taxa described above (**Fig. S4**). Synteny analysis of landmarks using chromoMap (41) revealed that chromosomes in these disparate taxa possess varying sizes of inversions relative to other taxa included in the analysis, whereas those species with more conserved UCE placements were clustered in the geno-metric space (**Fig. 5C-F**). Interestingly, as observed in the set 2 analyses, chromosomal inversions have been previously reported as important for shaping muroid rodent genomic landscapes (42).

Based on the *Drosophila* and mammal examples, as in morphological analysis, different axes reflect different aspects of variation in landmark position, often associated with chromosomal inversions. Given these results, UCE landmarks are clearly capable of capturing size, orientation, and (potentially) phylogenetic information (e.g. *D. neomexicana* + *D. americana*, Fig. 2; muriod rodent cluster, **Fig. 5B**) across large swaths of genomic architecture. However, the relative contribution of different mechanisms of genome evolution to disparity accumulation is still unclear. To provide a baseline for clearer interpretations of genomic disparity we conducted simulation studies using two ubiquitous causes of chromosome evolution.

Simulated and empirical data indicate genomic disparity increases through time. We simulated two common evolutionary shifts in genomic architecture that occur as species and populations diversify: insertion-deletion (indel) events and chromosomal inversions. We performed these simulations using chromosome set 1 from five species subsampled from the larger dataset, *Capra aegagrus*, *Homo sapiens*, *Mus pahari*, *Neomonachus schauinslandi*, and *Rhinopithecus roxellana* (**Table 1**). We selected these taxa based on phylogenetic diversity (representing Artiodactyla, Carnivora, Primates, and Rodentia) and higher disparity shown in PCAs (*R. roxellana*). We simulated evolutionary events using SimuG (43). We designed the experiment logarithmically, simulating 1, 10, 100, 1,000, and 10,000 random indels or inversions on the ancestral chromosome of the five focal species. We ran five replicated simulations for each of the five focal species. To assess the change in disparity between ancestral chromosomes and their simulated descendants, we used a simple pairwise metric of increase in disparity (ID):

$$ID = | PC_A - PC_B |$$

Where PC_A is the principal component score of the ancestral chromosome, PC_B is the principal component score of the descendent chromosome, and ID is calculated by taking the absolute difference of these values. As such, ID is an intuitive measurement of how much landmark placement changed during the simulations, where a value of 0 indicates no change and larger values correspond to increasingly disparate landmark placements. We predicted that if genomic disparity reflects the degree of evolutionary change of architecture between two genomes, there should be a positive correlation between ID and the number of simulated events. While analysis of individual PCs may be problematic when data have high effective dimensionality (44), here the dominance of PC1 and the overwhelming correlation with chromosome size on that primary axis, justifies separate analysis of these components. Nonetheless, this metric can be easily implemented for a multivariate analysis.

In both simulation studies, > 99% of variation in landmark placement was explained by the first two principal components (indels, PC1 – 80.7%, PC2 – 19.2%; inversions, PC1 – 76.3%, PC2 – 23.5%). As predicted, ID increased with the amount of simulated evolution confirming that, at least under neutral (random) conditions, disparity measures should correspond to relative amounts of genomic evolution (**Fig. 6A-D**). We compared the variance of log-transformed ID scores for indel versus inversion simulations using two-way ANOVAs. We found that on both axes, inversions caused a significantly higher increase in disparity than indels (PC1, $F = 13.52$, $p = 0.003$; PC2, $F = 33.51$, $p < 0.000$; SI Appendix, **Fig. S5**). This is an understandable result as inversions tend to change the orientation of landmarks substantially, whereas indel events only result in the amount of space between landmarks changing.

Given the simulation results, which imply disparity should increase as genomes diverge, we re-examined our empirical dataset for evidence of increased disparity over time. We did this by comparing intra- to inter-order disparity for PC1 and PC2, which together described > 99% of variance in both empirical chromosome sets. If disparity in UCE placement increases with evolutionary time, we expected that inter-order disparity should be significantly higher that intra-order disparity for both PCs because there has been more evolutionary divergence between mammalian orders than within them. Using nonparametric Wilcoxon tests, we found that inter-order disparity was significantly higher for both PC1 (chromosome set 1, $W = 102191$, $p < 0.0001$; chromosome set 2, $W = 102158$, $p < 0.0001$) and PC2 (chromosome set 1, $W = 102959$, $p < 0.0001$; chromosome set 2, $W = 101176$, $p < 0.0001$) than intra-order disparity (**Fig. 6E-H**). Combined with the simulation results, this suggests that disparity in chromosome architecture increases with increasing evolutionary divergence.

Interpreting genomic disparity in a macroevolutionary context. Genomic disparity is a continuous trait and thus can be used for phylogenetic comparative analyses. This attribute may be particularly helpful for studying clades/species with variable genome structures (e.g., varying degrees of sex chromosome differentiation [45]). We performed simulations of trait evolution using a posterior distribution of a birth-death model node-dated mammal phylogeny (100 trees) from previously published research (46) and trimmed the tree to match taxa for each chromosome set (**Table 1**). When a species was not represented on the phylogeny, a close relative within the same genus was used (additional details in SI Appendix).

We simulated trait evolution using disparity through time analyses performed in geiger 2.0.1 (47) and phytools 2.0 (48). Here, the mean disparity of a trait is measured at different slices in relative time where variation within each subtree is compared to the total variation across the tree. We used two standard evolutionary models (Brownian motion and Ornstein−Uhlenbeck; OU) that are often used to assess rates of trait evolution (e.g. (49)). For each model, we simulated disparity through time for a continuous variable on each of the 100 trees (in each chromosome set) and then calculated the mean disparity. We compared these 'forward through time' simulations with disparity through time as estimated when using the UCE-landmark PC2 scores as tip data on the phylogeny (i.e. backwards through time). We plotted all simulations so that we could compare estimates of genomic disparity to modelled trait evolution. We found that while the estimate for PC2 from chromosome set 1 was not a good fit for Brownian motion, it was a closer fit to Ornstein−Uhlenbeck (**Fig. 7**, left side). In contrast, the estimate from PC2 for chromosome set 2 was a better fit for Brownian motion simulations (**Fig. 7**, right side).

These results suggest an intriguing, but understandable pattern: the evolutionary accumulation of disparity in UCE placement can vary across chromosomes. Here, we observed disparity accumulation consistent with both Brownian motion (chromosome set 2) and adaptive optimum (OU) evolution (chromosome set 1). We also used stochastic character mapping (simMap; (50) to explore patterns of chromosome structural evolution using PC2 scores and the mammal phylogenies (46). We found that the chromosome set with Brownian motion-like structural evolution (set 2; **Fig. 8**, right side) had more phylogenetically restricted patterns than the OU-like chromosome set 1 (**Fig. 8**, left side). Specifically, the biggest shift in PC scores in chromosome set 2 was observed in the clade of muroid rodents containing the species that were found to be exceptionally disparate in the principal component analysis (**Fig. 5B**), whereas the stochastic character map of chromosome set 1 implied more dynamic structural evolution of chromosomes across placental mammalian orders.

Discussion

Using a morphometrics-inspired approach we identified structural variation associated with chromosome size, chromosome inversions, insertion-deletion events (=relative size change), and phylogeny. These qualities mirror the utility of morphological disparity analysis as applied to complex phenotypic structures. In a phenotypic context, disparity can be descriptive or serve as a proxy for phylogenetic affinities and ecology (23). It is widely used to understand adaptive radiation in the paleontological record (51). In a genomic context, we are optimistic that disparity can play a similar role in understanding how genome architecture varies among species and be used for quantitative comparisons with organismal evolution and adaptation. Below we discuss the value of landmark-based estimates of genomic disparity moving forward.

Genomic disparity is a powerful and accurate discovery tool. Genomic disparity enables a broad understanding of how similar (or dissimilar) chromosome structures are across different species based on the distribution of conserved landmark sequences like UCEs. This analytical quality is particularly advantageous for studying large-scale shifts in chromosome organization

across organisms relevant for the study of supergenes and homologous synteny blocks (*sensu* (17)). The approach does have several assumptions, most notably that the input genome assemblies are high-quality and chromosome-level. It also assumes that species are closely related enough that homologous synteny blocks have not been eroded by evolutionary divergence. For example, in the mammal examples, we needed to remove multiple UCE landmarks that were not present in all species and while we found that protein-coding content tended to be more similar within mammalian orders than between them, this was not a significant result (W = 499.5, p = 0.0722; nonparametric Wilcoxon tests; SI Appendix**, Fig. S6**). Presumably, both observations are explained by the loss of chromosome homology over evolutionary time; however, this phenomenon has strong parallels with landmarking of morphological structures where primary homologous landmarks (*sensu* (52)) are also lost over evolutionary time. Applying this method to polyploid organisms may be informative when chromosome phasing has been performed and the pipeline is applied separately to phased chromosome sets.

In addition to documenting variation in landmark placement across species (or populations), disparity analysis may be useful for identifying inter-chromosomal duplications, translocations, and potentially erroneous genome assemblies.

Putatively duplicated landmarks on different chromosomes are identified by our pipeline and removed so that landmarks are single copy orthologs. However, patterns of duplication may be of interest for other reasons and can be explored using the pipeline. Our ‘homology unknown’ pipeline branch can provide insights into landmark dynamics across different chromosomes within species. This is because our landmarking pipeline is applied iteratively to all chromosomes in a genome assembly. The presence/absence matrix used during the MMDS stage of the analysis (**Fig. 1**) can be applied to intraspecific genome assemblies to identify chromosomes with duplicated genome sections (assuming they contain the same landmark sequences), a commonly observed genomic phenomenon (e.g. (53)). When applied interspecifically, the MMDS analysis also has potential to identify subsets of landmarks that have been translocated (using clustering patterns and weighting scores across MMDS axes to identify chromosomes with multiple multivariate affinities; see SI Appendix). As such, rates of intra-chromosomal versus inter-chromosomal rearrangements among different groups could be measured quantitatively. For example, afrotherian mammals and marsupials are shown to have different genome reshuffling patterns at chromosome scales (54). Using our pipeline, this could be tested with whole genome assemblies and dense species sampling.

While chromosome-level genome assemblies are increasing in availability for groups like mammals, the quality of available chromosome sequences varies with different metrics used to assess completeness and accuracy (55). As such, it is plausible that some of the mammal genome assemblies we used contain assembly errors. At least two of the disparate outliers from the principal component analyses were suspicious in this regard (**Fig. 5**). First, the snub-nosed monkey (*Rhinopithecus roxellana*) was highly disparate in both chromosome sets which was unexpected given that it has a similar karyotype to other primates (2N = 44; (56)). Second, another disparate taxon in the chromosome set 1 analysis, the pig (*Sus scrofa*; **Fig. 5D**), has a history of assembly issues. While we used the latest assembly version ‘Sscrofa11.1’ produced by Warr et al. (57), the improved assembly still contains numerous gaps, which may explain the observed disparity from the principal component analysis. Both these cases require additional scrutiny to confirm if structural differences relate to improper assembly, however, we highlight that erroneous chromosome assemblies should be clearly identifiable through genomic disparity analysis, as their landmarks will be in substantially different compositions than the correctly assembled chromosomes of their close relatives.

A tool for studying structural change in genomes that complements existing methods. Dimensionality reduction is already in wide use in comparative genomics (e.g., uniform manifold approximation and projection, (58)). Genomic disparity analysis is therefore a natural companion method for assessing structural differences across the chromosomes of different species. It complements existing methods like WGA and synteny analysis by capturing a different dimension of how chromosomes vary across species.

To highlight differences in patterns captured by genomic disparity analysis and WGA, we used the five-species simulation dataset to directly compare UCE-based disparity to WGA sequence divergence levels. We line-wrapped chromosome assemblies using Bio Seq tool in Biopython (59), aligned each simulated chromosome with its corresponding ancestral chromosome using nucmer in Mummer 4.0 (60) and then computed pairwise genetic divergence measures using dnadiff in Mummer 4.0. Whereas substantial variance in genome architecture is detected between ancestral and descendent chromosomes in the simulations, WGA detects little to no change (**Table 2**). As such, disparity analysis captures a different dimension of genome evolution compared to alignment-based sequence divergence. This example also shows how changes to genomic architecture and sequence divergence are often decoupled.

Synteny analysis is a powerful tool for identifying similarities in gene order and arrangement. While synteny analysis does capture spatial variation (unlike WGA), it is known to be very sensitive to parameter settings (20). Furthermore, some methods of synteny inference are performed on post-WGA datasets meaning that structural information about the whole chromosomes has been removed *a priori* (16). In contrast, UCE-inferred estimates of genome structure offer at least two qualities that syntenic analyses lack (SI Appendix, **Fig. S7**). First, minor chromosomal rearrangements will be interpreted as breaks in synteny and ignored. In other words, inversions within syntenic regions not involving all the contiguous markers will break synteny. Second, large gap regions between contiguous orthologs can also break synteny (depending on parameter settings). The latter example highlights another important difference; disparity analyses will capture distance information between landmarks whereas synteny only provides landmark (ortholog) order information.

Avenues to expand the utility of genomic disparity. We see several ways that the approach described here could be further refined for more meaningful genomic comparisons. While we limited our validation analyses to using UCEs as landmarks and a handful of (largely homologous) chromosomes, landmarks could be any conserved sequences. Variation in the placement of conserved sequences can be used to study colocalization of transcribed genes (e.g. BUSCO [39]), cis-regulatory elements, enhancers, promoting regions, and more. The ethos of the approach is also scalable to three-dimensional genome structure once that information is more broadly available for many species (see (61, 62)). The extensive literature on geometric morphometrics offers substantial guidance for how 3D comparisons may be performed in the future (26). Landmark-based estimates of genomic disparity provide a rapid mechanism for comparisons of genome structure across species and populations, and we anticipate they will expedite discovery of novel patterns within comparative genomics.

Acknowledgments

We thank members of the Goswami Lab and NHM Herpetology Group for many helpful comments on experimental design and methods development. Several delegates at the 2023 Society for Integrative and Comparative Biology (Austin, Texas, USA) and 2023 Evolution (Albuquerque, New Mexico, USA) conferences provided valuable suggestions on earlier version of this work. We thank Christophe Liedtke for discussions on dimensionality reduction in genomics. We are indebted to Stephanie McClelland, Brant Faircloth and an anonymous reviewer

for comments that greatly improved the quality of this manuscript. We are grateful to Iain Milne for bioinformatics support through the UK Crop Diversity Bioinformatics HPC Resource (https://www.cropdiversity.ac.uk/). This study is part of a project funded by the European Commission through Marie Skłodowska-Curie Actions to A.V.M. (Grant Agreement ID: 101027832).

References

1. D. P. Genereux, *et al.*, A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**, 240–245 (2020).
2. S. Feng, *et al.*, Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).
3. A. B. Osmanski, *et al.*, Insights into mammalian TE diversity through the curation of 248 genome assemblies. *Science* **380**, eabn1430 (2023).
4. G. M. Rubin, *et al.*, Comparative Genomics of the Eukaryotes. *Science* **287**, 2204–2215 (2000).
5. X. Zhang, S. R. Wessler, Genome-wide comparative analysis of the transposable elements in the related species Arabidopsis thaliana and Brassica oleracea. *Proceedings of the National Academy of Sciences* **101**, 5589–5594 (2004).
6. Y. Yuan, *et al.*, Comparative genomics provides insights into the aquatic adaptations of mammals. *Proceedings of the National Academy of Sciences* **118**, e2106080118 (2021).
7. M. Lynch, B. Walsh, *The origins of genome architecture* (Sinauer associates Sunderland, MA, 2007).
8. K. Krupina, A. Goginashvili, D. W. Cleveland, Scrambling the genome in cancer: causes and consequences of complex chromosome rearrangements. Nature Reviews Genetics **25**, 196–210 (2024).
9. W. Hu, *et al.*, Nondestructive 3D Image Analysis Pipeline to Extract Rice Grain Traits Using X-Ray Computed Tomography. *Plant Phenomics* **2020**, 2020/3414926 (2020).
10. J. A. M. Graves, Evolution of vertebrate sex chromosomes and dosage compensation. *Nat Rev Genet* **17**, 33–46 (2016).
11. P. Banse, *et al.*, Forward-in-time simulation of chromosomal rearrangements: The invisible backbone that sustains long-term adaptation. *Molecular Ecology* **n/a**.
12. E. Lieberman-Aiden, *et al.*, Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
13. D. R. Schield, *et al.*, The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome Res.* **29**, 590–601 (2019).
14. J. Armstrong, *et al.*, Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
15. Y. Li, M. Zeng, F. Zhang, F.-X. Wu, M. Li, DeepCellEss: cell line-specific essential protein prediction with attention-based interpretable deep learning. *Bioinformatics* **39**, btac779 (2023).
16. K. Krasheninnikova, *et al.*, halSynteny: a fast, easy-to-use conserved synteny block construction method for multiple whole-genome alignments. *GigaScience* **9**, giaa047 (2020).
17. D. M. Larkin, *et al.*, Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res.* **19**, 770–777 (2009).
18. L. Chaney, A. R. Sharp, C. R. Evans, J. A. Udall, Genome Mapping in Plant Comparative Genomics. *Trends in Plant Science* **21**, 770–780 (2016).
19. J. T. Lovell, *et al.*, GENESPACE: syntenic pan-genome annotations for eukaryotes. 2022.03.09.483468 (2022).

20. D. Liu, M. Hunt, I. J. Tsai, Inferring synteny between genome assemblies: a systematic evaluation. *BMC Bioinformatics* **19**, 26 (2018).
21. S. Quigley, J. Damas, D. M. Larkin, M. Farré, syntenyPlotteR: a user-friendly R package to visualize genome synteny, ideal for both experienced and novice bioinformaticians. *Bioinformatics Advances* **3**, vbad161 (2023).
22. M. Foote, The Evolution of Morphological Diversity. *Annual Review of Ecology and Systematics* **28**, 129–152 (1997).
23. T. Guillerme, *et al.*, Disparities in the analysis of morphological disparity. *Biology Letters* **16**, 20200199 (2020).
24. A.-C. Fabre, *et al.*, Metamorphosis shapes cranial diversity and rate of evolution in salamanders. *Nature Ecology & Evolution* **4**, 1129–1140 (2020).
25. D. E. G. Briggs, R. A. Fortey, M. A. Wills, Morphological Disparity in the Cambrian. *Science* **256**, 1670–1673 (1992).
26. M. L. Zelditch, B. L. Lundrigan, T. Garland Jr., Developmental regulation of skull morphology. I. Ontogenetic dynamics of variance. *Evolution & Development* **6**, 194–206 (2004).
27. M. J. Christmas, *et al.*, Evolutionary constraint and innovation across hundreds of placental mammals. *Science* **380**, eabn3943 (2023).
28. G. Bejerano, *et al.*, Ultraconserved Elements in the Human Genome. *Science* **304**, 1321–1325 (2004).
29. W. Miller, *et al.*, 28-Way vertebrate alignment and conservation track in the UCSC Genome Browser. *Genome Res.* **17**, 1797–1808 (2007).
30. B. C. Faircloth, *et al.*, Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales. *Systematic Biology* **61**, 717–726 (2012).
31. A. Sandelin, *et al.*, Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics* **5**, 99 (2004).
32. N. K. Leclair, O. Anczuków, 'Poisoning' of the transcriptome by ultraconserved elements. *Nat Rev Mol Cell Biol* **23**, 777–777 (2022).
33. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
34. H. Li, Minimap2: pairwise alignment for nucleic acid sequences. *Bioinformatics* **34**, 3094-3100. (2018).
35. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. (2021).
36. N. Poikela, D. R. Laetsch, V. Hoikkala, K. Lohse, M. Kankare, Chromosomal inversions and the demography of speciation in *Drosophila montana* and *Drosophila flavomontana*. *Genome Biol. Evol*. **16**, evae024. (2024).
37. B. C. Faircloth, Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution* **8**, 1103–1112. (2017).
38. L. H. Yusuf, V. Tyukmaeva, A. Hoikkala, M. G. Ritchie, Divergence and introgression among the *virilis* group of Drosophila. *Evol. Lett.* **6**, 537–551. (2022).
39. M. Manni *et al*., BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* **38**, 4647–4654. (2021)
40. F. J. Martin, *et al.*, Ensembl 2023. *Nucleic Acids Research* **51**, D933–D941 (2023).
41. L. Anand, C. M. Rodriguez Lopez, ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. *BMC Bioinformatics* **23**, 33 (2022).
42. O. S. Harringmeyer, H. E. Hoekstra, Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nat Ecol Evol* **6**, 1965–1979 (2022).
43. J.-X. Yue, G. Liti, simuG: a general-purpose genome simulator. *Bioinformatics* **35**, 4442–4444 (2019).

44. J. C. Uyeda, D. S. Caetano, M. W. Pennell, Comparative Analysis of Principal Components Can be Misleading. *Systematic Biology* **64**, 677–689 (2015).
45. B. C. Phillips, N. Rodrigues, A. J. van Rensburg, N. Perrin, Phylogeography, more than elevation, accounts for sex chromosome differentiation in Swiss populations of the common frog (*Rana temporaria*). *Evolution* **74**, 644-654. (2020).
46. N. S. Upham, J. A. Esselstyn, W. Jetz, Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol* **17**, e3000494 (2019).
47. M. W. Pennell, *et al.*, geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* **30**, 2216–2218 (2014).
48. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other things): *phytools: R package*. *Methods in Ecology and Evolution* **3**, 217–223 (2012).
49. A. Goswami, *et al.*, Attenuated evolution of mammals through the Cenozoic. *Science* **378**, 377–383 (2022).
50. J. P. Bollback, SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* **7**, 88 (2006).
51. M. Ruta, K. D. Angielczyk, J. Fröbisch, M. J. Benton, Decoupling of morphological disparity and taxic diversity during the adaptive radiation of anomodont therapsids, *Proc. Roy. Soc. B.* **280**, 2013.1071. (2013).
52. M. C. C. De Pinna, Concepts and tests of homology in the cladistic paradigm. *Cladistics* **7**, 367–394 (1991).
53. S. Kirsch, *et al.*, Interchromosomal segmental duplications of the pericentromeric region on the human Y chromosome. *Genome Res.* **15**, 195–204 (2005).
54. L. Álvarez-González, *et al.*, Principles of 3D chromosome folding and evolutionary genome reshuffling in mammals. *Cell Rep* **41**, 111839 (2022).
55. P. Wang, F. Wang, A proposed metric set for evaluation of genome assembly quality. *Trends in Genetics* **39**, 175–186 (2023).
56. X. Liu, *et al.*, Effect of bFGF on fibroblasts derived from the golden snub-nosed monkey. *Primates* **62**, 369–378 (2021).
57. A. Warr, *et al.*, An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience* **9**, giaa051 (2020).
58. M. W. Dorrity, L. M. Saunders, C. Queitsch, S. Fields, C. Trapnell, Dimensionality reduction by UMAP to visualize physical and genetic interactions. *Nat Commun* **11**, 1537 (2020).
59. P. J. A. Cock, *et al.*, Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
60. G. Marçais, *et al.*, MUMmer4: A fast and versatile genome alignment system. *PLOS Computational Biology* **14**, e1005944 (2018).
61. O. Oluwadare, M. Highsmith, J. Cheng, An Overview of Methods for Reconstructing 3-D Chromosome and Genome Structures from Hi-C Data. *Biological Procedures Online* **21**, 7 (2019).
62. A. Lappala, *et al.*, Four-dimensional chromosome reconstruction elucidates the spatiotemporal reorganization of the mammalian X chromosome. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2107092118 (2021).

Collection: Methods for ecological and evolutionary data analysis

Paper type: Article

Figures and Tables

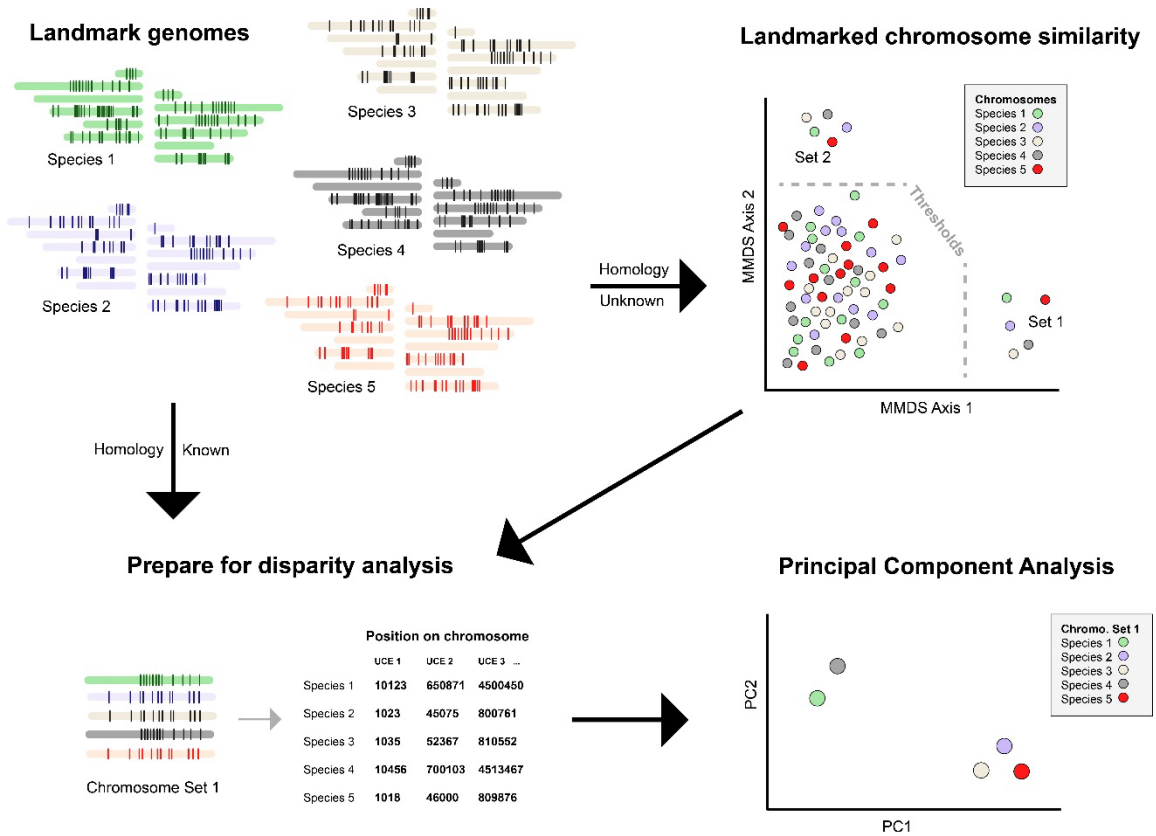

Figure 1. Schematic illustration of the landmarking pipeline for chromosome disparity analysis. The first step is reference mapping landmarks to chromosome-level genome assemblies of different species. If the homology of chromosomes is known, then the chromosomes are further prepared for disparity analysis. If homology is unknown, a presence-absence matrix is used to identify chromosomes with similar landmark compositions using a gap-based threshold for similarity. Chromosome sets are prepared for disparity analysis by extracting chromosome position information for each landmark. The resulting data matrix is used for multivariate Principal Component Analysis to visualize disparity across species.

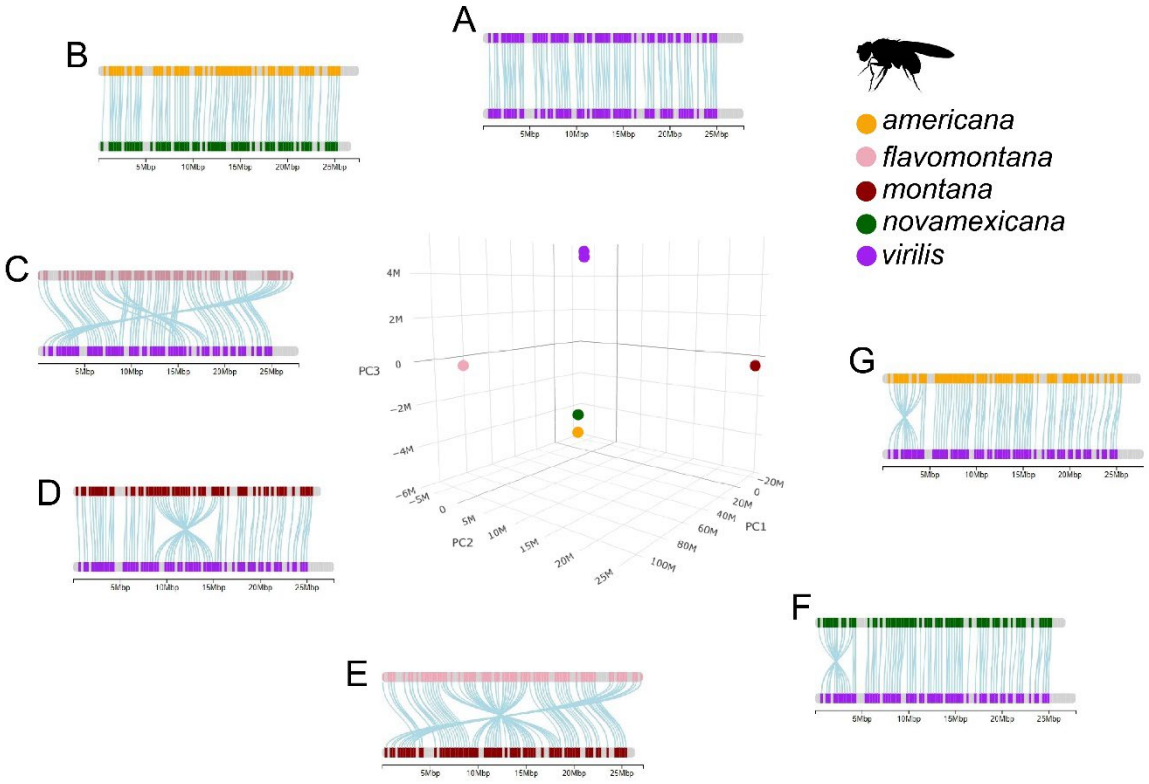

Figure 2. Genomic disparity analysis using 191 ultraconserved elements (UCEs) on chromosome 5 of the *Drosophila virilis* group, different colours correspond to different species in the complex. Synteny link maps around the main plot demonstrate what separation in geno-metric space corresponds to, in terms of landmark placement on chromosome 5. When samples are close to one another in space they are highly syntenic (A, B). Different axes correspond to discrete differences in landmark placement including two translocations (C), a pericentric inversion (D, E) and paracentric inversion (E, F). Synteny link maps were generated using chromoMap (41).

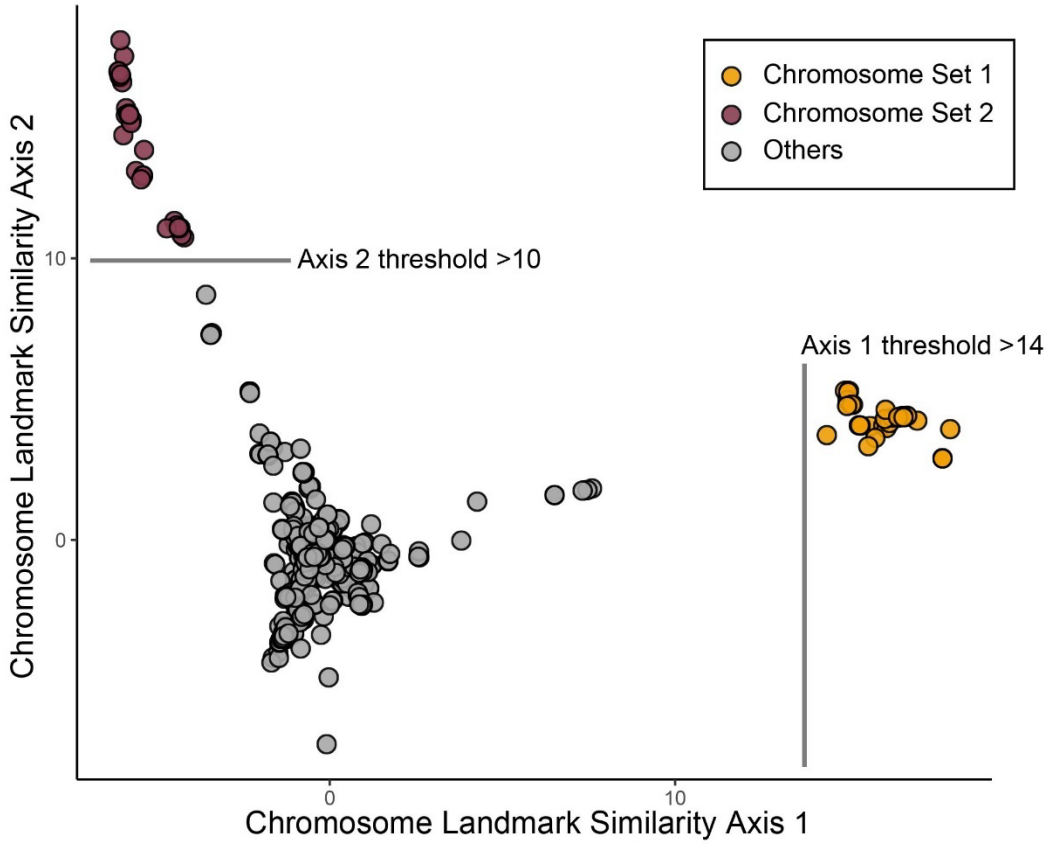

Figure 3. Metric multi-dimensional scaling (MMDS) of 1,004 landmarked placental mammal chromosomes based on patterns of ultraconserved element presence and absence. These chromosomes come from the genomes of 40 species (Table 1). MMDS axis thresholds are designated to establish 'chromosome sets' which contain a chromosome for most species included in the dataset. These thresholds correspond to chromosome sets that likely contain large homologous genomic blocks that are evolutionarily conserved and can be compared via genomic disparity analysis.

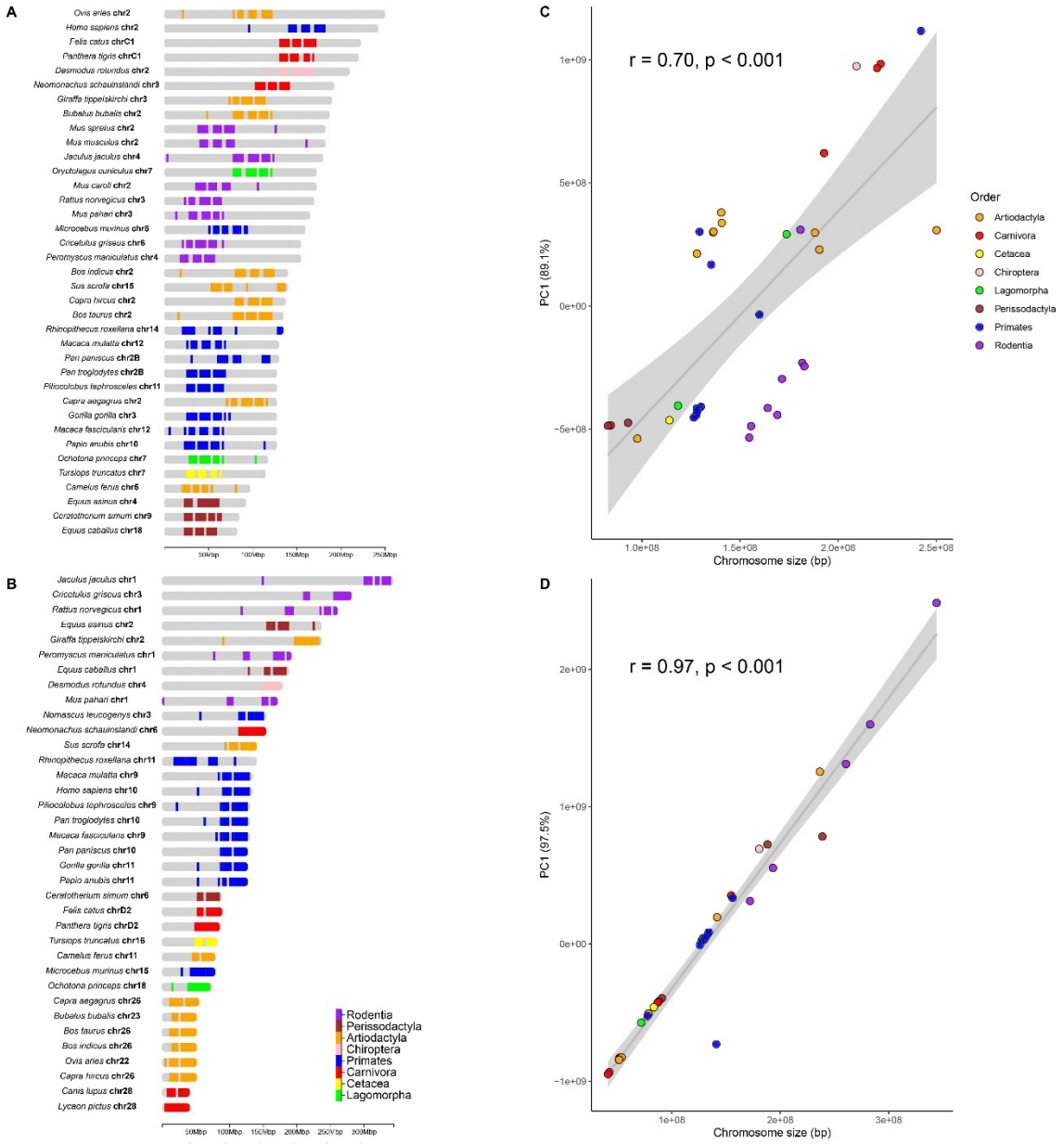

Figure 4. Distribution of ultraconserved element (UCE) landmarks for two placental mammal chromosome sets identified using metric multi-dimensional scaling (**Fig. 3**). Set 1 included 184 UCE landmarks from 37 species (A). Set 2 included 163 UCE landmarks from 36 species (B). The chromosome number (chr) is indicated next to each species name. Ordination results from a principal component (PC) analysis of landmark disparity comparing PC1 to chromosome size (with linear smoother functions added) for chromosome set 1 (C) and chromosome set 2 (D), respectively. Grey shading around regression lines corresponds to smoothed conditional means. Placental mammal orders included in this study are indicated by different colors.

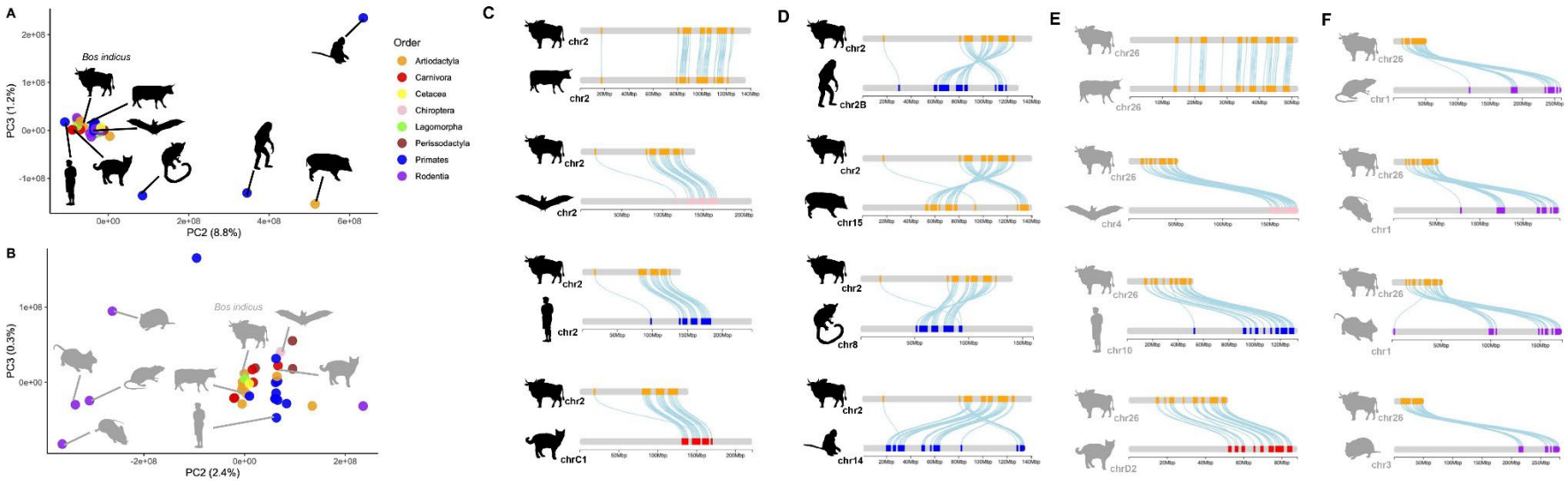

Figure 5. Ordination results from principal component (PC) analysis of landmark disparity comparing PC2 to PC3 for chromosome set 1 (A, 184 UCE landmarks from 37 species) and set 2 (B, 163 UCE landmarks from 36 species). Select synteny link maps to *Bos indicus* for species with similar (C) and dissimilar (D) PC2 scores in chromosome set 1 and similar (E) and dissimilar (F) PC2 scores for chromosome set 2. Placement in multivariate plot of each focal species is depicted with silhouetted icons (black for set 1, grey for set 2). The chromosome number (chr) is indicated next to each species silhouette in the synteny link maps. Synteny link maps were generated using chromoMap (41).

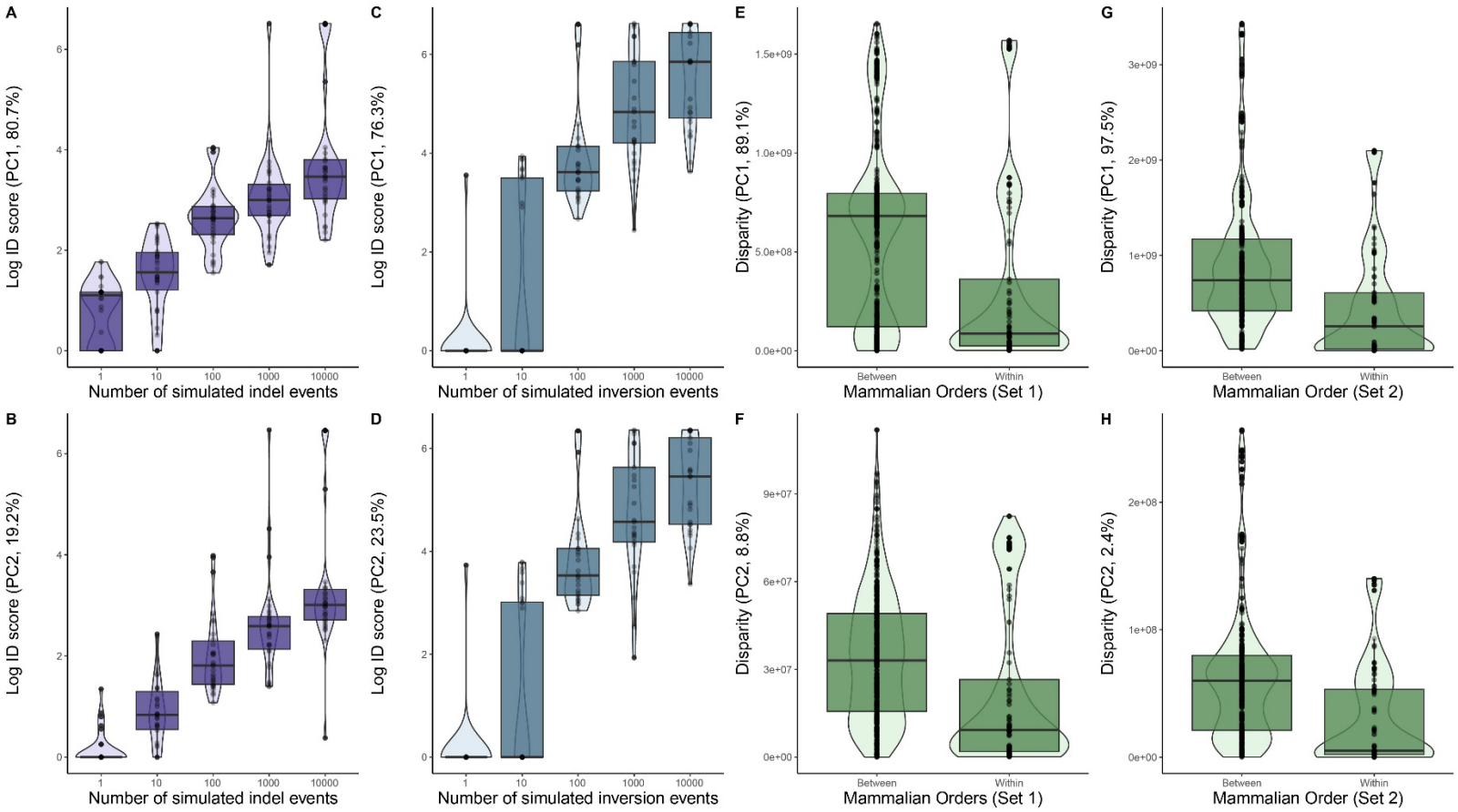

Figure 6. Log-transformed increase in disparity (ID) scores inferred from principal component (PC) analysis of ancestral and descendent insertion-deletion simulations for PC1 (A) and PC2 (B). Log-transformed ID scores inferred from inversion simulations for PC1 (C) and PC2 (D). Insertion-deletion simulation results are depicted in purple and inversion simulation results are depicted in blue. Observed empirical disparity scores for within and between mammalian orders for PC1 (E) and PC2 (F) of chromosome set 1 and PC1 (G) and PC2 (H) of chromosome set 2 are depicted in green.

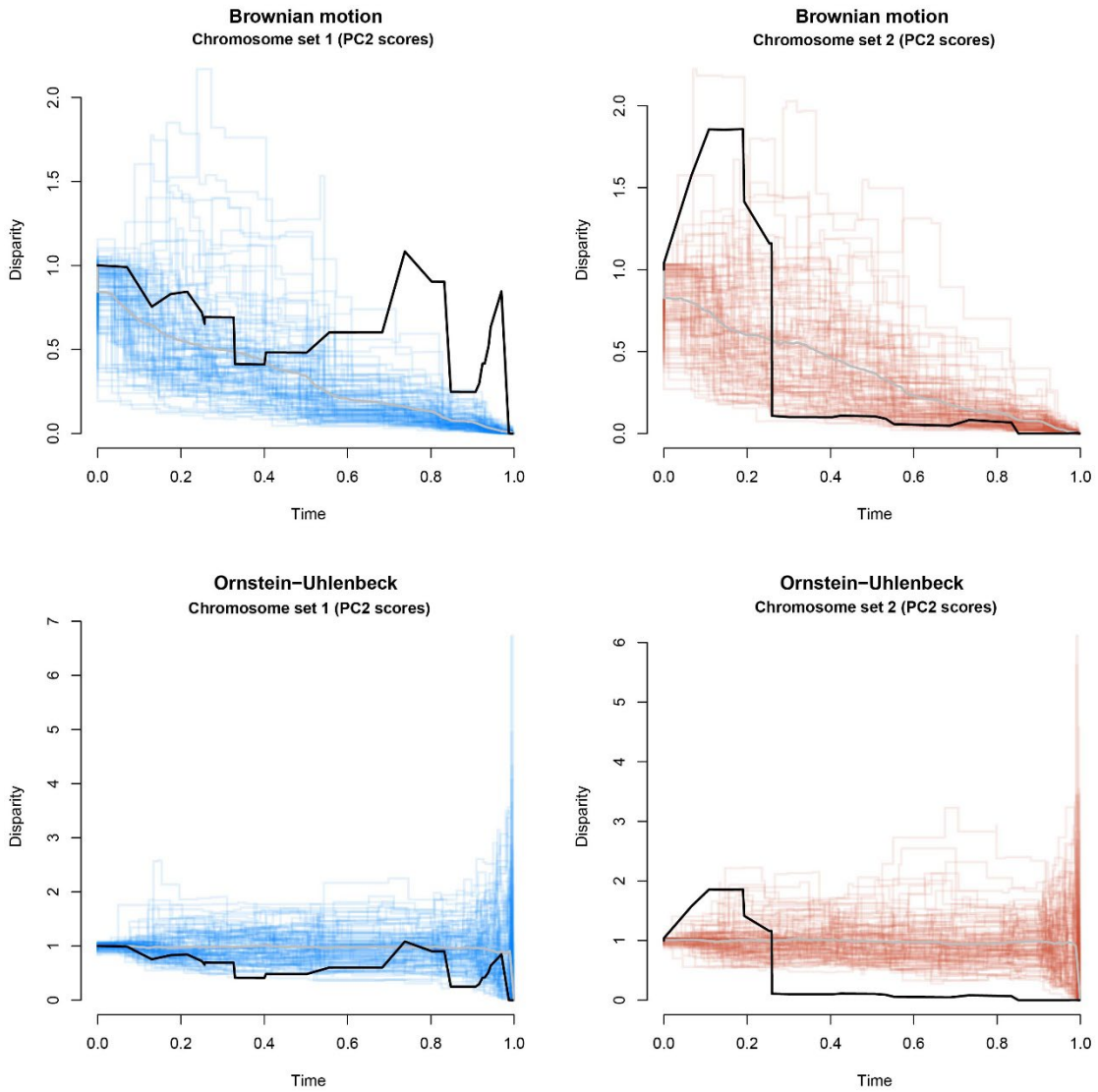

Figure 7. Disparity through time plots for Principal Component 2 (PC2) scores for chromosome set 1 (left side, blue) and chromosome set 2 (right side, red). Plots represent simulations conducted using models of Brownian motion (top) and Ornstein-Uhlenbeck adaptive optima (bottom). Black lines indicate the observed PC2 scores through time. Grey lines indicate the average disparity through time calculated from 100 simulations of trait evolution.

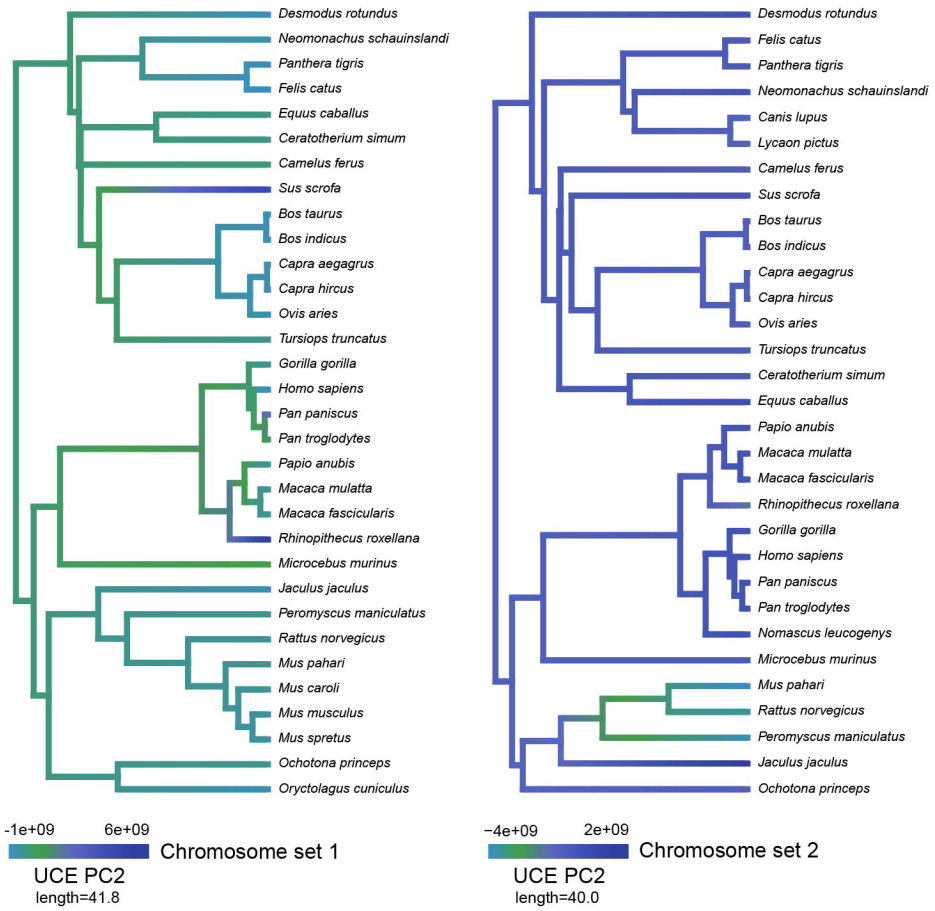

Figure 8. Stochastic character mapping of Principal Component 2 scores on mammalian phylogeny for chromosome set 1 (left) and chromosome set 2 (right). Phylogenies were obtained from Upham et al. (ref. 46).

Table 1. Chromosome sets identified from genome assemblies used to validate landmark-based estimates of genomic disparity. Those chromosomes appearing in bold text were used in the simulation/validation section of our study. CN = chromosome number (= reference number used in genomic assemblies), bp = base pairs, UCE = ultraconserved elements. References for genome assemblies are listed in the SI Appendix.

			Chromosome set 1 (184 UCEs)		Chromosome set 2 (162 UCEs)			
Species	Order	CN	Length	UCE landmarks	CN	Length	UCE landmarks	Reference (SI)
Bos indicus	Artiodactyla	2	140,680,885 bp	346	26	51,861,174 bp	235	ref. S16
Bos taurus	Artiodactyla	2	136,231,102 bp	353	26	51,992,305 bp	239	ref. S17
Bubalus bubalis	Artiodactyla	2	188,164,321 bp	453	23	52,095,120 bp	237	ref. S18
Camelus ferus	Artiodactyla	5	97,643,362 bp	355	11	78,620,977 bp	346	ref. S19
Capra aegagrus	Artiodactyla	2	128,023,632 bp	316	26	53,792,517 bp	233	ref. S20
Capra hircus	Artiodactyla	2	136,510,947 bp	356	26	51,421,553 bp	240	ref. S21
Giraffa tippelskirchi	Artiodactyla	3	190,444,849 bp	420	2	236,459,614 bp	562	ref. S22
Ovis aries	Artiodactyla	2	250,202,058 bp	569	22	51,512,491 bp	240	ref. S23
Sus scrofa	Artiodactyla	15	140,412,725 bp	393	14	141,755,446 bp	388	ref. S24
Canis lupus	Carnivora	N/A	N/A	N/A	28	42,479,636 bp	255	ref. S25
Felis catus	Carnivora	C1	221,611,373 bp	601	D2	88,083,857 bp	350	ref. S26
Lycaon pictus	Carnivora	N/A	N/A	N/A	28	41,181,707 bp	249	ref. S27
Neomonachus schauinslandi	Carnivora	3	192,837,575 bp	552	6	154,800,178 bp	565	ref. S28
Panthera tigris + P. leo	Carnivora	C1	219,880,392 bp	601	D2	86,999,983 bp	349	ref. S29
Tursiops_truncatus	Cetacea	7	114,015,307 bp	368	16	83,508,544 bp	355	ref. S30
Desmodus rotundus	Chiroptera	2	209,379,602 bp	507	4	180,635,192 bp	500	ref. S31
Ochotona princeps	Lagomorpha	7	118,371,147 bp	352	18	71,983,838 bp	319	ref. S32
Oryctolagus cuniculus	Lagomorpha	7	173,684,459 bp	441	N/A	N/A	N/A	ref. S33
Ceratotherium simum	Perissodactyla	9	84,132,471 bp	322	6	90,918,853 bp	357	ref. S34
Equus asinus	Perissodactyla	4	92,916,174 bp	324	2	238,820,862 bp	688	ref. S35
Equus caballus	Perissodactyla	18	82,641,348 bp	320	1	188,260,577 bp	614	ref. S36

Gorilla gorilla	Primates	3	127,988,552 bp	368	11	127,481,205 bp	418	ref. S37
Homo sapiens	Primates	2	242,193,529 bp	641	10	133,797,422 bp	423	ref. S38
Macaca fascicularis	Primates	12	127,712,166 bp	360	9	129,235,353 bp	426	ref. S39
Macaca mulatta	Primates	12	130,043,856 bp	361	9	134,124,166 bp	427	ref. S40
Microcebus murinus	Primates	8	159,902,609	448	15	77,693,121 bp	352	ref. S41
Nomascus leucogenys	Primates	N/A	N/A	N/A	3	156,076,922 bp	406	ref. S42
Pan paniscus	Primates	2B	129,381,034 bp	367	10	128,853,861 bp	416	ref. S43
Pan troglodytes	Primates	2B	128,755,405 bp	370	10	129,809,613 bp	421	ref. S44
Papio anubis	Primates	10	126,462,689 bp	358	11	125,913,696 bp	425	ref. S45
Piliocolobus tephrosceles	Primates	11	128,181,973 bp	358	9	131,283,114 bp	428	ref. S46
Rhinopithecus roxellana	Primates	14	135,287,999 bp	363	11	141,075,955 bp	425	ref. S47
Cricetulus griseus	Rodentia	6	155,611,870 bp	515	3	282,827,514 bp	720	ref. S48
Jaculus jaculus	Rodentia	4	180,757,125 bp	493	1	344,072,713 bp	816	ref. S49
Mus caroli	Rodentia	2	171,410,997 bp	536	N/A	N/A	N/A	ref. S50
Mus musculus	Rodentia	2	181,755,017 bp	543	N/A	N/A	N/A	ref. S52
Mus pahari	Rodentia	3	164,141,553 bp	519	1	172,151,128 bp	531	ref. S50
Mus spretus	Rodentia	2	182,815,943 bp	537	N/A	N/A	N/A	ref. S52
Peromyscus maniculatus	Rodentia	4	154,712,973 bp	504	1	193,310,054 bp	533	ref. S53
Rattus norvegicus	Rodentia	3	169,034,231 bp	517	1	260,522,016 bp	640	ref. S54

Table 2. Example of decoupling between genomic disparity and sequence divergence using simulated evolution of chromosomes. Comparison of average maximum increase in disparity (ID) ± S.E. and average maximum sequence divergence ± S.E. from whole chromosome alignments made using Mummer 4.0. Data are from simulated comparisons between ancestral chromosomes (set 1) and descendent chromosomes following simulation of random insertion-deletion (indel) and inversion events.

Species	Increase in disparity (ID)		Sequence divergence (dnadiff)	
	Indels	Inversions	Indels	Inversions
Capra aegagrus	4.21 ± 0.97	4.72 ± 1.31	0.005 ± 0.004	0.001 ± 0.0004
Homo sapiens	2,80 ± 0.36	3.08 ± 1.10	0.006 ± 0.003	0.001 ± 0.0003
Mus pahari	2.59 ± 0.40	3.49 ± 0.90	0.003 ± 0.003	0.000 ± 0.0001
Neomonachus schauinslandi	2.94 ± 0.50	4.32 ± 1.13	0.005 ± 0.003	0.002 ± 0.0009
Rhinopithecus roxellana	3.01 ± 0.57	3.26 ± 1.35	0.002 ± 0.004	0.002 ± 0.0001