Bioinformatics, 2026, **42**(6), btag315
https://doi.org/10.1093/bioinformatics/btag315
Published: 27 May 2026
Original Paper | **Genome Analysis**

OXFORD

Lift&Add—rapid and robust addition of new species to alignments of conserved non-coding sequences

Navya Shukla[1,2,6] **and Irene Gallego Romero**[1,3,4,5,*]

[1]Human Genomics and Evolution, St Vincent's Institute of Medical Research, Fitzroy, VIC 3065, Australia
[2]School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia
[3]Center for Genomics, Evolution and Medicine, University of Tartu, Tartu, 51010, Estonia
[4]Faculty of Health Sciences, Australian Catholic University, East Melbourne, VIC 3002, Australia
[5]School of Medicine, University of Melbourne, Melbourne, VIC 3053, Australia
[6]Present address: Children's Cancer Institute, 55 Botany St., Randwick, NSW 2031, Australia
*Corresponding author. Human Genomics and Evolution, St Vincent's Institute of Medical Research, Fitzroy, VIC 3065, Australia. Email: irene.gallego@svi.edu.au
Associate Editor: Pier Luigi Martelli

Abstract

Motivation: **Identifying sequence constraint across long evolutionary distances is a powerful method for the discovery of functional genomic sequences, especially putative non-coding elements. Conserved elements have been a mainstay of comparative genomic research, and can be further investigated for species-specific sequence acceleration to dissect the genetic basis of trait evolution. The conclusions of these comparative genomic studies are contingent on the number and range of species included in this phylogenetic analysis. However, while the number of metazoan genomes sequences is increasing rapidly, adding new genomes to existing whole-genome alignments remains computationally expensive.**

Results: **Here, we present a bioinformatic workflow, Lift&Add, that enables conserved elements, coding or non-coding, to be rapidly mapped to new genomes ("Lift") and subsequently be added to pre-existing multiple species alignments ("Add"), thus providing an avenue for easy exploration of these putative functional elements. Focusing here on a group of species that has been largely under-represented in genomic comparisons, the marsupials, we demonstrate the intuition behind this workflow and provide an example comparative genomic analysis that can be performed.**

Implementation and Availability: **Lift&Add is implemented as a series of scripts in Snakemake and bash, which can be downloaded from https://github.com/navyashukladr/Lift_and_Add.**

Introduction

The growing availability of genome-scale data in the last two decades has transformed both the breadth and resolution of comparative genetic investigations. Genome alignments of more than two species can clarify patterns of molecular evolution and have been particularly instrumental in the analysis of non-coding elements, by allowing the identification of intergenic elements that have undergone little sequence divergence over long evolutionary distances (Bejerano *et al*. 2004, Siepel *et al*. 2005, Woolfe *et al*. 2005, Miller *et al*. 2007, Visel *et al*. 2008, Lindblad-Toh *et al*. 2011, Christmas *et al*. 2023). These conserved non-coding elements are presumed to be of functional importance and can be further examined for lineage-specific variation—an excess of substitutions, also known as sequence acceleration. Acceleration can be indicative of positive selection, and has been used to isolate non-coding elements potentially underlying human-specific traits (Pollard *et al*. 2006, Prabhakar *et al*. 2008, Keough *et al*. 2023, Pal *et al*. 2025), wing development in bats (Booker *et al*. 2016) and hibernation in independent lineages of placental mammals (Ferris and Gregg 2019).

Over the last decade, the number of metazoan, and particularly vertebrate, genomes has risen rapidly (Feng *et al*. 2020, Rhie *et al*. 2021). However, generating the alignments of multiple genomes at the core of comparative genomics research is not trivial, and is made more difficult with increasing number, complexity and evolutionary distance between genomes (Saada *et al*. 2024). Consequently, updates to whole-genome alignments (WGAs) have been relatively slow (Kille *et al*. 2022) and not reflective of the large number of available high-quality genome sequences. A group of species that are particularly under-represented in WGAs are marsupial mammals. Existing genomic alignments of 60, 100 and 144 vertebrate taxa frequently used in comparative genomics include the same three species and assemblies—gray short-tailed opossum [monDom5, released in 2007 (Mikkelsen *et al*. 2007)], tammar wallaby [macEug2, released in 2009 (Murchison and Adams 2011)] and Tasmanian devil [sarHar1,

Received: 22 October 2025. **Revised:** 28 April 2026. **Accepted:** 1 May 2026

released in 2011 (Renfree *et al*. 2011)]. This reflects the historical sparsity of marsupial genomes, as until 2017 these were the only available. However, as of April 2025 there were reference genome assemblies for 27 marsupials available on NCBI. Yet these remain largely absent from most publicly available WGAs and can be challenging to incorporate into analyses.

Crucially, the range and number of taxa included in any comparative genomic study can profoundly affect inferences of constraint and divergence (Cooper *et al*. 2003, Boffelli *et al*. 2004, Eddy 2005, Stone *et al*. 2005, Hecker and Hiller 2020). For example, initial genomic comparisons of the naked mole rat with 10 other mammalian species identified a single amino acid change in a protein associated with hairlessness (HR) (Kim *et al*. 2011); this mutation was presumed to be phenotypically causal but subsequent investigation of the *HR* locus in 92 mammals found the same amino acid change in multiple other lineages that did not exhibit hairlessness (Delsuc and Tilak 2015). Similarly, when studying the impact of phylogeny on inferences of convergence in marine mammals, Thomas *et al*. (2017) found that as the number sampled taxa increased, the number of amino-acid substitutions called "convergent" decreased. A recent study by Bi *et al*. (2023) evaluated human acceleration in 49 primate species and showed that almost two-thirds of HARs curated from previous analyses (which sampled the primate phylogeny more sparsely) were in fact conserved among primates. Overall, these studies strongly suggest the importance of phylogenetic depth, in addition to breadth, particularly in the study of species-unique genetic innovation.

Previous work by our collaborators identified conserved non-coding elements that were under sequence acceleration in both the marsupial thylacine (*Thylacinus cynocephalus*) and placental wolf (*Canis lupus*), known as thylacine and wolf accelerated regions (TWARs) (Feigin *et al*. 2019), that could be potentially underlying their well-characterised convergent craniofacial morphology (Goswami *et al*. 2011, Feigin *et al*. 2018). This analysis however was limited to resources available at the time; Feigin *et al*. (2019) incorporated the thylacine genome into the 60-species WGA described above, which consists of 40 placental mammals but only 3 marsupial mammals (plus the thylacine) and 17 other vertebrate taxa. Intriguingly, the number of thylacine accelerated regions identified were almost tenfold (10 190) that of the wolf accelerated regions identified (1923). Concerned that the small number of marsupial genomes may have overestimated inferences of thylacine-specific acceleration, and motivated by the general challenges associated with studying non-coding elements in undersampled taxa, we devised an approach that would allow inclusion of novel genomes into comparative analysis of vertebrate conserved elements without requiring whole-genome alignments. We present a workflow—Lift&Add—that takes advantage of two genome annotation lift-over tools, (i) UCSC liftOver, which relies on existing pairwise genome alignments between genome assemblies and (ii) Liftoff, which maps genome annotations between assemblies with local alignment, as well as the sequence aligner MAFFT (i). We run and validate this workflow for a test dataset of alignments of conserved elements derived from the UCSC 60-way WGA of vertebrates, adding to this dataset 12 recently released marsupial genomes. We then briefly demonstrate a comparative genomic analysis that can be performed following the implementation of Lift&Add.

Methods

Defining a test dataset of vertebrate conserved elements for Lift&Add

We downloaded the WGA of 60 vertebrate species as well as the associated neutral model of evolution from the UCSC Genome Browser (https://hgdownload.soe.ucsc.edu/goldenPath/mm10/multiz60way/) (Rosenbloom *et al*. 2015). Alignments were filtered to remove sub-alignments with fewer than 30 species. Discrete conserved elements were predicted with the phastCons (PHAST version 1.5) `—most-conserved` option (Siepel *et al*. 2005, Hubisz *et al*. 2011). BEDtools (version 2.31.1) `—merge` and—`subtract` were used to merge elements within 10 bp of each other (Quinlan and Hall 2010). We filtered out elements shorter than 50 bp and those overlapping RepeatMasker annotations for mouse mm10 genome. The sequences were labelled with mouse coordinates (chr: start-stop) for identification. The UCSC mafFrags utility (https://github.com/ucscGenomeBrowser/kent) was subsequently used to extract discrete multiple species alignments (MSAs) for each vertebrate conserved element. The output Multiple Alignment Format (MAF) file was mapped to FASTA with the Galaxy MAF to FASTA tool (https://usegalaxy.org.au/root? tool_id=MAF_To_Fasta1), and split into a separate FASTA file for each vertebrate conserved element.

Workflow

We constructed a workflow consisting of a series of Snakemake and bash scripts to add sequences from target genomes to existing alignments of conserved regions (Table 1). The workflow is parallelized such that the key steps—mapping and alignment—can be implemented for multiple target genomes and chromosomes at the same time. The scripts and associated configuration files are available at https://github.com/navyashukladr/Lift_and_Add/. Details for each step are given in the following sections.

Snakefile, key steps

We implemented this Snakefile in parallel for each set of query and target genomes, that is, for the Tasmanian devil and its six target genomes, the Tammar wallaby and its six target genomes and the domestic dog and the wolf.

- *Mapping conserved elements from mouse to a query genome* We used UCSC liftOver (version 1.6.3) to map coordinates of the vertebrate conserved elements from the mouse (mm10) genome to the tammar wallaby (macEug2), Tasmanian devil (sarHar1) or domestic dog (canFam3) genomes. All chain files were downloaded from the UCSC Genome Browser (https://hgdownload.soe.ucsc.edu/goldenPath/mm10/liftOver/). We set the liftOver parameter -minMatch at 0.5, which requires 50% coverage for mapped sequences between the reference and target genomes. This was set at a more conservative threshold than the default 0.10 for inter-species coordinate conversion with liftOver, in order to obtain a high confidence test dataset. The output BED files are mapped to a Gene Transfer Format (GTF) with bed2gff.py [from CGAT

(Uchiyama *et al*. 2006), https://github.com/CGATOxford/cgat/blob/master/CGAT/scripts/bed2gff.py].

- *Mapping conserved elements from query to target marsupial genomes with Liftoff* We next used two rounds of Liftoff (Shumate and Salzberg 2021) to map coordinates of vertebrate conserved elements from a reference species to a closely related target species. Liftoff uses local alignment with minimap2 (Li 2018) to map sequences between the query and target genome. Inputs required include the query and target genomes and the GTF file of elements to be mapped. The Liftoff output consists of a GTF file of mapped elements and a list of unmapped ones. We used the -exclude-partial parameter to exclude partial alignments from the output as recommended by Liftoff—these are alignments that have lower than 50% alignment coverage and/or sequence identity between the query and target. We performed two rounds of Liftoff—in the first round, we mapped elements with the parameter -flank set at 0 (default). This parameter adjusts the amount of flanking genomic sequence added to a query element as a proportion of its total length. Next, for unmapped elements, we repeat Liftoff with -flank set a 1, which adds the maximum possible amount of flanking genomic sequence to the query element. We combined the mapped elements from both rounds of Liftoff into a single GTF file.
- *Getting sequences from each target genome*. We converted the combined GTF file of mapped elements from Liftoff to BED with BEDOPS (version 2.4.4.1) gtf2bed (Neph *et al*. 2012). We then extracted the nucleotide sequences for all the mapped elements with the BEDTools getfasta utility (Quinlan and Hall 2010). We obtained individual FASTA files per element with the UCSC faSplit utility (https://github.com/ucscGenomeBrowser/kent). Lastly, an output table is created summarising the species to which each element was mapped.

Adding new sequences to multiple species alignments of vertebrate conserved elements with MAFFT

For each element that was successfully mapped to at least one species, we generated compiled FASTA files consisting of sequences from all mapped species (3.mv files.sh, 4.combine_seqs.sh, Table 1). We then added the new species sequences to the original 60-way alignments with MAFFT—`add` (5.mafft_add.sh, 1). We used the following parameters:

- `—adjustdirectionaccurately`, to account for reverse complement sequences.
- `—keeplength`, to retain the structure of the original MSA, such that only conserved sequences are kept and any flanking sequence that is not conserved is trimmed.
- `—localpair`, L-INS-i local pairwise alignment strategy, the most accurate MAFFT algorithm appropriate for alignment of fewer than 200 sequences.

This step can be run in parallel per element if multiple cores are specified. As quality control for this step, we computed pairwise distances between species alignments with the ape package (version 5.8–1) in R (version 4.40), using `model = "raw"` in the function `dist.dna()`. Under this model, pairwise distance is simply defined as proportion of sites that differ between a pair of aligned sequences. For every alignment of a conserved element, we calculated the pairwise alignment distance between Tasmanian devil (sarHar1) or tammar wallaby (macEug2) and each new marsupial species in the alignment. For comparison, we also calculated the pairwise alignment distances between the Tasmanian devil (sarHar1), tammar wallaby (macEug2) and opossum (monDom5).

Runtime

Table 2 summarises the estimated runtime for each Lift&Add step. For the MAFFT step, run-time depends on the number of sequences to re-align rather than the number of species.

Using RERConverge to correlate evolutionary rates shifts with thylacine-wolf convergence

We followed two RERconverge (version 0.1.0) walkthroughs—`PhangornTreeBuildingWalkthrough.pdf` and (ii)

Table 1 Summary of scripts used to add genomic sequences from new genomes to alignments of vertebrate conserved elements.

Scripts	Steps
1. make_dir.sh	Makes individual output directories for each target species, per chromosome
2. Snakefile	Maps with liftOver (v.1.6.3) and then Liftoff (v.0.1.6) elements from reference to target genome. Extracts sequences of target coordinates from their genomes with BEDtools (v.2.31.1) and splits the output into individual fasta files per conserved element with UCSC faSplit. Per query genome, multiple target genomes can be specified. This workflow can be run in parallel across genomes and chromosomes if multiple cores are provided
3. mv_edit_files.sh	Moves individual conserved element fasta files into the appropriate directory. Also changes the header of each fasta file to the correct species name
4. combine_seqs.sh	Combines all mapped sequences (across different species) for each conserved element into a single fasta file
5. mafft_add.sh	Adds new genome sequences to existing alignments with MAFFT. Multiple cores can be specified to run each single re-alignment in parallel

Table 2 Estimated run-time for the Lift (liftOver and Liftoff, Snakefile) and Add (MAFFT) steps.

Step	Number of target genomes	Number of regions aligned	Number of cores	Approximate run-time (min)
liftOver and Liftoff	1	–	1	5
liftOver and Liftoff	6	–	10	10
MAFFT	6	1000	10	20
MAFFT	6	10 000	10	80

`FullWalkthroughUTD.pdf`. Firstly, we estimated for each MSA of a vertebrate conserved element a phylogenetic tree with the `estimatePhangronTreeAll()` function. We chose the widely used general time reversible (GTR) substitution model for phylogenetic estimation, which is also recommended for DNA sequences by RERConverge. RERConverge then was used to estimate the consensus tree from the individual trees, followed by calculation of the relative evolutionary rates (RERs) for all branches for each genomic element with `getAllResiduals()`. The old genomes for the Tasmanian devil (sarHar1), wallaby (macEug2) and domestic dog (canFam3) were excluded when calculating these RERs to remove redundancy, as newer genomes of these species (or of a closely related subspecies for the dog) were also present in the alignments.

We encoded thylacine-wolf convergence as a binary phenotype with the function `foreground2tree()`, setting `clade = 'terminal'`, as we were testing for rate shifts only in the terminal thylacine and wolf branches and not in the internal branches leading to the thylacine-wolf common ancestor (Partha *et al.* 2017, Kowalczyk *et al.* 2019). Supplementary Figure 1 displays the tree model for the binary phenotype, with the thylacine and wolf foreground lineages highlighted. With the function `correlateWithBinaryPhenotype()`, the RERs were associated with the binary trait. We set the parameter `min.pos = 20`, which requires a minimum of 20 species to be present in a MSA included in this analysis. The default setting is `min.pos = 10`, but given that the maximum number of species in this analysis was 73, this was increased to 20 to exclude highly sparse alignments.

Results

Optimising liftoff parameters to map short elements over long evolutionary distances

We sought to develop a computational workflow that would allow us to incorporate newly available marsupial genomes into comparative analysis of vertebrate conserved elements. The workflow has been summarised in Fig. 1, and is described in detail in the following sections. Briefly, our approach, which we call Lift&Add, uses UCSC liftOver, Liftoff and MAFFT to allow for straightforward incorporation of genome sequence alignments at broadly conserved regions. Most existing whole genome alignments of vertebrates are referenced to a human or mouse genome assembly, as are conserved elements derived from these alignments. UCSC liftOver can move annotations between genome assemblies; however, it requires chain files, that is, existing pairwise whole genome alignments between genomes. Liftoff, on the other hand, applies pairwise local alignment with minimap2 to map elements between a query genome and a target genome, an approach that works best between closely related taxa. Our workflow uses a combination of liftOver and Liftoff to map elements to new genomes in two steps, leveraging available chain files to minimise divergence times for Liftoff (Shumate and Salzberg 2021). Mapped elements are subsequently compiled and added to conserved sequence alignments with MAFFT (Katoh and Standley 2013). The final output consists of new alignments for vertebrate conserved elements covering an expanded breadth of taxa.

Existing documentation and published use of Liftoff is limited to genes, exons and transcripts. Thus we first evaluated Liftoff's performance with our mostly non-coding dataset by comparing its performance to that of to UCSC liftOver. In order to create a small test dataset to evaluate Liftoff with, we began by identifying 17 155 discrete vertebrate conserved elements >50 bp aligning to the shortest mouse chromosome, chromosome 19, in the original 60-way WGA (without the thylacine) with PHAST. Annotating our dataset against mm10, which the WGA is referenced to, we determined that 28.0% of elements (4796) overlapped protein coding sequences (CDS), 5.5% (949) 3′ or 5′ untranslated regions (UTRs) and the remaining 66.5% (11 410) were intergenic or unannotated. We then used liftOver and existing chain files to map 13 178 of these elements from mouse (mm10) to the opossum (monDom5) (Supplementary Fig. 2, Supplementary Table 1).

To evaluate the performance of Liftoff on this heterogenous test dataset, we then mapped these elements from the opposum (monDom5) to the Tasmanian devil (sarHar1) with either liftOver or Liftoff, and compared the results (Fig. 2A; Duchêne *et al.* 2018). Given the long divergence time between these taxa [time to most recent common ancestor (TMRCA) ≈80–90 million years], substantially fewer elements mapped with Liftoff (5699 of 13 178, 43.2%) than with liftOver (11 207, 85.0%) (Fig. 2B). Of the 5469 elements successfully mapped with both liftOver and Liftoff, 5448 had overlapping output coordinates between the two methods (with an average 99.0% of the liftOver output bases overlapping with Liftoff output bases). Similar proportions of protein-coding (44.2%), UTR (42.4%) and intergenic/unannotated (51.4%) elements were successfully mapped with Liftoff (Supplementary Table 2); however Liftoff performed significantly worse with shorter elements [mean size (bp) mapped elements 200.5, unmapped elements = 104.3; pairwise *t*-test *P*-value $< 2.2 \times 10^{-16}$] (Fig. 2C). We performed an identical comparison for three other pairs of mammalian genomes (each

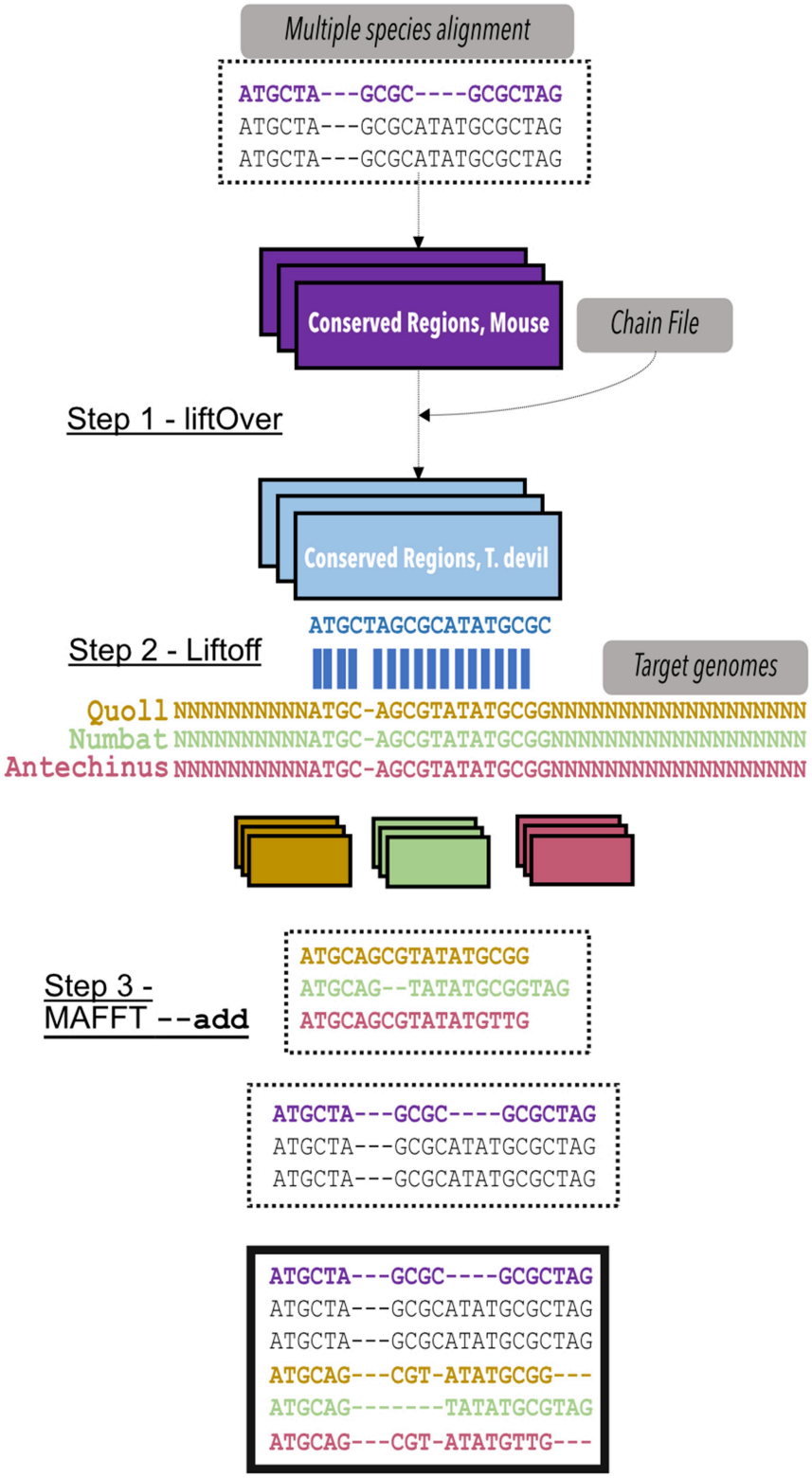

Figure 1 Workflow to add sequences from new genomes to alignments of conserved elements. First, liftOver is used to map mouse coordinates of vertebrate conserved elements to marsupials with available chain files (here, the Tasmanian devil). Second, we map coordinates of vertebrate conserved elements between marsupials with Liftoff. The target genomes here are new marsupial assemblies that are not present in the whole-genome alignment or have chain files with mouse. Third, for each mapped element we add sequences from the newer marsupial genomes to the existing multiple-species alignments.

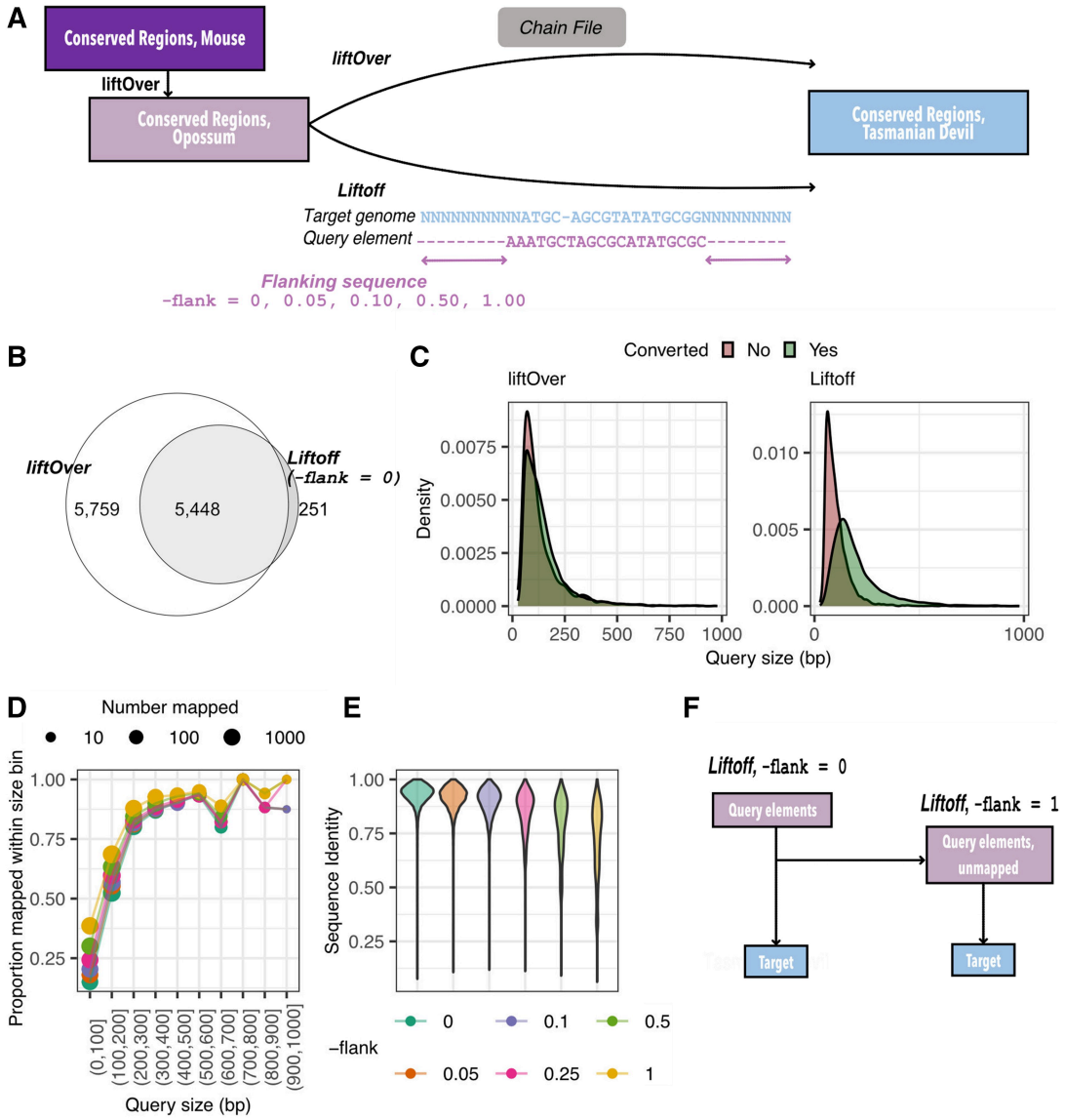

Figure 2 Testing Liftoff. (A) Coordinates for vertebrate conserved elements were mapped from opossum to Tasmanian devil with both liftOver and Liftoff. (B) Overlap between output from liftOver and Liftoff. (C) Distribution of sizes of elements that did and did not successfully map from opossum to devil with liftOver and Liftoff. Elements greater than $>$ 1000 bp in length are excluded. (D) Proportion of opossum query elements in each 100 bp size interval successfully mapped at different thresholds of `-flank`. (E) Sequence identity between query opossum and target Tasmanian devil elements at different threshold of `-flank`. (F) Final Liftoff strategy used in our workflow.

diverged approximately 50–60 MYA), and observed this same dependence on size (Supplementary Fig. 3).

Liftoff allows flanking genomic sequence to be symmetrically added to query elements (parameter `-flank`, default 0). We found that increasing the total amount of flanking sequence added stepwise (`-flank` = 0.05, 0.10, 0.25, 0.5 and 1, where these values are multiples of the original query length) consistently increases the number of elements mapped (Supplementary Fig. 4A, Supplementary Table 3). Elements only mapping at `-flank` $>$ `0` were significantly shorter than those mapped at `-flank = 0` (pairwise *t*-test *P*-value $< 2.2 \times 10^{-16}$), demonstrating that adding flanking sequence improves Liftoff performance for shorter elements, primarily for those between 0 and 200 bp (Fig. 2D, Supplementary Fig. 4B). However, genomic sequence flanking vertebrate conserved elements is likely to be under less constraint itself and therefore may reduced alignment specificity. Indeed, with added flanking sequence the average alignment coverage and sequence identity between the aligned query and target elements decreases, and the number of elements that are partially mapping (i.e. with alignment coverage and/or sequence identity $\leq$ 0.5, as defined by Liftoff) increasing from 44 at `-flank=0` to 806 at `-flank=1` (Fig. 2E, Supplementary Fig. 4C, Supplementary Table 3).

To minimize partial mapping, we tested a successive approach, whereby we first use Liftoff to map elements between query and target with no flanking sequence (`-flank=0`) (Supplementary Fig. 5). Subsequently, for sequences that do not map/partially map, we repeat Liftoff ("re-map") with `-flank=1` to maximize recovery. Initial Liftoff maps 5699 elements (43.2% of the opossum dataset) to the Tasmanian devil genome. An additional 2344

elements (31.3% of previously unmapped elements) re-map with the inclusion of flanking sequence. Only 191 elements displayed partial mapping (0.02% of 7999); excluding these, the total number of successfully mapped elements was 7805 (59.2% of the query dataset). This was higher than a single round of Liftoff at any threshold of `-flank` (Supplementary Table 3). Of these 7805 elements, 7370 have overlapping coordinates with the output of liftOver from opossum to the Tasmanian devil (average overlap 99.2%), 391 were mapped with Liftoff but not liftOver and 44 display disagreement between the two methods.

Lifting vertebrate conserved elements to new genome assemblies

Our results with the Tasmanian devil and opossum suggested that Liftoff could be used to map these vertebrate conserved elements (both coding and non-coding elements) between more closely related taxa. Focusing still on marsupial taxa, we used UCSC liftOver to map the test dataset of conserved elements to the two marsupial genomes with publicly available mm10 chain files—the Tasmanian devil (sarHar1/Devil_ref_v7.0) and tammar wallaby (macEug2/Meug_2.0) (Supplementary Fig. 2, Supplementary Table 1). We then curated 10 new genome assemblies of species from 8 marsupial families not represented in the UCSC 60-way WGA of vertebrates [including the newer thylacine genome assembly (Feigin *et al*. 2022)] as well as newer assemblies for Tasmanian devil (mSarHar1.11, 2022) and tammar wallaby (mMacEug1, 2021) (Supplementary Table 4) (Rosenbloom *et al*. 2015). The older tammar wallaby assembly [macEug2 (Murchison and Adams 2011)] was used as the Liftoff query genome for the six new marsupials within Diprotodontia, while the Tasmanian devil (sarHar1) (Renfree *et al*. 2011) was used as query genome for the five marsupials within Dasyuromorphia. Our final taxa, the monito del monte (*Dromiciops gliroides*, mDroGli1) is the only extant species in the order Microbiotheria, and equally divergent from Tasmanian devil and tammar wallaby (TMRCA 60–70 million years) (Duchêne *et al*. 2018). We used the Tasmanian devil genome (sarHar1) as query genome in this instance due to its higher contiguity and completeness relative to the tammar genome (macEug2) (Murchison and Adams 2011, Renfree *et al*. 2011, Nishimura *et al*. 2017).

Starting with the same set of mm10 chromosome 19 elements as above, we mapped with liftOver 12 345 and 10 692 to the Tasmanian devil (sarHar1) and tammar wallaby (macEug2) genomes, respectively (Table 3). With Liftoff, we then mapped all 12 345 vertebrate conserved elements from Tasmanian devil to one or more of its target genomes, and 10 688 (99.9%) conserved elements from tammar wallaby to one or more of its target genomes (Fig. 3A and B, Supplementary Fig. 6A and D). For both groups, a majority of the elements mapped to all six species—8564 (69.4%) from the Tasmanian devil and 8098 (75.7%) from the tammar wallaby. Elements that mapped to all new target species from the Tasmanian devil or the tammar wallaby were significantly longer than those that did not, consistent with size being a primary driver of success with Liftoff (pairwise *t*-test Dasyuromorphia: P-value$= 1.386 \times 10^{-5}$, Diprotodontia: P-value$= 8.718 \times 10^{-145}$) (Supplementary Fig. 6E and F). No bias was observed in the type of elements—CDS, UTR or intergenic/un-annotated—that were successfully mapped (Fig. 3C and D).

The percentage of elements mapped was highly negatively correlated with the evolutionary distance between query and target genome [Spearman's ρ −0.95 (P-value $= 1.46 \times 10^{-6}$)] (Fig. 3E). This fits expectations of sequence decay over time, as even for ultraconserved elements, alignability across species decreases proportional to the phylogenetic distance between them (Miller *et al*. 2007, Cummins *et al*. 2024). Nearly all query elements mapped across different assemblies of the same species (98.9% from sarHar1 to the mSarHar1.11 for Tasmanian devil, and 99.9% from macEug2 to mMacEug1 for tammar wallaby), whereas the lowest percentage of mapped elements was between the distantly related Tasmanian devil and monito del monte (74.9%; estimated divergence time ∼ 60–70 MYA). Additionally, the proportion of elements that could not be mapped with `-flank=0`, but mapped at `-flank=1` was

Table 3 Liftoff results for coordinate conversion from the Tasmanian devil/tammar wallaby to 12 new marsupial assemblies.

Reference	Common Name	Estimated divergence from reference (MYA)	Mapping at `-flank=0`	Mapping at `-flank=1` but not `-flank=0`	Total mapped	Percentage mapped (%)
Tasmanian devil (sarHar1)	Tasmanian devil (mSarHar1.11)	0	11 893	326	12 219	99.0
	Eastern quoll (DasVivv1.0)	0–10	11 410	904	12 314	99.7
	Brown antechinus (AStuM)	10–20	10 589	1601	12 190	98.7
	Numbat (mMyrFas1)	25–35	8747	2357	11 104	89.9
	Tasmanian tiger (ThyCyn2.0)	25–35	8805	2456	11 261	91.2
	Monito del monte (mDroGli1)	60–70	6674	2576	9250	74.9
Tammar wallaby (macEug2)	Tammar wallaby (mMacEug1)	0	10 338	347	10 685	99.9
	Woylie (mBetpen)	20–30	8876	1653	10 529	98.5
	Sugar glider (Puasm1.0)	45–55	6721	2544	9265	86.7
	Common brushtail possum (mTriVul1)	45–55	7128	2503	9631	90.1
	Koala (phaCin4)	50–60	6768	2492	9260	86.6
	Common wombat (GCA900497805.2)	50–60	6724	2553	9277	86.8

correlated positively with divergence time (Spearman's $\rho = 0.98$, P-value $= 3.181 \times 10^{-8}$), confirming that as sequence divergence between query and target increases, adding flanking sequence and increasing query length has a greater impact on local alignment quality (Supplementary Fig. 6G). Lastly, there was no overall association between contig N50 and percentage of query elements mapped (Spearman's $\rho = -0.19$, P-value = .54), likely due to the short sizes of the query vertebrate conserved elements [query element mean length, Tasmanian devil (sarHar1): 161.9 bp, tammar wallaby (macEug2) 147.2 bp] (Fig. 3F).

Aggregating results from the two clades, we successfully mapped 13 297 vertebrate conserved elements from Tasmanian devil (sarHar1) and tammar wallaby (macEug2) to at least one of 12 new genomes. Excluding the newer assemblies for Tasmanian devil (mSarHar1.11) and tammar wallaby (mMacEug1), 13 288 (99.9%) mapped to at least one of 10 new genomes (Supplementary Fig. 7). Of these, 5986 (45.0%) mapped to all 10 species. Only 77% of the 17 155 vertebrate conserved elements from mouse chromosome 19 mapped to at least one new marsupial genome (excluding newer assemblies); considering only alignment blocks that contain a marsupial (14 425), we were able to map the vast majority (92.1%) to new marsupial genomes.

To confirm the generalisability of this approach, we also mapped elements between other mammalian query and target genomes, including at least one pair from each of the placental superorders and a monotreme example. These results are presented in Supplementary Table 5.

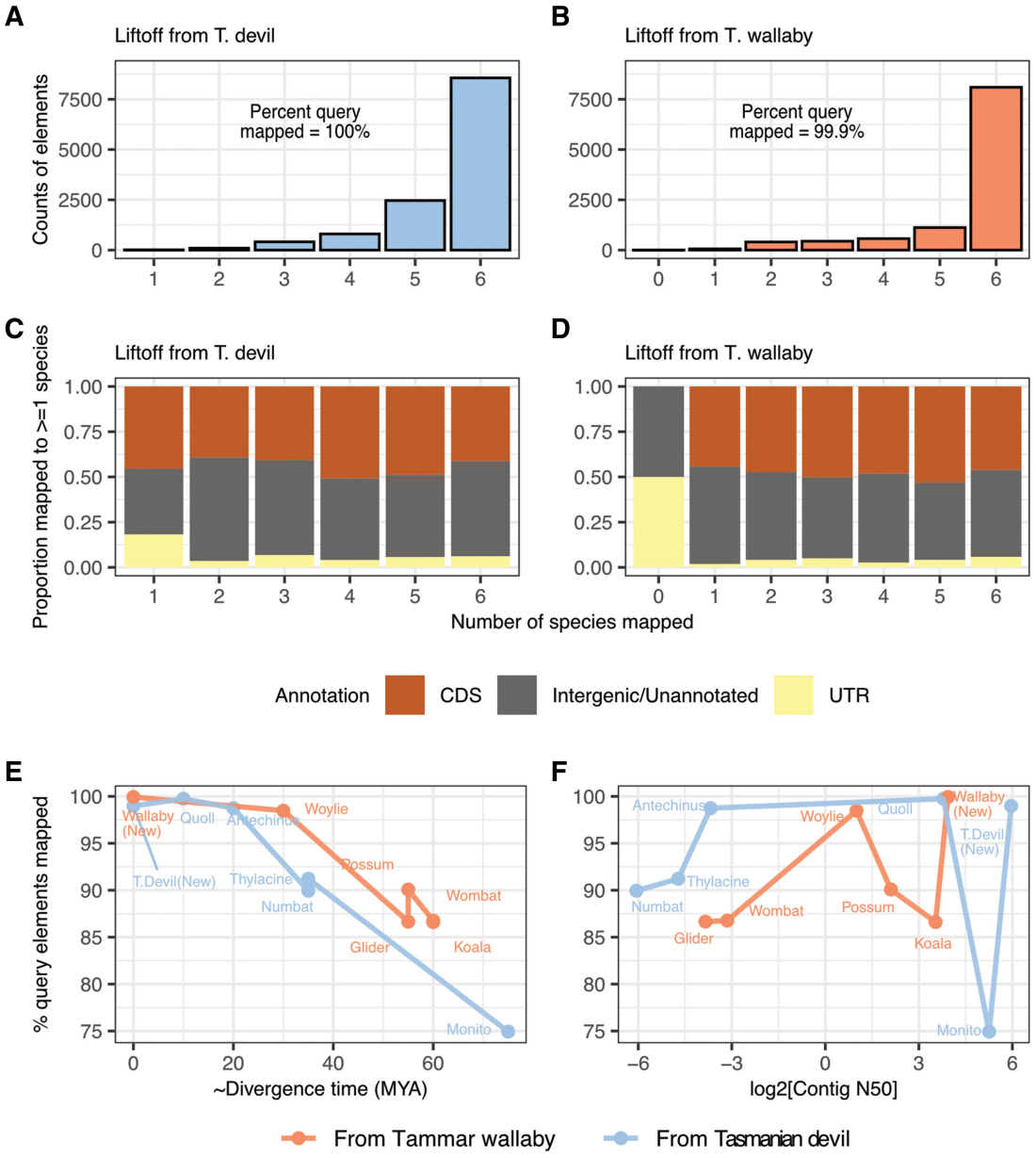

Figure 3 Mapping conserved elements from Tasmanian devil (sarHar1) and tammar wallaby (macEug2) to target marsupial genomes. (A and B) Counts of query elements that were mapped to 0 or more species. (C and D) Annotations for query elements that mapped to 0 or more species. (E) Estimated divergence time versus percentage of query elements mapped. (F) Genome contiguity (Contig N50) versus percentage of query elements mapped.

Lift&Add results correlate with inter-species karyotypic relationships

As a means of confirming the accuracy of our approach, we examined the genomic order of mapped elements in the new target genomes (as illustrated in Supplementary Fig. 8). Karyotypes are highly conserved across species in Dasyuromorphia (chromosome number $2n = 14$) (Deakin 2018). Thus, we expected the genomic order of mapped elements to be co-linear between the query Tasmanian devil genome and the target Dasyuromorphia genomes. However, the query Tasmanian devil (sarHar1) genome assembly was only scaffold-level. Instead, we assessed co-linearity between the newer chromosome-level Tasmanian devil genome assembly (mSarHar1.11) to the other Dasyuromorphia target genomes, noting that since our test dataset includes only conserved elements in mouse chromosome 19, we do not expect to cover the whole genome in any species. As expected, we observed near-perfect co-linearity in genomic order of conserved elements (Spearman's ρ = 0.98, *P*-value $< 2.2 \times e^{-16}$) between the closely related Tasmanian devil and quoll genomes (divergence time ≈5–10 MYA) (Deakin 2018) (Fig. 4A). A perfect inversion in the order of elements from chromosome 6 when comparing between mSarHar1 and the Eastern quoll (DasVivv1.0) could be indicative of structural variation but could also originate from differences in chromosome assembly and annotation.

Comparing the results of mapping to the new Tasmanian devil assembly (mSarHar1.11) and the more distantly related extinct thylacine (divergence time ∼ 25–35 MYA) we again found a high degree of co-linearity of mapped elements (Spearman's $\rho = 0.96$, *P*-value $< 2.2 \times e^{-16}$) (Fig. 4). However, there are more elements that show discordant genomic order between the new Tasmanian devil genome and the thylacine genome. The thylacine chromosome-level assembly (ThyCyn2.0, 2022) was derived from short contigs (contig N50 0.015 Mb) that were assembled into larger scaffolds with reference to mSarHar1.11 (Feigin *et al.* 2022); as such we speculate that this discordance was likely due to instances of mis-assembly of the thylacine genome due to the challenging nature of the original data. By examining the coordinates of these conserved elements between the two genomes (Supplementary Fig. 9), we find support for this theory, with three clusters of elements from likely repetitive regions in the Tasmanian devil mapping to variable positions in the thylacine.

Lastly, Spearman's correlation in the genomic order of mapped elements between the Tasmanian devil and distantly related monito del monte (which also has a karyotype $2n = 14$) was 0.52 (*P*-value $< 2.2 \times e^{-16}$) (Fig. 4C). Elements from chromosome 1 and chromosome 6 display linear agreement in order; a near-perfect inversion of coordinates for elements in chromosome 2 lowered the overall correlation. As with the quoll comparison, this likely reflects differences in genome assembly.

As with the Tasmanian devil, the older tammar wallaby assembly (macEug2) was also scaffold-level, but the *newer* assembly (mMacEug1) was chromosome-level; we thus again compared order of mapped elements in this new assembly to their order in the brushtail possum genome (mTriVul1), the only other Diprotodontia chromosome-scale assembly. The Diprotodontia taxa display high variability in karyotypes with several instances of chromosomal reshuffling in their evolutionary history, particularly in the Macropodiformes, and the brushtail possum has two more chromosomes ($2n = 20$) than the tammar wallaby ($2n = 16$) (TMRCA 45–55 MYA) (O'Neill *et al.* 2004, Deakin 2018). Overall, we observed a strong (but negative) correlation in genomic order between species (Spearman's $\rho = -0.93$, *P*-value $< 2.2 \times e^{-16}$) (Fig. 5A), driven primarily by strong collinearity between tammar wallaby chromosomes 1 and 2 and brushtail possum chromosomes 8 and 6 respectively, although tammar chromosome 1 and possum chromosome 8 are inverted relative to each other. However, results for elements mapped to tammar wallaby chromosomes 3 or X were more variable.

To better understand their complex cytogenetic relationship, we referred to chromosome painting results from (Rens *et al.* 1999) (Fig. 5B, Supplementary Fig. 10, Supplementary Table 6). In this work, Rens *et al.* (1999) showed that chromosome 2 of the tammar wallaby hybridized to chromosomes 6 and 3 of the

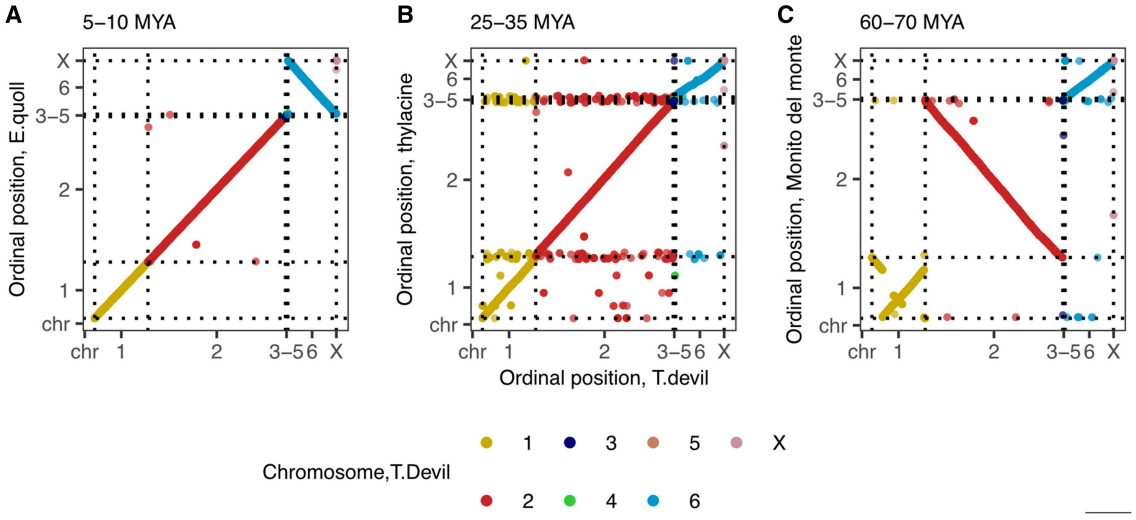

Figure 4 Correlation in genomic order of mapped elements between (A) Tasmanian devil (mSarHar1.11) and Eastern quoll (DasVivv1.0), (B) Tasmanian devil and Thylacine (ThyCyn2.0) (C) Tasmanian devil and Monito del monte (mDroGli1). Elements are coloured by their chromosomal location in the Tasmanian devil.

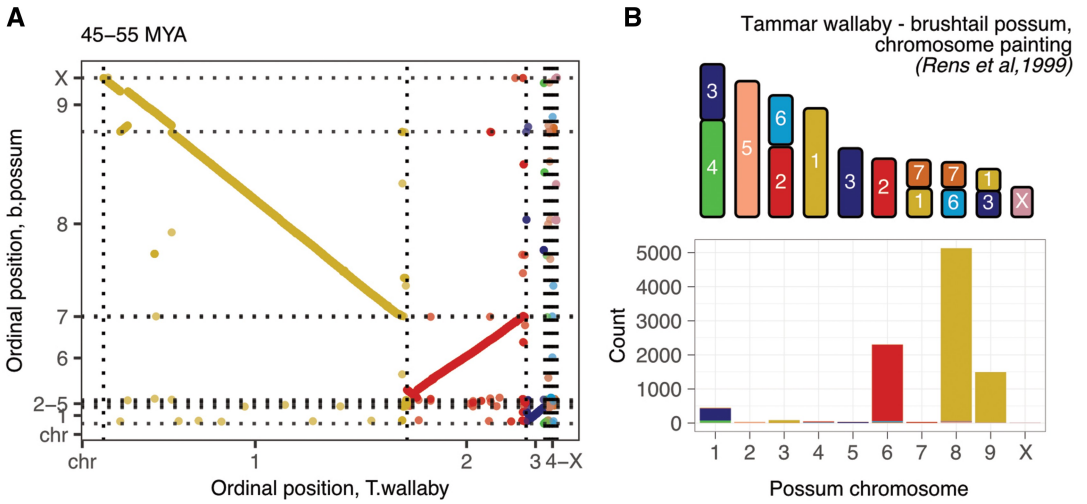

Figure 5 Tammar wallaby and brushtail possum. (A) Correlation in genomic order of mapped elements between tammar wallaby (mMacEug1) and brushtail possum (mTriVul1). Elements are coloured by their chromosomal location in the tammar wallaby. (B) Relationship between the tammar wallaby and brushtail possum karyotypes, as determined by Rens *et al*. (1999), and counts of test elements mapping to each brushtail possum chromosome, coloured by their location in the tammar wallaby genome.

brushtail possum, aligning with our observation that 95.9% of elements from tammar wallaby chromosome 2 mapped to chromosome 6 of the brushtail. Chromosome 3 of the tammar wallaby hybridized to brushtail possum chromosomes 1, 5 and 9 in the original chromosome painting results; in our data 91.4% of tammar wallaby chromosome 3 elements mapped to brushtail possum chromosome 1 and an additional 6.2% mapped to brushtail possum chromosome 5. However, chromosome painting results suggested that tammar wallaby chromosome 1 is alignable to chromosome 4, 7 and 9 of the brushtail possum, in our data the majority (98.4%) of elements from tammar wallaby chromosome 1 map to either chromosome 8 and 9 of the brushtail possum. Looking closer, we found that chromosome 7 and 8 of the brushtail are 275 and 267 Mb respectively, and likely difficult to differentiate cytogenetically. Altogether, we were able to show that our results agreed with expectations of both karyotype conservation in Dasyuromorphia and karyotype rearrangement in Diprotodontia.

Generating new multiple-species alignments with MAFFT

Having mapped vertebrate conserved elements to new marsupial genomes, the next step in our pipeline is to incorporate the relevant sequences back to the corresponding multiple-species alignments with MAFFT—`add`, as a prerequisite for testing for acceleration. An additional parameter with MAFFT, `—keep-length` allows the structure of the original alignment to be preserved when new sequences are added: if the new marsupial sequences are longer than the alignment, MAFFT trims the unaligned bases, such that only the original sequence is retained, controlling for the inclusion of flanking sequence in the steps above.

To evaluate the resulting alignments, we leveraged sequences from existing marsupials in the 60-way WGA, reasoning that they should align well with added marsupial sequences (Fig. 6). For every alignment of a conserved element, we calculated the pairwise alignment distance (proportion of sites that differ between two sequences in an alignment) between the Tasmanian devil (sarHar1) or tammar wallaby (macEug2) sequences and each new marsupial sequence (Fig. 6B and C). Overall, the sequences of the added Dasyuromophia species were significantly more similar to Tasmanian devil (sarHar1) than sequences from the Diprotodontia species or the opossum (Supplementary Table 7). Conversely, sequences of all added Diprotodontia species were significantly more similar to tammar wallaby (macEug2) than sequences from the Dasyuromorphia species, the monito del monte or the opossum (Supplementary Table 8). For each species, the mean pairwise distance from the Tasmanian devil was positively correlated with divergence time [Spearman's ρ 0.80 (*P*-value .0009)], as was the mean pairwise distance from the tammar wallaby [Spearman's ρ 0.86 (*P*-value $= 7.31 \times 10^{-5}$)] (Fig. 6D and E). These observations suggested that the new alignments were largely congruous with phylogenetic expectations.

Comparing relative rates of evolution with RERConverge

Lastly, we examined how output alignments from our Lift&Add workflow can be explored for variation in evolutionary rates, using here the R package RERConverge to revisit thylacine and wolf acceleration (Partha *et al*. 2017, Kowalczyk *et al*. 2019). The wolf genome is also not present in the 60-way vertebrate WGA. Thus, as in the original analysis, we added the gray wolf [mCanLor1.2 released 2021, contig N50 34.4 Mb and scaffold N50 65.8 Mb (Sinding *et al*. 2021)] to our alignments with Lift&Add, mapping the test dataset of 17 155 conserved elements from the mm10 chromosome 19 to the dog genome (canFam3) with UCSC liftOver, followed by Liftoff of elements from the dog genome to the target gray wolf genome. Of the total 17 067 (99.9%) vertebrate conserved elements that mapped to the wolf

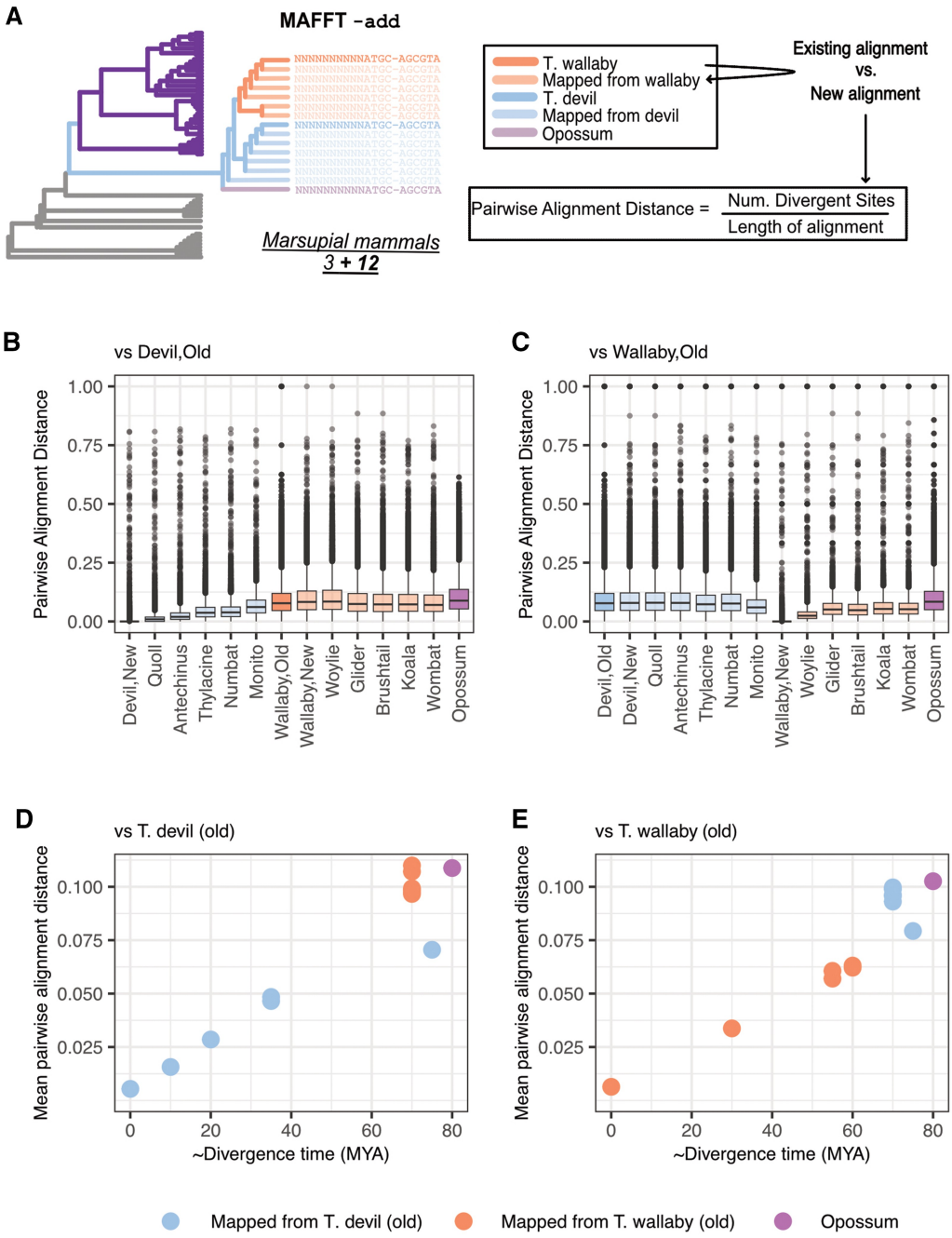

Figure 6 Alignment with MAFFT. (A) MAFFT—add was used to incorporate sequences from 12 new marsupials genomes to existing sequence alignments for vertebrate conserved elements. To QC the resulting alignments, we calculated the pairwise alignment distance between existing and added sequences. Pairwise alignment distance between (B) the Tasmanian devil and each new and old marsupial genome and (C) the tammar wallaby and each new and old marsupial genome. (D) Estimated divergence time from the Tasmanian devil versus average pairwise distance (across all elements) from the Tasmanian devil. (E) Estimated divergence time from the tammar wallaby versus average pairwise distance from the tammar wallaby.

genome, 13 200 also mapped to at least one of the 10 marsupial taxa not already present in the 60-way WGA (i.e. excluding the new Tasmanian devil and tammar wallaby genome). With MAFFT, we added conserved sequences from the wolf to the multiple sequence alignments of each of these 13 200 elements. Our final alignments contained 13 new genomes: this new chromosome-level assembly of the gray wolf and the 12 marsupial genomes detailed above.

RERConverge is specifically designed to correlate convergent phenotypic traits (binary or continuous data) with shifts in evolutionary rates (Partha *et al.* 2017, Kowalczyk *et al.* 2019). As in the original Feigin *et al.* (2019) study, we here looked at thylacine-wolf convergence as a binary trait (coded as 1 for thylacine and wolf, and 0 for all other taxa). We first estimated individual gene trees for each conserved element, constraining the species topology. Subsequently, we averaged branch lengths of these

individual gene trees to derive a single consensus tree, the branches of which described the overall evolutionary rate for this dataset (Kowalczyk *et al*. 2019) (Supplementary Fig. 11). Comparing this consensus tree to the phylogenetic model for neutral evolution derived from the original 60-way WGA (Rosenbloom *et al*. 2015), we observed positive correlation in phylogenetic distances for each pair of taxa [Spearman's ρ 0.98 (*P*-value $< 2.22 \times e^{-16}$)]. These pairwise distances are on average 0.22 times shorter than those in the neutral tree, reflecting the evolutionary constraint on these conserved genomic regions.

After estimating the consensus tree, we then calculated element-specific rates of evolution—the relative evolutionary rates (RERs), where a positive value is indicative of acceleration and a negative one of constraint—for every element in our test dataset (Kowalczyk *et al*. 2019, Redlich *et al*. 2024). This included elements overlapping (with between 47% to 100% overlap) six thylacine and wolf accelerated regions (TWARs) from Feigin *et al*. (2019) which aligned to mouse chromosome 19 (Fig. 7A–F). By using Lift&Add we were able to incorporate between 4 and 9 additional marsupial genomes to these TWAR elements. Both the thylacine and wolf had positive RERs for all six TWAR-corresponding elements. This confirms accelerated evolution of these elements in the thylacine and wolf, as was previously determined in Feigin *et al*. (2019).

However, multiple other marsupial taxa are also under accelerated evolution for these TWAR-corresponding elements. In elements equivalent to TWAR 1, 2, 3 and 4 from chromosome 19, the thylacine is not the marsupial species with highest RER (Fig. 7A–D, marsupials highlighted in purple). In fact, multiple other marsupials have positive RERs greater than or comparable to the thylacine. These results highlight the importance of densely sampling more closely related taxa in comparative genomic analyses for specific identification of lineage-specific variation. Across 8505 conserved elements aligned to at least 20 vertebrate species, we were unable to identify any elements where RERs were correlated with the thylacine-wolf binary convergence (at a FDR of 5%), including any of the six TWARs included in this re-analysis.

Discussion

The power of comparative genomics analyses to resolve evolutionary constraint and acceleration is a function of the sequence diversity that is sampled (Cooper *et al*. 2003, Boffelli *et al*. 2004, Delsuc and Tilak 2015, Hecker and Hiller 2020). While large vertebrate and mammalian alignments such as the UCSC Multiz alignments and Zoonomia 241-way alignment prioritize breadth (which provides greater power to detect conservation), in the investigation of species-specific genetic changes, deep phylogenetic sampling is valuable (Boffelli *et al*. 2004, Stone *et al*. 2005, Christmas *et al*. 2023, Raney *et al*. 2024). However, generating alignments of multiple genomes or adding a whole genome to an existing alignment is not trivial, and is made more difficult with increasing number, complexity and evolutionary distance between genomes (Saada *et al*. 2024). Here we have presented a bioinformatic approach, Lift&Add, designed for vertebrate conserved elements that (i) robustly maps them to several new genome assemblies and (ii) can add sequences of conserved elements from these new genomes back to existing alignments, bypassing the need for whole-genome alignment.

We hypothesized that Liftoff (Shumate and Salzberg 2021) could be feasibly used to map both exonic and non-coding vertebrate conserved elements between genome assemblies, considering that these vertebrate conserved elements (i) are defined by alignment and a high degree of sequence similarity across distant taxa, (ii) are on average shorter than most exons, (iii) and thus require fewer structural considerations for alignment than genes and transcripts. Success with local alignment is greatest over shorter evolutionary distances; we thus adopted a two-step approach, using an initial liftOver step to map the elements in mouse coordinates to query marsupial genomes. We note that this approach, although designed and tested on multiz alignments, can be extended to Cactus reference-free alignments with halLiftover replacing liftOver in the initial mapping step.

As expected, the percentage of mapped elements with Liftoff was negatively correlated with divergence distance from the query genome. Yet, even at the longest evolutionary distance between query and target genome (60–70 MYA, between the Tasmanian devil and monito del monte) 75% of conserved elements remained mappable. Crucially, we observed no bias for successful mapping of protein-coding or non-coding elements. Examining positional information of mapped elements between genomes [as in the example given by Shumate and Salzberg (2021)], we observed concordance with expected cytogenetic relationships between taxa.

The biggest predictor of successful mapping with Liftoff was the length of query elements. Addition of flanking sequence to short (50–200 bp) conserved elements was associated with improved mapping. The sequence aligner used by Liftoff, minimap2, can map short reads but displays better performance with longer reads and has a minimum recommended read length of 100 bp (Li 2018). This could underlie the poor performance of Liftoff for shorter elements. It can be argued that adding flanking sequence to a vertebrate conserved element introduces non-conserved bases, promoting spurious mapping. However, by comparing Liftoff and liftOver output, we confirmed that our two-step approach displays comparable accuracy with liftOver. Furthermore, the ability to align a conserved element across species decays proportional to the phylogenetic distance between them (Miller *et al*. 2007, Cummins *et al*. 2024). As such, even if genomic sequence flanking a query element was not determined to be conserved across vertebrates, it is likely to be more conserved among just the marsupial species. Nonetheless, we emphasise that in usage of this workflow, evolutionary distance between query and target genome should be kept as small as possible for maximum recovery.

To ensure robust mapping, we exclude elements that map to the target genome with sequence identity lower than 50% of the query element, as recommended by Liftoff (Shumate and Salzberg 2021). While reducing incorrect mapping, this limits investigation of sequence variation unique to the added taxa. Elements that are excessively divergent in the query or target genomes, such as those containing large deletions or insertions, will not be mapped. As such, our Lift&Add workflow is best suited to add phylogenetic context to a multiple species alignment or to investigate more subtle sequence changes in the added taxa—which will also likely be more frequent in highly constrained genomic regions. Additionally, we are also limited

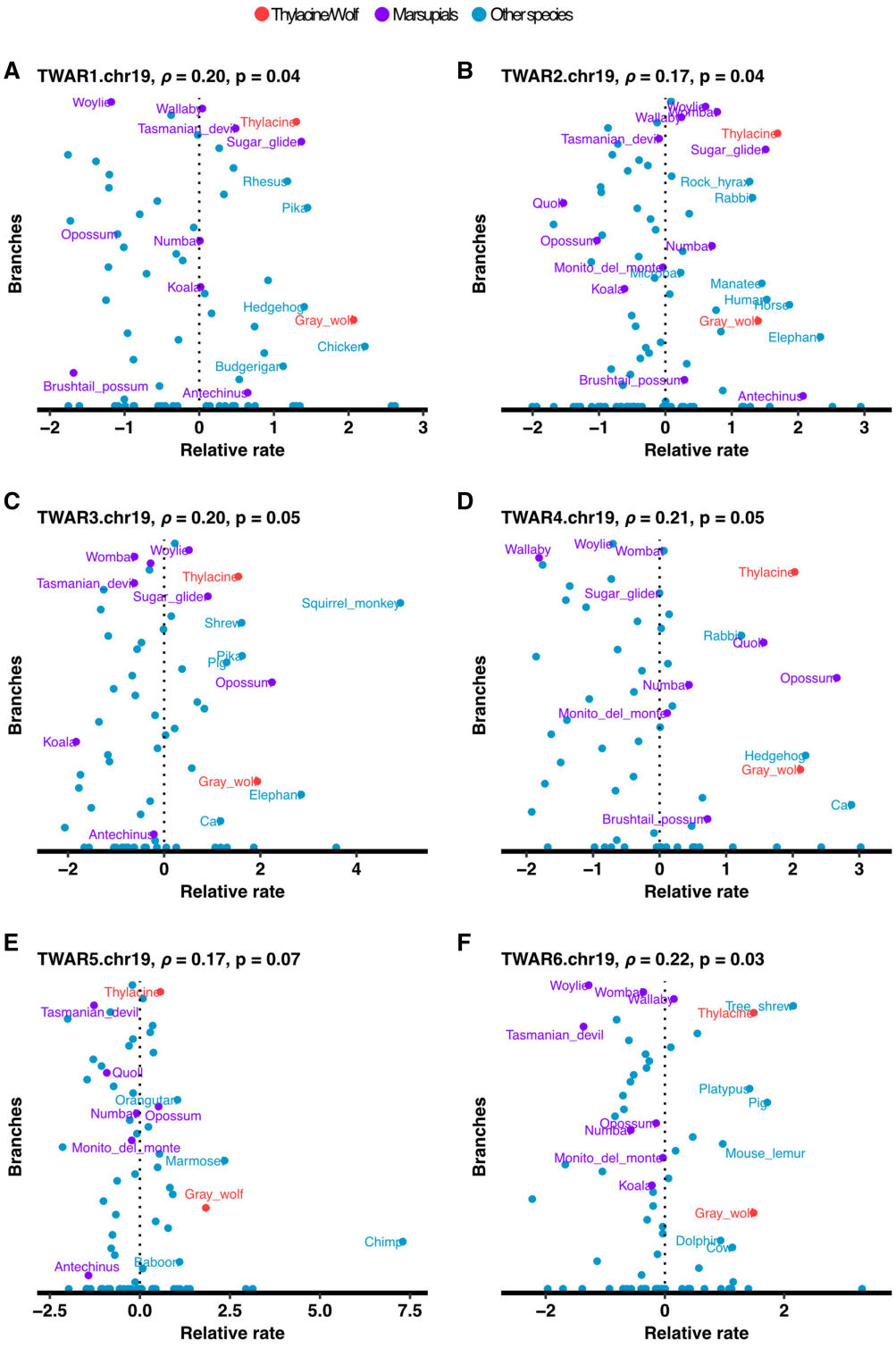

Figure 7 Relative evolutionary rates and phenotype association output (ρ and *P* values) for elements overlapping TWARs. ρ is the correlation between the relative evolutionary rate and the convergent phenotype and *P* is the nominal *P*-value for this association.

by the availability and quality of chain files, required for the initial liftOver step, generated through pairwise whole-genome alignments and thus only available for limited pairs of genome assemblies.

Lastly, we also present here an example of an investigation that can be carried out following the use of Lift&Add, using RERConverge to explore evolutionary rates in our test dataset to re-evaluate the evidence for convergent acceleration between the gray wolf and the thylacine. We reanalysed six TWARs from Feigin *et al*. (2019) within the context of additional marsupial sequences. Encouragingly, all six were accelerated both in thylacine and the wolf. Four of these, however, also had positive

relative rates in other marsupial species. Although narrow, this example analysis hints at the importance of dense sampling in the study of lineage-specific acceleration and other similar comparative genomic inferences.

Conclusion

The workflow presented here allows enhanced study of vertebrate conserved elements, facilitating computationally inexpensive incorporation of sequences from newly sequenced genomes into phylogenetic comparisons.

Acknowledgement

We thank Charlie Feigin, Laura E. Cook and members of the Gallego Romero Lab for their analysis advice.

Author contributions

Navya Shukla (Conceptualization [lead], Data curation [lead], Formal analysis [lead], Investigation [lead], Methodology [lead], Project administration [equal], Software [lead], Validation [lead], Visualization [lead], Writing—original draft [lead], Writing—review & editing [lead]) and Irene Gallego Romero (Conceptualization [supporting], Funding acquisition [lead], Project administration [equal], Resources [lead], Supervision [lead], Writing—original draft [supporting], Writing—review & editing [supporting])

Conflict of interests

The authors declare that they have no competing interests.

Funding

NS was supported by a Melbourne Research Scholarship and the David Lachlan Hay Memorial Grant from the University of Melbourne. St Vincent's Institute acknowledges the infrastructure support it receives from the National Health and Medical Research Council Independent Research Institutes Infrastructure Support Program and from the Victorian Government through its Operational Infrastructure Support Program.

Data availability

Code for Lift&Add is available at https://github.com/navyashukladr/Lift_and_Add/ and has been archived on Zenodo (https://doi.org/10.5281/zenodo.19568005). All plots and statistical testing detailed were implemented in R (version 4.2.1). Analysis scripts are available at https://gitlab.unimelb.edu.au/igr-lab/thylacine_canid_convergence_reanalysis/-/tree/master/analysis.

References

Bejerano G, Pheasant M, Makunin I *et al.* Ultraconserved elements in the human genome. *Science* 2004;**304**:1321–5.

Bi X, Zhou L, Zhang J-J *et al.* Lineage-specific accelerated sequences underlying primate evolution. *Sci Adv* 2023;**9**:eadc9507.

Boffelli D, Nobrega MA, Rubin EM. Comparative genomics at the vertebrate extremes. *Nat Rev Genet* 2004;**5**:456–65.

Booker BM, Friedrich T, Mason MK *et al.* Bat accelerated regions identify a bat forelimb specific enhancer in the HoxD locus. *PLoS Genet* 3 2016;**12**:e1005738. https://doi.org/10.1371/journal.pgen.1005738

Christmas MJ, Kaplow IM, Genereux DP, Zoonomia Consortium§ *et al.* Evolutionary constraint and innovation across hundreds of placental mammals. *Science (1979)* 2023;**380**:eabn3943.

Cooper GM, Brudno M, Green ED, NISC Comparative Sequencing Program *et al.* Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. *Genome Res* 2003;**13**:813–20.

Cummins M, Watson C, Edwards RJ *et al.* The evolution of ultraconserved elements in vertebrates. *Mol Biol Evol* 2024;**41**:msae146.

Deakin JE. Chromosome evolution in marsupials. *Genes (Basel)* 2018;**9**:72.

Delsuc F, Tilak M-K. Naked but not hairless: the pitfalls of analyses of molecular adaptation based on few genome sequence comparisons. *Genome Biol Evol* 2015;**7**:768–74.

Duchêne DA, Bragg JG, Duchêne S *et al.* Analysis of phylogenomic tree space resolves relationships among marsupial families. *Syst Biol* 2018;**67**:400–12.

Eddy SR. A model of the statistical power of comparative genome sequence analysis. *PLoS Biol* 2005;**3**:e10.

Feigin CY, Newton AH, Doronina L *et al.* Genome of the Tasmanian tiger provides insights into the evolution and demography of an extinct marsupial carnivore. *Nat Ecol Evol* 2018;**2**:182–92. ISSN 2397-334X. https://doi.org/10.1038/s41559-017-0417-y

Feigin CY, Newton AH, Pask AJ. Widespread cis-regulatory convergence between the extinct Tasmanian tiger and Gray wolf. *Genome Res* 2019;**29**:1648–58. ISSN 1549-5469. https://doi.org/10.1101/gr.244251.118

Feigin CY, Frankenberg S, Pask AJ. A chromosome-scale hybrid genome assembly of the extinct Tasmanian tiger (*Thylacinus cynocephalus*). *Genome Biol Evol* 2022;**14**:evac048.

Feng S, Stiller J, Deng Y *et al.* Dense sampling of bird diversity increases power of comparative genomics. *Nature* 2020;**587**:252–7.

Ferris E, Gregg C. Parallel accelerated evolution in distant hibernators reveals candidate cis elements and genetic circuits regulating mammalian obesity. *Cell Rep* 2019;**29**:2608–20.e4. https://doi.org/10.1016/j.celrep.2019.10.102

Goswami A, Milne N, Wroe S. Biting through constraints: cranial morphology, disparity and convergence across living and fossil carnivorous mammals. *Proc Biol Sci* 2011;**278**:1831–9. https://doi.org/10.1098/rspb.2010.2031

Hecker N, Hiller M. A genome alignment of 120 mammals highlights ultraconserved element variability and placenta-associated enhancers. *Gigascience* 2020;**9**:giz159.

Hubisz MJ, Pollard KS, Siepel A. Phast and rphast: phylogenetic analysis with space/time models. *Brief Bioinform* 2011; **12**:41–51.

Katoh K, Standley DM. Mafft multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–80.

Keough KC, Whalen S, Inoue F, Zoonomia Consortium§ *et al.* Three-dimensional genome rewiring in loci with human accelerated regions. *Science (1979)* 2023;**380**:eabm1696.

Kille B, Balaji A, Sedlazeck FJ *et al.* Multiple genome alignment in the telomere-to-telomere assembly era. *Genome Biol* 2022; **23**:182.

Kim EB, Fang X, Fushan AA *et al.* Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* 2011;**479**:223–7.

Kowalczyk A, Meyer WK, Partha R *et al.* RERconverge: an r package for associating evolutionary rates with convergent traits. *Bioinformatics* 2019;**35**:4815–7. https://doi.org/10.1093/bioinformatics/btz468

Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**34**:3094–100.

Lindblad-Toh K, Garber M, Zuk O, Genome Institute at Washington University *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 2011; **478**:476–82.

Mikkelsen TS, Wakefield MJ, Aken B, Broad Institute Whole Genome Assembly Team *et al.* Genome of the marsupial monodelphis domestica reveals innovation in non-coding sequences. *Nature* 2007;**447**:167–77.

Miller W, Rosenbloom K, Hardison RC *et al.* 28-way vertebrate alignment and conservation track in the UCSC genome browser. *Genome Res* 2007;**17**:1797–808.

Murchison EP, Adams DJ. Sequencing skippy: the genome sequence of an Australian kangaroo, *Macropus eugenii*. *Genome Biol* 2011;**12**:123.

Neph S, Kuehn MS, Reynolds AP *et al.* Bedops: high-performance genomic feature operations. *Bioinformatics* 2012;**28**:1919–20.

Nishimura O, Hara Y, Kuraku S. Gvolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics* 2017;**33**:3635–7.

O'Neill R, Eldridge M, Metcalfe C. Centromere dynamics and chromosome evolution in marsupials. *J Hered* 2004;**95**: 375–81.

Pal A, Noble MA, Morales M *et al.* Resolving the three-dimensional interactome of human accelerated regions during human and chimpanzee neurodevelopment. *Cell* 2025; **188**:1504–23.e27.

Partha R, Chauhan BK, Ferreira Z *et al.* Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *eLife* 2017;**6**:e25884. https://doi.org/10.7554/eLife.25884

Pollard KS, Salama SR, King B *et al.* Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* 2006;**2**: e168. https://doi.org/10.1371/journal.pgen.0020168

Prabhakar S, Visel A, Akiyama JA *et al.* Human-specific gain of function in a developmental enhancer. *Science* 2008;**321**: 1346–50.

Quinlan AR, Hall IM. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841–2.

Raney BJ, Barber GP, Benet-Pagès A *et al.* The UCSC genome browser database: 2024 update. *Nucleic Acids Res* 2024; **52**:D1082–D1088.

Redlich R, Kowalczyk A, Tene M *et al.* Rerconverge expansion: using relative evolutionary rates to study complex categorical trait evolution. *Mol Biol Evol* 2024;**41**:msae210.

Renfree MB, Papenfuss AT, Deakin JE *et al.* Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biol* 2011;**12**:R81.

Rens W, O'Brien PC, Yang F *et al.* Karyotype relationships between four distantly related marsupials revealed by reciprocal chromosome painting. *Chromosome Res* 1999;**7**:461–74.

Rhie A, McCarthy SA, Fedrigo O *et al.* Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 2021;**592**:737–46.

Rosenbloom KR, Armstrong J, Barber GP *et al.* The UCSC genome browser database: 2015 update. *Nucleic Acids Res* 2015; **43**:D670–D681.

Saada B, Zhang T, Siga E *et al.* Whole-genome alignment: methods, challenges, and future directions. *Appl Sci* 2024;**14**:4837.

Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics* 2021;**37**:1639–43.

Siepel A, Bejerano G, Pedersen JS *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005;**15**:1034–50.

Sinding M-HS, Gopalakrishnan S, Raundrup K, D. T. of Life Consortium W. S. I. T. of Life *et al.* The genome sequence of the grey wolf, *Canis lupus* linnaeus 1758. *Wellcome Open Res* 2021;**6**:310.

Stone EA, Cooper GM, Sidow A. Trade-offs in detecting evolutionarily constrained sequence by comparative genomics. *Annu Rev Genomics Hum Genet* 2005;**6**:143–64.

Thomas GW, Hahn MW, Hahn Y. The effects of increasing the number of taxa on inferences of molecular convergence. *Genome Biol Evol* 2017;**9**:213–21.

Uchiyama I, Higuchi T, Kobayashi I. Cgat: a comparative genome analysis tool for visualizing alignments in the analysis of complex evolutionary changes between closely related genomes. *BMC Bioinformatics* 2006;**7**:472–17.

Visel A, Prabhakar S, Akiyama JA *et al.* Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet* 2008;**40**:158–60.

Woolfe A, Goodson M, Goode DK *et al.* Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* 2005;**3**:e7.