

# Platypus and echidna genomes reveal mammalian biology and evolution

<https://doi.org/10.1038/s41586-020-03039-0>

Received: 4 December 2019

Accepted: 30 July 2020

Published online: 6 January 2021

Open access

 Check for updates

Yang Zhou<sup>1,2,36</sup>, Linda Shearwin-Whyatt<sup>3,36</sup>, Jing Li<sup>4,36</sup>, Zhenzhen Song<sup>1,5</sup>, Takashi Hayakawa<sup>6,7</sup>, David Stevens<sup>3</sup>, Jane C. Fenelon<sup>8</sup>, Emma Peel<sup>9</sup>, Yuanyuan Cheng<sup>9</sup>, Filip Pajpach<sup>3</sup>, Natasha Bradley<sup>3</sup>, Hikoyu Suzuki<sup>10</sup>, Masato Nikaido<sup>11</sup>, Joana Damas<sup>12</sup>, Tasman Daish<sup>3</sup>, Tahlia Perry<sup>3</sup>, Zexian Zhu<sup>4</sup>, Yuncong Geng<sup>13</sup>, Arang Rhie<sup>14</sup>, Ying Sims<sup>15</sup>, Jonathan Wood<sup>15</sup>, Bettina Haase<sup>16</sup>, Jacquelyn Mountcastle<sup>16</sup>, Olivier Fedrigo<sup>16</sup>, Qiye Li<sup>1</sup>, Huanming Yang<sup>1,17,18,19</sup>, Jian Wang<sup>1,17</sup>, Stephen D. Johnston<sup>20</sup>, Adam M. Phillippy<sup>14</sup>, Kerstin Howe<sup>15</sup>, Erich D. Jarvis<sup>21,22</sup>, Oliver A. Ryder<sup>23</sup>, Henrik Kaessmann<sup>24</sup>, Peter Donnelly<sup>25</sup>, Jonas Korlach<sup>26</sup>, Harris A. Lewin<sup>12,27,28</sup>, Jennifer Graves<sup>29,30,31</sup>, Katherine Belov<sup>9</sup>, Marilyn B. Renfree<sup>9</sup>, Frank Grutzner<sup>3,32</sup>, Qi Zhou<sup>4,32,33</sup> & Guojie Zhang<sup>1,2,34,35</sup>✉

Egg-laying mammals (monotremes) are the only extant mammalian outgroup to therians (marsupial and eutherian animals) and provide key insights into mammalian evolution<sup>1,2</sup>. Here we generate and analyse reference genomes of the platypus (*Ornithorhynchus anatinus*) and echidna (*Tachyglossus aculeatus*), which represent the only two extant monotreme lineages. The nearly complete platypus genome assembly has anchored almost the entire genome onto chromosomes, markedly improving the genome continuity and gene annotation. Together with our echidna sequence, the genomes of the two species allow us to detect the ancestral and lineage-specific genomic changes that shape both monotreme and mammalian evolution. We provide evidence that the monotreme sex chromosome complex originated from an ancestral chromosome ring configuration. The formation of such a unique chromosome complex may have been facilitated by the unusually extensive interactions between the multi-X and multi-Y chromosomes that are shared by the autosomal homologues in humans. Further comparative genomic analyses unravel marked differences between monotremes and therians in haptoglobin genes, lactation genes and chemosensory receptor genes for smell and taste that underlie the ecological adaptation of monotremes.

The iconic egg-laying monotremes of Australasia represent one of the three major mammalian lineages. The monotreme lineage comprises two extant families, the semi-aquatic Ornithorhynchidae (platypus) and the terrestrial Tachyglossidae (echidna). At present, the single species of platypus has a restricted distribution in Eastern Australia, whereas four echidna species (*T. aculeatus* and three *Zaglossus* spp.) are present in Australia and New Guinea (Supplementary Information). Platypuses and echidnas feature radical differences in diet (carnivorous compared with insectivorous), neurophysiology (electroreception-oriented compared with olfaction-oriented), as well as specific intraspecific conflict and defence adaptations<sup>1</sup>. Owing to their distinct ecological, anatomical and physiological features, monotremes are interesting mammals well-suited for the study of the evolution of ecological adaptation. Of particular interest are their sex chromosomes, which originated independently from those of therian mammals through additions of autosomes onto an ancestral XY pair, resulting in a multiple sex chromosome system that assembles as a chain during meiosis<sup>3</sup>.

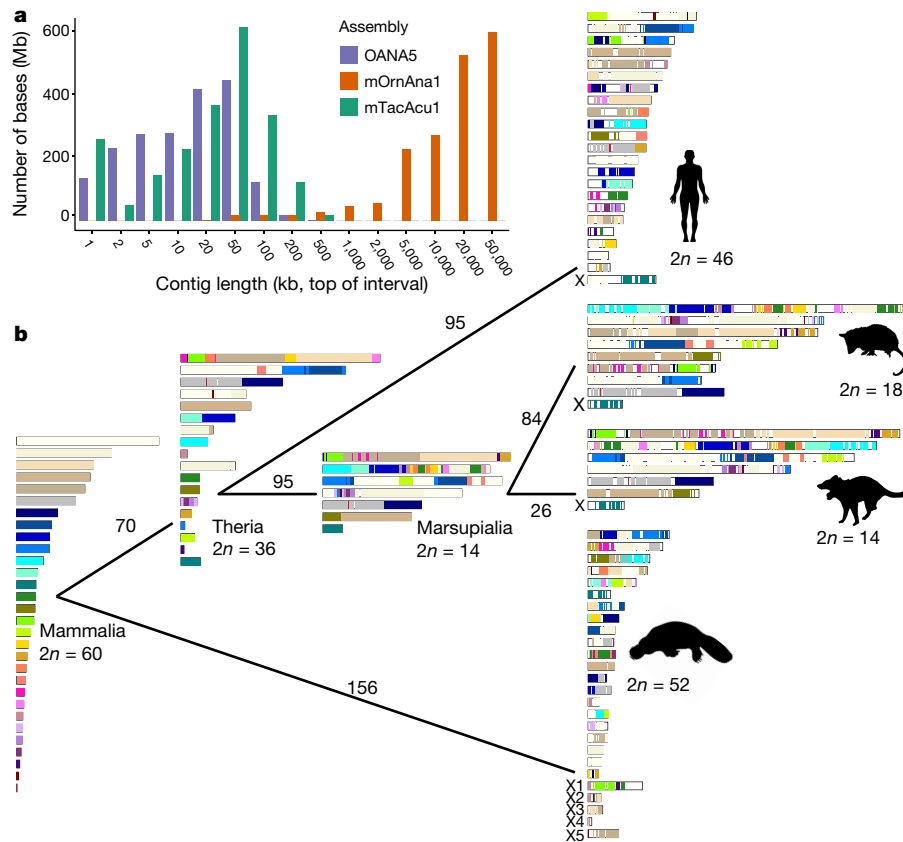
The previous female platypus genome assembly (OANA5) provided many important insights into monotreme biology and mammalian evolution. However, only about 25% of its sequence was assigned to

chromosomes<sup>2</sup>. The incomplete platypus assembly without Y chromosome sequences and lack of an echidna genome have limited the interpretation of the evolution of mammals and monotremes. Here we combined PacBio long-read, 10× linked-read, chromatin conformation (Hi-C) and physical map data to produce a highly accurate chromosome-scale assembly of the platypus genome. We also produced a less-continuous assembly for the short-beaked echidna, which enables us to infer the genomic changes that occurred in the ancestral monotremes and other mammals.

## Chromosome-scale monotreme genomes

Our new male platypus genome assembly (mOrnAna1) shows a 1,390-fold improvement for the contig N50 and a 49-fold improvement for the scaffold N50 compared with the previous Sanger-based assembly (OANA5) (Fig. 1a). We performed extensive error correction and manual curation to polish and anchor the assembly at the chromosome scale (Extended Data Fig. 1a, b). Ambiguous chromosome assignments were resolved with fluorescence in situ hybridization (FISH) experiments (Extended Data Fig. 1c, d). We also produced a male echidna

A list of affiliations appears at the end of the paper.



**Fig. 1 | Chromosome assembly of monotreme and mammalian genome evolution. a**, The contig length distribution among the three monotreme assemblies shows a large improvement in the sequence continuity of the platypus assembly, and at least equivalent quality of the echidna assembly. **b**, Mammalian karyotype evolution trajectory.  $2n = 60$  ancestral karyotypes were inferred for the common ancestor of mammals. Conserved blocks were

colour-coded in accordance with their chromosomal source in the mammalian ancestor. Numbers of estimated rearrangements are shown for each branch. Silhouettes of the human and opossum are from <https://www.flaticon.com/>. Silhouettes of the platypus and Tasmanian devil are created by S. Werning and are reproduced under the Creative Commons Attribution 3.0 Unported licence (<http://creativecommons.org/licenses/by/3.0/>).

genome (mTacAcu1) from a variety of short- and long-insert-size libraries, and further scaffolded it using the same methods as in platypus. The resulting mTacAcu1 sequence shows better sequence continuity than OANA5, with a scaffold N50 size of 32.51 Mb (Supplementary Table 2).

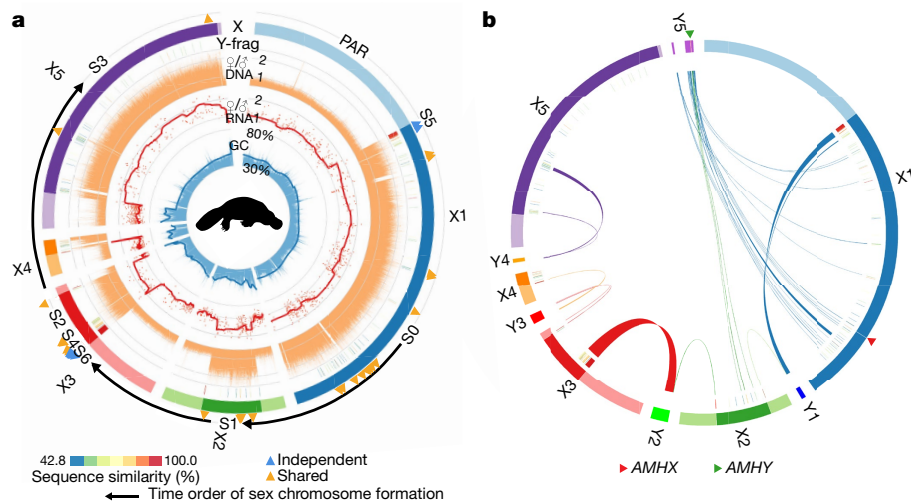
To study the origin and evolution of monotreme sex chromosomes, we greatly improved the assembly of the platypus sex chromosomes. We anchored 172 Mb (92% compared to 22% in OANA5) X-borne sequences to chromosomes (Supplementary Tables 4, 6). This includes one 1.6-Mb segment that was previously misassigned to chromosome 14 (Extended Data Fig. 1e). We determined all of the pseudoautosomal regions (PARs) except for X4, on the basis of the different read coverage between sexes and representation of FISH markers (Supplementary Table 3). We also mapped 92% of the platypus Y-borne sequences to the five Y chromosomes using PacBio reads produced using Y-borne bacterial artificial chromosome (BAC) clones<sup>4</sup> (Supplementary Tables 5, 6). Owing to a lack of echidna linkage markers, we used the platypus X chromosomes as a reference to anchor a similar length (177 Mb, 96%) of X chromosomes and identified 8.6 Mb Y-borne sequences in echidna.

In the final curated platypus genome (mOrnAna1) 98% of the sequence was assigned to the 21 autosomes, 5 X and 5 Y chromosomes (Supplementary Table 7), with putative telomeres and centromeres annotated for half of the chromosomes (Supplementary Table 8). mOrnAna1 fills around 90% of the gaps in OANA5 (Supplementary Table 9), recovering 161 Mb of previously missed genomic sequences, most of which are long interspersed nuclear elements (LINE)/L2 and short interspersed nuclear elements (SINE)/MIR (Supplementary Tables 10, 11). We also removed 68 Mb of redundant sequences in OANA5 (Extended Data

Fig. 1f–h). The repeat elements comprising about half of the monotreme genomes are dominated by LINE/L2 elements that are more similar to reptile genomes than therian mammals (which comprise mostly LINE/L1)<sup>2</sup> (Supplementary Table 12). The highly continuous assembly also substantially improves gene annotation. We identified 20,742 and 22,029 protein-coding genes in mOrnAna1 and mTacAcu1, respectively (Supplementary Table 13). Specifically, 19,576 coding exons from 8,303 platypus genes were recovered from the gapped regions of OANA5. Among them, 454 genes were completely missed in OANA5, and 3,961 fragmented genes in OANA5 now have complete open-reading frames. We corrected 2,395 genes that were previously split or misannotated in OANA5 (Extended Data Fig. 1i, j).

### Insights into mammalian genome evolution

Our phylogenomic reconstruction shows that monotremes diverged from therians around 187 million years ago, and the two monotremes diverged around 55 million years ago (Extended Data Fig. 2a). This estimate provides a date for the monotreme–therian split that is earlier than previous estimates (about 21 million years ago)<sup>2</sup>, but agrees with recent analyses of few genes and fossil evidence<sup>6</sup>. We also inferred that monotremes had similar genome substitution rates (approximately  $2.6 \times 10^{-3}$  substitutions per site per million years) compared with other mammals (Supplementary Table 15). About 14 Mb of mammalian specific highly conserved elements were identified by comparison among vertebrates (Methods): around 90% of elements were located in non-coding regions (Extended Data Fig. 2c), and are associated with genes that are enriched



**Fig. 2 | Origin and evolution of the sex chromosomes of the platypus.**

**a**, Genomic composition of the platypus sex chromosomes. From the outer to inner rings: the X chromosomes with the PARs (light colours) and SDRs (dark colours) labelled; the assembled Y chromosome fragments within SDRs showing the colour-scaled sequence divergence levels with the homologous X chromosomes; female-to-male (F/M) ratios of short sequencing-read coverage in non-overlapping 5-kb windows; F/M expression ratios (each red dot is one gene) of the adult kidney and the smoothed expression trend; and GC content in non-overlapping 2-kb windows. In addition, we labelled the positions on the

X chromosome ring of the gametologue pairs that have suppressed recombination before the divergence of monotremes ('shared', orange triangles) or after the divergence ('independent', blue triangles). **b**, Homology between X and Y chromosomes of platypus. In particular, most of Y5 shows homology with X1 and X2, which suggests an ancestral ring conformation of the platypus sex chromosomes. We also labelled the position of the putative sex-determining gene *AMH*. The platypus silhouette is created by S. Werning and is reproduced under the Creative Commons Attribution 3.0 Unported licence (<http://creativecommons.org/licenses/by/3.0/>).

in processes such as brain development (Extended Data Fig. 2d, e, Supplementary Results and Supplementary Tables 18–20).

Next we used chromosome information from human, opossum, Tasmanian devil, platypus, chicken and common wall lizard genomes to reconstruct the mammalian ancestral karyotype (Methods). This analysis reveals 30 mammalian ancestral chromosomes (MACs) ( $2n = 60$ ) at a resolution of 500 kb, covering around 66% of the human genome and approximately 67% of the platypus genome (Fig. 1b and Supplementary Tables 24–26). Of these, 25 MACs were maintained without breaks in a single chromosome of the therian ancestor, and 17 of them have fused with other MACs in therians. Sixteen MACs were still maintained in a single human chromosome, but only MAC28 had not undergone any intrachromosomal rearrangements during therian evolution (Extended Data Fig. 2f, g). We detected at least 918 chromosome breakage events, and confirmed that the X chromosome in humans was derived from the fusion of an original therian X chromosome with an autosomal region after the divergence from marsupials<sup>7</sup> (Fig. 1b and Extended Data Fig. 2f, g). The five X chromosomes in platypus were derived from different MACs by multiple fusion and translocation events.

We found that gene families associated with the immune response and hair growth were expanded considerably in the mammalian ancestor, perhaps contributing to the evolution of immune adaptation and fur, respectively, in mammals (Supplementary Table 30). We further manually annotated major histocompatibility complex (MHC) genes and other immune genes (Supplementary Results). As in nonmammalian vertebrates, the monotreme MHC class Ia genes colocalize with antigen-processing genes and MHC class II genes (Extended Data Fig. 3a and Supplementary Table 31). The defensin genes gave rise to unique defensin-like peptides (*OavDLP* genes) in platypus venom<sup>8</sup>. By contrast, echidna has only one single *OavDLP* pseudogene (Extended Data Fig. 3f–h), suggesting the loss of the key venom gene family in this species.

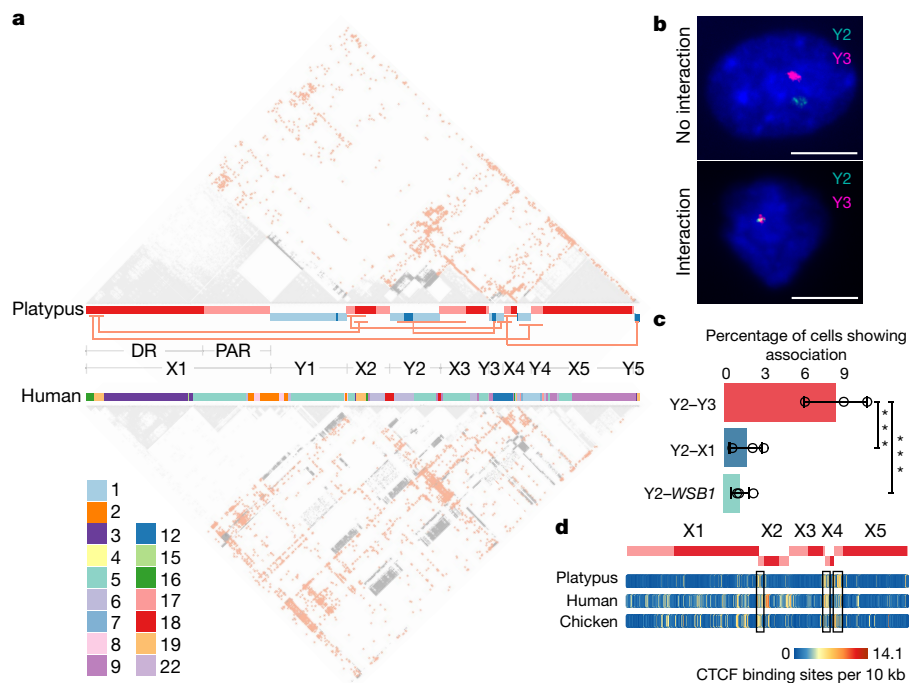
### Monotreme sex chromosome evolution

To elucidate the detailed genomic composition of the monotreme sex chromosomes, we compared regions that share sequences between the

sex chromosomes—that is, the PARs—with regions that have become sexually differentiated (SDRs). PAR boundaries show a sharp shift in the female-to-male sequencing coverage ratio as expected (Fig. 2a and Extended Data Fig. 4a). Both monotremes showed generally nonbiased gene expression levels between sexes within PARs, but pronounced female-biased expression within SDRs, indicating the absence of complete chromosome-wide dosage compensation in monotremes as previously suggested<sup>9</sup> (Extended Data Fig. 4b).

The short PARs of platypus chromosomes X2–X5 have a significantly higher GC content (one-sided Wilcoxon rank-sum test,  $P < 0.01$ ) than the SDRs or the longer PARs (Extended Data Fig. 4c), which probably reflects strong GC-biased gene conversion that is caused by a high recombination rate<sup>10</sup>. This is similar to the pattern of the short GC-rich human PAR, the recombination rate of which is 17-fold higher than the genome-wide average<sup>11</sup>. Notably, chicken orthologous sequences of these monotreme PARs are all located on the microchromosomes, which also have a high GC content<sup>12</sup> (one-sided Wilcoxon rank-sum test,  $P < 0.01$ ) (Extended Data Fig. 4c, d). This highly conserved recombination landscape might be partially selected in monotremes for maintaining the sequence polymorphism and balanced dosage of MHC genes, which reside in the PARs of the chromosome X3–Y3 and Y4–X5 pairs in platypus<sup>13</sup> (Extended Data Fig. 3a). The regional selection for high recombination may also counteract further expansion of SDRs on these sex chromosomes.

Sex chromosomes of both eutherians and birds formed through stepwise suppression of recombination, resulting in a pattern of pairwise sequence divergence between SDRs termed 'evolutionary strata'<sup>14,15</sup>. We identified at least seven strata in monotremes, named S0 to S6 from the oldest to the youngest strata (Fig. 2a and Extended Data Fig. 4a), by ranking their levels of pairwise synonymous sequence divergence between the X–Y gametologues and the phylogeny (Extended Data Fig. 5a, b). All but the most recent strata (S5 and S6) are shared by platypus and echidna. However, the PARs that border S5 and S6, as well as the shorter PARs of chromosomes X2 and X5 (Extended Data Fig. 5c, d), formed independently after their divergence. Overall, the distribution of evolutionary strata suggested a time order of incorporating different



**Fig. 3 | Interactions between the platypus sex chromosomes.**

**a**, Interchromosomal interactions among the platypus sex chromosomes detected by Hi-C data of liver tissue in platypus (top) and human (bottom). The bars between the Hi-C panels show the platypus sex chromosomes and their orthologues in the human genome. Grey, intrachromosomal interactions; red, interchromosomal interactions. Red lines link the regions with significantly high interchromosomal interactions. The interchromosomal interactions seem to be conserved in mammals, as indicated by the homologous chromosomal fragments of the human and platypus sex chromosomes and their Hi-C contact patterns. **b**, FISH with BAC probes to detect sex chromosomes Y2, Y3 or X1 and autosome chromosome 17 (*WSBI*) in interphase platypus fibroblasts. Examples show no interaction between chromosomes Y2

and Y3 (top,  $n = 593$ , 3 independent experiments) and interaction (bottom,  $n = 56$ , 3 independent experiments). Scale bars, 10  $\mu\text{m}$ . **c**, The significantly higher frequency of interaction between Y2 and Y3 than that between Y2 and X1, and between Y2 and *WSBI* (chromosome 17).  $n = 185, 206, 258$  cells for the three independent replicate experiments of Y2–Y3,  $n = 258, 250, 205$  cells for the three independent replicate experiments of Y2–X1,  $n = 298, 262, 220$  cells for the three independent replicate experiments of Y2–*WSBI*. Data are mean  $\pm$  s.d. \*\*\* $P < 0.001$  (Y2–Y3 versus Y2–X1,  $P = 0.0004675$ ; Y2–Y3 versus Y2–*WSBI*,  $P = 6.376 \times 10^{-5}$ ), one-sided Fisher's exact test. **d**, Putative CTCF-binding-site density plot showing its enrichment among homologous regions in the platypus, human and chicken genomes.

ancestral autosomes into the sex chromosome chain: it started from the S0 region of X1 containing a sex-determining gene (see below), followed by X2, X3 and X5. X4 and individual regions of X3 and X1 underwent suppression of recombination after the monotreme divergence.

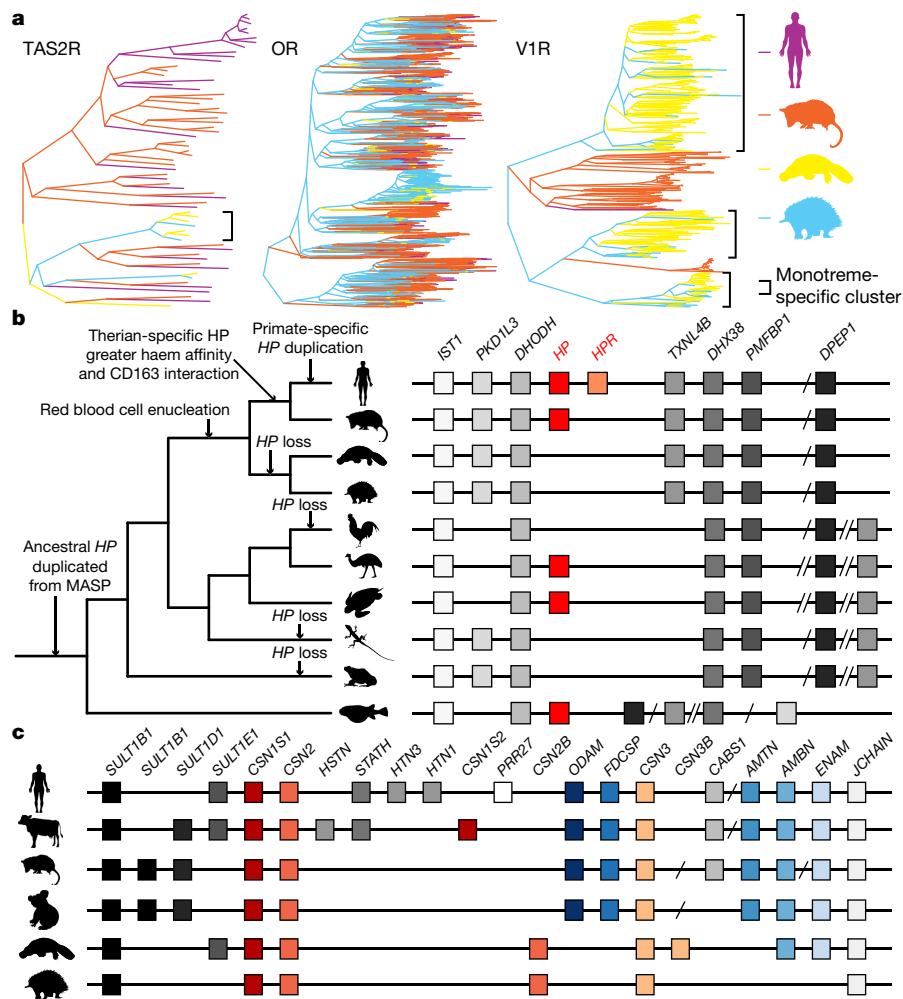
Despite episodes of independent evolution, most sex chromosome regions of the platypus and echidna are homologous (Extended Data Fig. 6a), suggesting that the complex formed in the monotreme ancestor<sup>16</sup>. To reconstruct its origin, we projected the platypus sex chromosomes onto their chicken homologues (Supplementary Table 39). This refined homology map (Extended Data Fig. 4d) suggests that both fusions and reciprocal translocations among the ancestral micro- and macrochromosomal fragments gave rise to the monotreme sex chromosome complex. The platypus X chromosomes contain homologous sequences of the entire or partial chicken microchromosomes 11, 16, 17, 25 and 28. These microchromosomes also have orthologues in the spotted gar<sup>17</sup>, suggesting that they were ancestral vertebrate microchromosomes, and fused in the ancestral monotreme or mammalian chromosomes. Evidence of reciprocal translocations came from the observation that parts of every two neighbouring sex chromosomes are homologous to two adjacent regions of the same chicken chromosome (Extended Data Fig. 6c, d). For example, platypus chromosomes X1 and X2 are both homologous to parts of chicken microchromosome 12 and chromosome 13, whereas X2 and X3 are both homologous to chicken chromosome 2.

Notably, X1 at one end of the meiotic chain and Y5 at the other share this alternately overlapping relationship, and both are homologous to chicken microchromosome 28. Indeed, most of the genes on Y5

are not found on its pairing partner X5, but on X1 (Fig. 2b and Supplementary Table 40). Chromosomes X1 and Y5 do not pair at meiosis, but this homology suggests that the origin of the extant monotreme sex chromosome complex involved the opening of the ancestral chromosomal ‘ring’ as degeneration proceeded<sup>18</sup>. A conserved vertebrate sex-determining gene, the anti-Müllerian hormone, is located on chromosome Y5 (*AMHY*) and S0 of chromosome X1 (*AMHX*)<sup>14</sup> (Fig. 2b). The ancestral X1–Y5 pairing region that encompasses *AMH* could, therefore, be the site at which homologous recombination was first suppressed. The degeneration of chromosome Y5 then caused the loss of homology with X1 and led to the break of the chromosome ring. Indeed, synonymous substitution rates (dS) between the retained X1–Y5 gametologue pairs are significantly higher (one-sided Wilcoxon ranked-sum test,  $P < 0.01$ ) than those of any other sex chromosome pairs (Extended Data Fig. 6e). A chromosome ring configuration has been reported in plants<sup>19</sup>, but not in any animal species. Alternatively, the ancestral ring structure might have evolved after the emergence of the proto-X1–Y5 pair by translocations that involve other autosomes, so that sexually antagonistic alleles could be linked to the sex-determining genes<sup>20</sup>.

### Interactions between sex chromosomes

The platypus sex chromosomes exhibit an unusual association with each other compared to autosomes during and after meiosis<sup>21</sup>. As little is known about their spatial organization in platypus somatic cells, we investigated this using Hi-C data (male liver) and chromosomal FISH with sex-chromosome-specific and autosomal BAC probes (male



**Fig. 4 | Genomic features related to biological characteristics of the monotremes.** **a**, Differences in numbers of TAS2R, OR and V1R genes between platypus and echidna. **b**, Phylogeny and synteny of the *HP* gene. Regions are not drawn to scale. **c**, Synteny conservation of the region surrounding caseins (*CSN* genes) and the ancestral teeth genes (*ODAM*, *FDCSP*, *AMTN*, *AMBN* and *ENAM*). Silhouettes of the human, opossum, koala and frog are from <https://www.flaticon.com/>. Silhouettes of the platypus and Tasmanian devil are created by S. Werning and the emu silhouette is created by D. Naish (vectorized by T. M. Keeseey); all three silhouettes are reproduced under the Creative Commons Attribution 3.0 Unported licence (<http://creativecommons.org/licenses/by/3.0/>).

fibroblasts). Notably, Hi-C data showed that chromosomes Y2 and Y3 undergo frequent interchromosomal interactions, whereas autosomes confine their interactions mostly within chromosomes (Fig. 3a and Extended Data Fig. 7a–d). FISH showed that chromosomes Y2 and Y3 signals overlapped more frequently (5.2- and 7.6-fold) than signals between chromosomes Y2 and X1 or Y2 and an autosome (chromosome 17) ( $P = 8.67 \times 10^{-4}$  and  $8.57 \times 10^{-5}$ , respectively) (Fig. 3b, c and Supplementary Table 41). These interactions allow us to predict a zigzag three-dimensional conformation of the sex chromosomes at interphase (Extended Data Fig. 7e). A similar pattern was also present in echidna (Extended Data Fig. 7f). Notably, the high interaction frequency is conserved in human orthologous autosomal regions (Fig. 3a), suggesting functional importance unrelated to the evolution or function of sex chromosomes.

We further examined the distribution of putative binding sites of the CTCF protein, which is usually enriched at the boundaries of topologically associated domains (TADs) and mediates both intra- and interchromosomal interactions<sup>22</sup>. This revealed considerable enrichment of putative CTCF-binding sites at the TAD boundaries of the platypus genome (Extended Data Fig. 7g), which are more enriched along the interacting sex chromosomes X2 and X4, as well as along their orthologous regions in human and chicken (Fig. 3d and Extended

Data Fig. 7h). These results suggest that an ancestral interaction landscape facilitated by local enrichment of CTCF-binding sites could have promoted the reciprocal translocations between spatially adjacent autosomal fragments that gave rise to the sex chromosome complex in the monotremes.

### Eco-evolutionary adaptation of diet

Platypuses consume aquatic invertebrates whereas echidnas feed predominantly on social insects. Although the recent ancestor of monotremes had adult teeth, both extant monotremes lack teeth<sup>23</sup>. Of eight genes involved in tooth development<sup>24</sup>, four genes were lost in both monotreme genomes, suggesting that the loss occurred in their recent common ancestors (Extended Data Fig. 8a and Supplementary Table 42), consistent with other toothless or enamel-less eutherians<sup>25</sup>. Echidnas (but not platypuses) further lost two enamel genes. Analysis of genes involved in stomach function revealed that the considerable loss of digestive genes (reported in platypus<sup>26</sup>) is shared with echidna and probably occurred in the monotreme ancestor, although *NGN3*—which is essential for stomach and pancreas development—has been maintained in both species (Extended Data Fig. 8b–g and Supplementary Table 43).

Chemosensory systems mediate animal behaviour that is essential for survival and reproduction through the direct interaction with environmental chemical cues<sup>27</sup>. For example, eutherian mammals have more than 25 copies of bitter taste receptor genes (*TAS2R* genes)<sup>27,28</sup>, whereas this gene family is considerably smaller in monotremes (Extended Data Fig. 9a) with only 7 in platypus (Supplementary Tables 44, 45). The number is reduced to three in echidna (Fig. 4a and Supplementary Results). This reduction is also observed in pangolins, which suggests convergent evolution that results from the insectivore diet of both echidnas and pangolins<sup>29</sup>.

The nasal cavity of the platypus is closed off during diving and the size of the main olfactory bulb of the platypus is much smaller than that of the echidna<sup>1</sup>. Consistent with this, the number of olfactory receptors (OR genes) in platypus (299) is much smaller than in echidna (693) (Fig. 4a and Supplementary Table 46). The difference in the large olfactory bulb and OR repertoire in echidna may contribute to the ability to search for odours of underground prey, whereas the platypus relies on electroreception to detect prey in the water. However, the size of the accessory olfactory bulb is larger in the platypus than in the echidna<sup>1</sup>. The accessory olfactory bulb receives projections from the vomeronasal organ, and there is a marked expansion of the number of vomeronasal type-1 receptors (V1R genes) in the platypus (262) compared with the echidna (28) (Fig. 4a and Supplementary Table 47). Vomeronasal receptors probably have important roles in courtship, parental care, induction of lactation and milk ejection in monotremes<sup>23</sup>. Therefore, the diversification of the olfactory bulb and accessory olfactory bulb systems in monotremes provide an interesting example of the eco-evolutionary trade-off. V1R amplification has been associated with the size of the vomeronasal organ and nocturnal activity<sup>30</sup>. This is also consistent with the fact that the platypus closes its eyes when diving and therefore relies entirely on other senses underwater and in the burrow.

## Haemoglobin degradation in monotremes

The semi-aquatic lifestyle of the platypus is supported by particularly high haemoglobin levels and large numbers of small red blood cells<sup>31</sup>. The haemoglobin–haem detoxification system in mammals provides efficient clearance to minimize oxidative damage<sup>32</sup> in which haptoglobin is the haemoglobin chaperone<sup>32</sup> and free haem is bound by haemopexin and alpha-1 microglobulin<sup>33</sup>.

Both the haemopexin and alpha-1 microglobulin genes are found in the monotreme genomes, whereas the haptoglobin gene is absent (Fig. 4b, Extended Data Fig. 10a, b and Supplementary Table 48), which suggests that monotremes evolved a haemoglobin clearance system that is different from that of other mammals. Haptoglobin evolved in the common ancestor of vertebrates from an immune gene of the MASP family<sup>33</sup> but has neofunctionalized in mammals to bind to haemoglobin with a higher affinity and to bind to the CD163A receptor, which is also absent in monotremes, for clearance in macrophages<sup>34</sup>. The absence of the haptoglobin gene and *CD163A* in monotremes suggests that the neofunctionalization of haptoglobin happened after the divergence of monotremes from therians, not before it as previously thought<sup>34</sup>, and long after the evolution of enucleated red blood cells in the common ancestor of mammals<sup>35</sup>. Several nonmammalian vertebrates have lost haptoglobin, including chicken<sup>34</sup> (Fig. 4b), in which an alternative, secreted CD163 family member, *PITS4*, is the haemoglobin-binding chaperone<sup>33</sup>. Phylogenetic analysis shows that monotremes lack genes that cluster with haptoglobin in the MASP family or a *PITS4* orthologue (Extended Data Fig. 10c–e and Supplementary Table 50). We confirmed the expansion of the CD163 family in platypus<sup>2</sup> (ten members) and found five in echidna, compared with two and three in humans and mice, respectively (Extended Data Fig. 10e, f). As mammalian CD163A can bind to haemoglobin in the absence of haptoglobin<sup>36</sup> and one CD163 family member has become

the haemoglobin chaperone in chicken, the CD163 family protein(s) may have evolved this role in monotremes.

## Transition from oviparity to viviparity

Monotremes provide the key to understanding how viviparity evolved in mammals. They are not as dependent on egg proteins as egg-laying avian and reptilian species owing to their nutrient acquisition from uterine secretions<sup>23,37</sup>, and the subsequent reliance of the young on lactation. Whereas reptiles have three functional copies of the major egg protein vitellogenin (*VTG*)<sup>38</sup>, in monotremes we found only one functional copy (*VTG2*) (Extended Data Fig. 10g and Supplementary Table 52) and a partial sequence for *VTG1*.

Similar to marsupials, monotremes have an extended lactation period and the composition of the milk changes dynamically as the development progresses to match the changing needs of the young<sup>37</sup>. *SPINT3*, a major milk-specific protein that is present in early lactation of therians with a probable role in the protection of immunoincompetent young in marsupials<sup>39</sup>, is absent in monotremes. Syntenic analysis confirmed that this region is conserved in platypus but contains two copies of a new protein that contains a Kunitz domain (Extended Data Fig. 10h and Supplementary Table 53). The Kunitz family is a rapidly evolving family, and one of the new members could have an immunoprotective function similar to *SPINT3* in monotremes.

The monotreme genomes contain most of the milk genes that have been identified in therian mammals<sup>38,40</sup>. Most mammals have three casein genes<sup>41</sup>, which encode the most abundant milk proteins secreted throughout lactation (Fig. 4c). In addition to these genes, monotremes have extra caseins that are not found in therian mammals, with unknown functions, an extra copy of *CSN2* (*CSN2B*) (previously reported<sup>40</sup>) and *CSN3* (*CSN3B*) in platypus (described here), which has the classic structure of *CSN3*<sup>42</sup> (Extended Data Fig. 10i and Supplementary Table 54).

All caseins are members of the secretory calcium-binding phosphoprotein (SCPP) gene family and are thought to have evolved from other SCPP genes, namely the teeth-related gene *ODAM* through its derivatives *FDCSP* and *SCPPPQ1*<sup>42</sup>. As reported above (see 'Eco-evolutionary adaptation of diet'), extant monotremes appear to have lost both *ODAM* and *FDCSP*. Syntenic analysis showed that the additional monotreme casein genes (*CSN2B* and *CSN3B*) are found in the same therian chromosomal region as *ODAM* and *FDCSP* and within the casein locus (Fig. 4c), providing further evidence that caseins evolved from odontogenic genes.

## Summary

Complete and accurate reference genomes and annotations are critical for evolutionary and functional analyses. It remains a challenge to produce a highly accurate chromosome-level assembly, particularly for differentiated sex chromosomes. We have produced a high-quality platypus genome using a combination of single-molecule sequencing technology and multiple sources of physical mapping methods to assign most of the sequences to a chromosome-scale assembly. This permits better-resolved analyses of the origin and diversification of the complex sex chromosome system that evolved specifically in monotremes. We delineate ancient and lineage-specific changes in the sensory system, haemoglobin degradation and reproduction that represent some of the most fascinating biology of platypus and echidna. The new genomes of both species will enable further insights into therian innovations and the biology and evolution of these extraordinary egg-laying mammals.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-03039-0>.

- Ashwell, K. *Neurobiology of Monotremes: Brain Evolution in Our Distant Mammalian Cousins* (CSIRO PUBLISHING, 2013).
- Warren, W. C. et al. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**, 175–183 (2008).
- Grützner, F. et al. In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. *Nature* **432**, 913–917 (2004).
- Kortschak, R. D., Tsend-Ayush, E. & Grützner, F. Analysis of SINE and LINE repeat content of Y chromosomes in the platypus, *Ornithorhynchus anatinus*. *Reprod. Fertil. Dev.* **21**, 964–975 (2009).
- Boissinot, S. & Sookdeo, A. The evolution of LINE-1 in vertebrates. *Genome Biol. Evol.* **8**, 3485–3507 (2016).
- Phillips, M. J., Bennett, T. H. & Lee, M. S. Molecules, morphology, and ecology indicate a recent, amphibious ancestry for echidnas. *Proc. Natl Acad. Sci. USA* **106**, 17089–17094 (2009).
- Bellott, D. W. et al. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494–499 (2014).
- Whittington, C. M. et al. Defensins and the convergent evolution of platypus and reptile venom genes. *Genome Res.* **18**, 986–994 (2008).
- Julien, P. et al. Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol.* **10**, e1001328 (2012).
- Rousselle, M., Laverré, A., Figue, E., Nabholz, B. & Galtier, N. Influence of recombination and GC-biased gene conversion on the adaptive and nonadaptive substitution rate in mammals versus birds. *Mol. Biol. Evol.* **36**, 458–471 (2019).
- Hinch, A. G., Altemose, N., Noor, N., Donnelly, P. & Myers, S. R. Recombination in the human pseudoautosomal region PARI. *PLoS Genet.* **10**, e1004503 (2014).
- Burt, D. W. Origin and evolution of avian microchromosomes. *Cytogenet. Genome Res.* **96**, 97–112 (2002).
- Dohm, J. C., Tsend-Ayush, E., Reinhardt, R., Grützner, F. & Himmelbauer, H. Disruption and pseudoautosomal localization of the major histocompatibility complex in monotremes. *Genome Biol.* **8**, R175 (2007).
- Cortez, D. et al. Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**, 488–493 (2014).
- Zhou, Q. et al. Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science* **346**, 1246338 (2014).
- Veyrunes, F. et al. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res.* **18**, 965–973 (2008).
- Braasch, I. et al. The spotted gar genome illuminates vertebrate evolution and facilitates human–teleost comparisons. *Nat. Genet.* **48**, 427–437 (2016).
- Gruetzner, F., Ashley, T., Rowell, D. M. & Marshall Graves, J. A. How did the platypus get its sex chromosome chain? A comparison of meiotic multiples and sex chromosomes in plants and animals. *Chromosoma* **115**, 75–88 (2006).
- Gołczyk, H., Massouh, A. & Greiner, S. Translocations of chromosome end-segments and facultative heterochromatin promote meiotic ring formation in evening primroses. *Plant Cell* **26**, 1280–1293 (2014).
- de Waal Malefijt, M. & Charlesworth, B. A model for the evolution of translocation heterozygosity. *Heredity* **43**, 315–331 (1979).
- Casey, A. E., Daish, T. J., Barbero, J. L. & Grützner, F. Differential cohesin loading marks paired and unpaired regions of platypus sex chromosomes at prophase I. *Sci. Rep.* **7**, 4217 (2017).
- Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Griffiths, M. *The Biology of Monotremes* (Academic, 1978).
- Meredith, R. W., Zhang, G., Gilbert, M. T., Jarvis, E. D. & Springer, M. S. Evidence for a single loss of mineralized teeth in the common avian ancestor. *Science* **346**, 1254390 (2014).
- Springer, M. S. et al. Odontogenic ameloblast-associated (ODAM) is inactivated in toothless/enamelless placental mammals and toothed whales. *BMC Evol. Biol.* **19**, 31 (2019).
- Ordoñez, G. R. et al. Loss of genes implicated in gastric function during platypus evolution. *Genome Biol.* **9**, R81 (2008).
- Hayakawa, T., Suzuki-Hashido, N., Matsui, A. & Go, Y. Frequent expansions of the bitter taste receptor gene repertoire during evolution of mammals in the Euarchontoglires clade. *Mol. Biol. Evol.* **31**, 2018–2031 (2014).
- Johnson, R. N. et al. Adaptation and conservation insights from the koala genome. *Nat. Genet.* **50**, 1102–1111 (2018).
- Liu, Z. et al. Dietary specialization drives multiple independent losses and gains in the bitter taste gene repertoire of Laurasiatherian mammals. *Front. Zool.* **13**, 28 (2016).
- Hunnicut, K. E. et al. Comparative genomic analysis of the pheromone receptor class 1 family (VIR) reveals extreme complexity in mouse lemurs (genus, *Microcebus*) and a chromosomal hotspot across mammals. *Genome Biol. Evol.* **12**, 3562–3579 (2020).
- Johansen, K., Lenfant, C. & Grigg, G. C. Respiratory properties of blood and responses to diving of platypus *Ornithorhynchus anatinus* (Shaw). *Comp. Biochem. Physiol.* **18**, 597–608 (1966).
- Alayash, A. I. Haptoglobin: old protein with new functions. *Clin. Chim. Acta* **412**, 493–498 (2011).
- Wicher, K. B. & Fries, E. Haptoglobin, a hemoglobin-binding plasma protein, is present in bony fish and mammals but not in frog and chicken. *Proc. Natl Acad. Sci. USA* **103**, 4168–4173 (2006).
- Redmond, A. K. et al. Haptoglobin is a divergent masp family member that neofunctionalized to recycle hemoglobin via CD163 in mammals. *J. Immunol.* **201**, 2483–2491 (2018).
- Huttenlocker, A. K. & Farmer, C. G. Bone microvasculature tracks red blood cell size diminution in Triassic mammal and dinosaur forerunners. *Curr. Biol.* **27**, 48–54 (2017).
- Schaer, D. J. et al. CD163 is the macrophage scavenger receptor for native and chemically modified hemoglobins in the absence of haptoglobin. *Blood* **107**, 373–380 (2006).
- Griffiths, M. *Echidnas* (Pergamon, 1968).
- Brawand, D., Wahli, W. & Kaessmann, H. Loss of egg yolk genes in mammals and the origin of lactation and placentation. *PLoS Biol.* **6**, e63 (2008).
- Pharo, E. A. et al. The mammary gland-specific marsupial *ELP* and eutherian *CTI* share a common ancestral gene. *BMC Evol. Biol.* **12**, 80 (2012).
- Lefèvre, C. M., Sharp, J. A. & Nicholas, K. R. Characterisation of monotreme caseins reveals lineage-specific expansion of an ancestral casein locus in mammals. *Reprod. Fertil. Dev.* **21**, 1015–1027 (2009).
- Holt, C., Carver, J. A., Ercroft, H. & Thorn, D. C. Invited review: Caseins and the casein micelle: their biological functions, structures, and behavior in foods. *J. Dairy Sci.* **96**, 6127–6146 (2013).
- Kawasaki, K., Lafont, A. G. & Sire, J. Y. The evolution of milk casein genes from tooth genes before the origin of mammals. *Mol. Biol. Evol.* **28**, 2053–2061 (2011).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

<sup>1</sup>BGI-Shenzhen, Shenzhen, China. <sup>2</sup>Villum Center for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark. <sup>3</sup>School of Biological Sciences, The Environment Institute, The University of Adelaide, Adelaide, South Australia, Australia. <sup>4</sup>MOE Laboratory of Biosystems Homeostasis and Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, China. <sup>5</sup>BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, China. <sup>6</sup>Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Japan. <sup>7</sup>Japan Monkey Centre, Inuyama, Japan. <sup>8</sup>School of BioSciences, The University of Melbourne, Melbourne, Victoria, Australia. <sup>9</sup>School of Life and Environmental Sciences, The University of Sydney, Sydney, New South Wales, Australia. <sup>10</sup>digzyme Inc, Tokyo, Japan. <sup>11</sup>School of Life Science and Technology, Tokyo Institute of Technology, Tokyo, Japan. <sup>12</sup>The Genome Center, University of California, Davis, CA, USA. <sup>13</sup>Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA. <sup>14</sup>Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. <sup>15</sup>Tree of Life Programme, Wellcome Sanger Institute, Cambridge, UK. <sup>16</sup>The Vertebrate Genome Lab, The Rockefeller University, New York, NY, USA. <sup>17</sup>James D. Watson Institute of Genome Sciences, Hangzhou, China. <sup>18</sup>University of the Chinese Academy of Sciences, Beijing, China. <sup>19</sup>Guangdong Provincial Academician Workstation of BGI Synthetic Genomics, BGI-Shenzhen, Shenzhen, China. <sup>20</sup>School of Agriculture and Food Sciences, The University of Queensland, Gatton, Queensland, Australia. <sup>21</sup>Laboratory of Neurogenetics of Language, The Rockefeller University, New York, NY, USA. <sup>22</sup>Howard Hughes Medical Institute, Chevy Chase, MD, USA. <sup>23</sup>San Diego Zoo Global, Escondido, CA, USA. <sup>24</sup>Center for Molecular Biology of Heidelberg University (ZMBH), DKFZ-ZMBH Alliance, Heidelberg, Germany. <sup>25</sup>Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>26</sup>Pacific Biosciences, Menlo Park, CA, USA. <sup>27</sup>Department of Evolution and Ecology, College of Biological Sciences, University of California, Davis, CA, USA. <sup>28</sup>Department of Reproduction and Population Health, School of Veterinary Medicine, University of California, Davis, CA, USA. <sup>29</sup>Research School of Biology, Australian National University, Canberra, Australian Capital Territory, Australia. <sup>30</sup>Institute for Applied Ecology, University of Canberra, Canberra, Australian Capital Territory, Australia. <sup>31</sup>School of Life Sciences, La Trobe University, Melbourne, Victoria, Australia. <sup>32</sup>Department of Neuroscience and Developmental Biology, University of Vienna, Vienna, Austria. <sup>33</sup>Center for Reproductive Medicine, The 2nd Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China. <sup>34</sup>State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. <sup>35</sup>Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China. <sup>36</sup>These authors contributed equally: Yang Zhou, Linda Shearwin-Whyatt, Jing Li. <sup>✉</sup>e-mail: frank.grutzner@adelaide.edu.au; zhouqi1982@zju.edu.cn; guojie.zhang@bio.ku.dk

## Methods

### Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

### Ethics and sample collection

Pmale08, Pmale09 and Emale12 were collected under AEC permits S-49-2006, S-032-2008 and S-2011-146 at Upper Barnard River (New South Wales, Australia) during the breeding season. Emale01 was collected under San Diego Zoo Global IACUC approval 18-024 and vouched at San Diego Natural History Museum.

### Sequencing and assembling

Skeletal muscle of Pmale09 was used for PacBio, 10X and BioNano genome sequencing and the liver of Pmale09 was used for Hi-C (Phase Genomics); the liver of Pmale08 was used for Chicago Hi-C (Dovetail Genomics). Heart muscle of Emale01 was used for a variety of library construction and Illumina sequencing analyses. Muscle of Emale12 was used for 10X and BioNano genome sequencing and liver of Emale12 was used for Hi-C (Phase Genomics). Echidna RNA was extracted from brain, cerebellum, kidney, liver, testis and ovary and sequenced using a previously published procedure<sup>43</sup>. Platypus Y chromosome BAC isolation via hybridization was performed using a previous published procedure<sup>4</sup> and sequenced with PacBio. The platypus genome was assembled following VGP assembly pipeline v.1.0. The echidna genome was assembled using Platanus<sup>44</sup> (v.1.2.1) and followed by three steps of scaffolding in the order of 10X, BioNano and Hi-C. Manual curation was performed for both assemblies. Details are available in the Supplementary Methods.

### Sex-borne sequence identification

Female and male reads were mapped to the genome using BWA ALN<sup>45</sup> (v.0.7.12). The read depth of each sex was calculated in 5-kb non-overlapping windows to identify X-borne sequences and 2-kb non-overlapping windows to identify Y-borne sequences, normalized against the median depth. To identify X-borne sequences, we calculated the female-to-male (F/M) depth ratio of regions that were covered by both sexes in each scaffold, requiring a minimum coverage of 80%, and assigned sequences to X-borne if the depth ratio ranged between 1.5 and 2.5. To identify Y-borne sequences, we calculated the F/M depth ratio as well as the F/M coverage ratio and assigned scaffolds to Y-borne if either ratio was within the range of 0.0–0.3. Parameter evaluation details are available in the Supplementary Methods.

### Chromosome anchoring

We collected 75 BAC and 179 marker genes (Supplementary Table 3) and ordered them according to their relative order from those papers. Protein sequences of the gene markers were compared to mOrnAna1 using TBLASTN and the best hit was kept, after which the markers were analysed using GeneWise<sup>46</sup> (v.2.4.1) to obtain the location within a scaffold. BAC-end reads were mapped to the assembly using BWA MEM<sup>47</sup> (v.0.7.12) and the best hits were kept. We also used the anchored sequences of OANA5 except for the sequence of chromosome 14 to anchor the scaffolds into chromosomes. Scaffolds were orientated and ordered first based on the order of FISH or gene markers then on the order in OANA5. All identified PARs were included in chromosome X. We collected assembled Y contigs from a previously published study<sup>4</sup> and generated some Y-BAC PacBio sequencing data. Assembled Y contigs were mapped to the platypus assembly using BWA MEM and Y-BAC PacBio reads were mapped using minimap2 (v.2.13)<sup>48</sup>. As evidence of both Y2 and Y3 were found on scaffold\_229\_arrow\_ctg1 and scaffold\_269\_arrow\_ctg1 and the covered regions overlapped, these two scaffolds were excluded from the chromosome Y classification. Classified Y-borne scaffolds failed to anchor and orient due to the lack

of information. We also curated and anchored some echidna X-borne scaffolds to chromosome X based on Mashmap<sup>49</sup> (v.2.0) one-to-one results with platypus<sup>50</sup>.

### Annotation

We identified repetitive elements in both assemblies using the same pipeline, which included homologue-based and de novo prediction. For the homology-based method, we used default repeat library from Repbase (v.21.11)<sup>51</sup> for RepeatMasker (v.4.0.6)<sup>52</sup>, trf (v.4.07)<sup>53</sup> and Proteinmasker (v.4.0.6)<sup>52</sup> to annotate. For the de novo method, we first ran RepeatModeler (v.1.0.8) to construct the consensus sequence library for each monotreme using their genome as input, then aligned the genome against each consensus library to identify repeats using RepeatMasker. Gene annotation was performed by merging the homology, de novo prediction and transcriptome analyses to build a consensus gene set of each species. Protein sequences from human, mouse, opossum, platypus, chicken and green lizard (*Anolis carolinensis*) from Ensembl<sup>54</sup> (release 87) were aligned to the genome using TBLASTN<sup>55</sup> (v.2.2.26) ( $e < 1 \times 10^{-5}$ ). Candidate gene regions were refined using GeneWise for more accurate gene models. We randomly selected 1,000 high score homology-based genes to train Augustus<sup>56</sup> (v.3.0.3) for de novo prediction on a repeat N-masked genome. We also mapped RNA-sequencing reads of the platypus from a previously published study<sup>57</sup> and echidna to their respective assemblies using HISAT2<sup>58</sup> (v.2.0.4), and constructed transcripts using stringTie<sup>59</sup> (v.1.2.3). Results from these three methods were merged into a nonredundant gene set. Possible retrogenes were filtered according to their hit to SwissProt database<sup>60</sup> (release 2015\_12) or Iprscan<sup>61</sup> (v.5.16-55.0). We used the SwissProt database ( $e < 1 \times 10^{-5}$ ) to annotate the function of the genes. Iprscan was used to annotate the GO of genes. Detailed descriptions of the manual annotation, curation and phylogenetic analysis of genes related to imprinting, immune system, reproduction and haemoglobin degradation can be found in the Supplementary Methods.

### Gap analysis

We identified gap-filling regions using an alignment-based strategy similar to a previously published study<sup>62</sup>. We considered gaps for which both flanking regions mapped to mOrnAna1 as closed gaps. Only properly closed gaps defined by (1) both flanking regions were aligned but did not overlap and (2) closed gap size were within 100 times the estimated gap size in OANA5 were considered for repeat and gene improvement analysis.

### Redundant sequences analysis

We performed two rounds of Mashmap with parameters '-f one-to-one -s 2000' using mOrnAna1 as reference and OANA5 as query. A one-to-one relationship was obtained in the first round of Mashmap. In the second round of Mashmap, those OANA5 sequences that were unmapped in the first round of mapping were used as query. Candidate redundant sequences were obtained from the second round Mashmap result, but excluded regions that were gaps in OANA5. Female and male reads were then mapped to OANA5 and mOrnAna1 using BWA ALN and normalized by the mode depth.

### Gene set comparison

We performed LASTZ<sup>63</sup> (v.1.04.00) alignment using OANA5 as reference with parameter set '-hspthresh=4500 --gap=600,150 --ydrop=15000 --notransition --format=axt' and a score matrix for the comparison of closely related species to generate a chain file for gene location liftover from OANA5 to mOrnAna1. Gene coordinates in OANA5 were first converted to mOrnAna1 using in-house-generated scripts with the chain file. We searched for overlap between the converted OANA5 gene set and mOrnAna1 gene set. Fragmented genes were defined as multiple converted OANA5 genes that overlapped with a single mOrnAna1 gene. A one-to-one gene pair between the two gene sets was defined



# Article

as the liftover of the OANAS gene when it overlapped with only one mOrnAna1 gene. Only one-to-one pairs were used for the comparison of open-reading frame completeness. We defined a gene as having a complete open-reading frame if its first codon is a start codon and the last codon is a stop codon.

## Identification of one-to-one orthologues and syntenic blocks between the human sequence and sequences of other species

We defined one-to-one orthologues between the human sequence and the sequences of other species by considering both reciprocal best BLASTP hits (RBH) and syntenic, taking the human sequence as reference, as previously described<sup>64</sup>. First, we conducted BLASTP for all protein sequences from human and other species including mouse, opossum, platypus, echidna, chicken and green lizard with an *e*-value cut-off of  $1 \times 10^{-7}$ , and combined local alignments with the SOLAR (<http://treesoft.svn.sourceforge.net/viewvc/treesoft/>). Next, we identified RBH orthologues between human and every other species on the basis of the following parameters: alignment score, alignment rate and identity. From these RBH orthologues, we retained those pairs with conserved syntenic across species. Syntenic was determined based on their flanking genes. If RBH orthologous gene pairs shared the same flanking genes, we retained the genes for downstream analyses. Finally, we merged pairwise orthologue lists according to the human coordinates. In this way, we produced the final one-to-one orthologue set across species.

We used the human genome as the reference and aligned it with other species using LASTZ with parameter set ‘--hsptresh=4500 --gap=600,150 --ydrop=15000 --notransition --format=axt’ and a score matrix for the comparison of closely related species. Alignments were converted into ‘chain’ and ‘net’ results with different levels of alignment scores using utilities of the UCSC Genome Browser (<http://genome.ucsc.edu/index.php/>), and the pairwise syntenic blocks between genomes of each species and the human genome were extracted according to the net result. Only alignments larger than 10 kb were kept. The syntenic blocks were further cleaned of overlapping genes. N50 and the total length of the syntenic block inferred from each human–species pair were calculated based on the human coordinates.

## Phylogenetic analysis

The phylogenetic tree was constructed using concatenated four-degenerated sites from the 7,946 one-to-one orthologues using RaxML<sup>65</sup> (v.8.2.4) with parameter set ‘-m GTRCAT -# 100 -p 12345 -x 12345 -f a’ and chicken and green lizard were specified as the outgroup. MCMCtree in PAML<sup>66</sup> (v.4.7) was used to estimate divergence time of each species with calibration points obtained from a previously published study<sup>67</sup> using the same data. Points and time range included the most recent common ancestor of human–mouse, 85–94 million years ago; human–opossum 150–167 million years ago; human–platypus, 163–191 million years ago, human–chicken, 297–326 million years ago, anole–chicken, 276–286 million years ago. The seed used for MCMC was 1192664277.

## Substitution rate analysis

We first performed pairwise whole-genome LASTZ alignment using 12 mammals (*Macaca mulatta*, *Tupaia belangeri*, *Mus musculus*, *Canis lupus familiaris*, *Myotis lucifugus*, *Bos taurus*, *Sorex araneus*, *Loxodonta africana*, *Dasyurus novemcinctus*, *Monodelphis domestica*, *O. anatinus* and *T. aculeatus*) with the human genome as the reference genome, with the parameter set ‘--step=19 --hsptresh=2200 --inner=2000 --ydrop=3400 --gappedthresh=10000 --format=axt’ and a score matrix for the comparison of distantly related species. Pairwise alignments were merged using MULTIZ<sup>68</sup> (v.11.2). The four-degenerated site alignment was extracted based on the human gene set (Ensembl release 87), concatenated and fed to phyloFit in the PHAST package<sup>69</sup> (v.1.5) for the calculation of branch lengths (substitution per site). The substitution

rate was calculated by dividing the branch length to the mammalian common ancestor to the mammal–reptile divergence time.

## Gene family analysis

Gene families across the seven species were generated using orthoMCL<sup>70</sup> (v.2.0.9) with BLASTP results ( $e < 1 \times 10^{-7}$ ) and was fed to CAFÉ<sup>71</sup> (v.4.2) along with the phylogenetic tree. We first estimated the assembly error by excluding families with more than 100 members. Then the estimated rate was used to infer the family size at every node for each family. The ancestral node gene number of families with more than 100 members among extant species were inferred separately. We extracted genes based on the human gene set for GO enrichment ( $\chi^2$  test) of the significantly expanded family (Viterbi  $P < 0.05$ ) for the mammalian ancestor. A false-discovery rate (FDR) adjustment was used for multiple-test corrections in GO enrichment analyses.

## Mammalian-specific highly conserved element analysis

We used the same MULTIZ alignment of the substitution rate analysis and identified mammalian-specific highly conserved elements (MSHCEs) using a similar strategy as has previously been described<sup>72</sup>. At least 80% of species and at least one species in eutherians, marsupials and monotremes were required to be present in alignments. Type-I MSHCEs were defined as HCEs to which no outgroup could be aligned; type-II MSHCEs were HCEs that were significantly conserved ( $P < 0.01$ ) in mammals compared to mammals + outgroup calculated using phyloP (Benjamini–Hochberg adjusted). We considered four sets of outgroup combinations: (1) green lizard only; (2) chicken only; (3) two reptiles and one frog; and (4) two reptiles, one frog and one fish, and only kept those that were significantly conserved in all four sets of statistical tests (Benjamini–Hochberg adjusted  $P < 0.01$ ). Only elements  $\geq 20$  bp were kept for further analysis.

To annotate MSHCEs to possible functional elements, we used the human annotation (Ensembl release 87) as a reference and classified the elements into the coding sequence, 5′ and 3′ untranslated regions, non-coding RNA, pseudogene, intron, upstream 10-kb region (from start codon), downstream 10-kb region (from the stop codon) and intergenic regions, with the same hierarchical order if the regions overlapped. Genes located within the upstream or downstream 10-kb range of MSHCEs were considered to be MSHCE-associated genes, and ordered by the length of the element. The top-300 MSHCE-associated genes were used in the GO enrichment analysis ( $\chi^2$  test, FDR-adjusted) and visualized using REVIGO<sup>73</sup>.

## Mammalian karyotype reconstruction

We used pairwise LASTZ alignments of the opossum, Tasmanian devil, platypus, chicken and common wall lizard (*Podarcis muralis*) genomes to the human genome as input. Echidna was not used here as most of the sequences were not anchored to chromosomes, which would lead to a more fragmented reconstruction. With the net and chain results, conserved segments that were uniquely and universally presented in all six species were obtained using inferCARs<sup>74</sup> (release 2006-Jun-16). Marsupial and therian ancestral karyotypes were inferred using ANGES<sup>75</sup> (v.1.01) using the branch-and-bound algorithm, and the resulting continuous ancestral regions (CARs) were further reorganized based on the previously predicted configuration<sup>76</sup> (Supplementary Tables 22, 23). We replaced the conserved segments of the human, opossum and Tasmanian devil genomes with those of the reconstructed therian ancestral karyotype and reconstructed marsupial ancestral karyotype using ANGES with the same parameters except setting the target reconstruction node to mammalian ancestor. We reorganized CARs on the basis of gene syntenic among ingroups and outgroups inferred using MCScanX<sup>77</sup> (release 08-05-2012), requiring that there is syntenic across CARs in at least one ingroup–outgroup pair (Supplementary Tables 22, 23). Pairwise MCScanX was run among the six species with BLASTP ( $e < 1 \times 10^{-7}$ ).

Rearrangement events in each lineage were inferred using GRIMM<sup>78</sup> (v.2.1) by taking the karyotypes of the most recent ancestor and the child as input. The breakpoint number in each lineage was calculated on the basis of the output of GRIMM using an in-house-generated script, in which one breakpoint was counted in fission, two breakpoints were counted in translocation, and one or two breakpoints were counted in inversion, depending on whether the inversion happened at the end of the chromosome. Calculations were done using resolutions of 500 kb and 300 kb, and using the raw ANGES output and reorganized output, respectively (Supplementary Table 28). Differences in breakpoint rates compared to the average of all branches were tested as previously described<sup>79</sup>.

### Gametologue identification

We used BLASTP to compare all Y-borne genes to all X-borne genes ( $e < 1 \times 10^{-5}$ ) and kept the best hit for each Y-borne gene. Candidate gametologue pairs were further confirmed if both of the genes were mapped to the same gene in NCBI or the SwissProt database. Four gametologues (platypus *AMHX* and *FEMICX* from OANAS, and *SDHAY* and *HNRNP KY* from ref. <sup>14</sup>) were added as they were missing in mOrnAna1. Translated genes were aligned using PRANK<sup>80</sup>, filtered using Gblock<sup>81</sup>, and converted back into the alignment of the coding sequence. dS was calculated using codeml in PAML with 'runmode=-2'.

### Demarcate evolutionary strata

We aligned all platypus Y-borne scaffolds (N-masked) to all platypus X-borne sequences (N-masked), and aligned all echidna Y-borne scaffolds (N-masked) to all echidna X-borne sequences (N-masked), using LASTZ with the parameter set '-step=19 --hspthresh=2200 --inner=2000 --ydrop=3400 --gappedthresh=10000 --format=act' and a score matrix set for the comparison of distantly related species. On the basis of the net and 'maf' results, the identity of each alignment block was calculated in a 2-kb non-overlapped window and the aligned Y-borne sequences were oriented along the X chromosomes. Identity along X chromosomes was colour-coded for visualization.

### Expression calculation

RNA-sequencing reads of platypus (SRP102989) and echidna were mapped to the genome using HISTA2. Uniquely mapped reads were used in the calculation and normalization of the reads per kilobase per million reads (RPKM) using DESeq<sup>82</sup> (v.1.28.0) to generate an expression matrix for each species. For tissues that were available in both sexes, we computed the median RPKM of each X-borne gene, and computed its F/M RPKM ratio (requiring RPKM in both sexes to be  $\geq 1$ ) to determine dosage-compensation status. We used the median expression value in each tissue to calculate the tissue specificity index TAU<sup>83</sup> for each gene. We defined tissue-specific expression as a gene that shows at least twofold higher expression in tissue with the highest expression than in any other tissue, the highest RPKM  $> 1$  and TAU  $> 0.8$ .

### Building genome-wide Hi-C interaction maps

Genome-wide interaction maps at a 100-kb resolution were generated for platypus, echidna and human (SRX641267) with HiC-Pro<sup>84</sup> (v.2.10.0). For echidna, we only retained scaffolds  $> 10$  kb as the large number of short scaffolds would cause ICE normalization failure. The normalized sex chromosomes submatrix was extracted for quantification and plotting with ggplot2 (v.3.2.1). For human, we used the scaled homologous sequences of platypus for quantification and plotting.

### Identification of TADs and CTCF-binding sites

HiC-Pro interaction maps were transformed to h5 format using hicConvertFormat and fed to hicFindTADs with the parameters '--outPrefix TAD --numberOfProcessors 32 --correctForMultipleTesting fdr' to identify TADs with HiCExplorer<sup>85</sup> (v.3.0). The human CTCF motif<sup>86</sup> was used

as a bait by fimo in MEME<sup>87</sup> (v.4.12.0) to identify putative CTCF-binding sites. CTCF densities in every 100 kb non-overlapping sliding window along the platypus sex chromosomes or scaled homologous sequences of echidna, human and chicken were compared.

### FISH

BACs were obtained from the Children's Hospital Oakland Research Institute from the platypus BAC library CH236: CH236-775N6 13q2; CH236-9713 15p1 and CUGI BAC/EST resource centre from the platypus BAC library Oa\_Ab: Oa\_Bb-155A12 autosomal (*WSBI*); Oa\_Bb-145P09 Y2; Oa\_Bb-397121 Y3. The Super\_Scaffold\_40-specific probe was amplified from platypus genomic DNA. Gene ENSOANT 0000009075.3 was amplified using primers GTCTAAAGACAAGTG TACATCTGTGAC and GTGACTTCTCTTGCGAACACAC. The 3.9-kb product was cloned into pGEM-T Easy (Promega). BAC probes were directly labelled with dUTP Alexa Fluor 594-dUTP, aminoallyl-dUTP-XX-ATTO-488 (Jena Bioscience) using the Nick Translation Kit (Roche Diagnostics) and the Super\_Scaffold\_40-specific probe labelled with biotin using the Biotin-Nick Translation Mix (1175824919, Roche Diagnostics). The FISH protocol was carried out on cultured fibroblasts from platypus (authenticated by karyotype, not mycoplasma tested) obtained from animals captured at the Upper Barnard River (New South Wales, Australia) during the breeding season (AEC permits S-49-2006, S-032-2008 and S-2011-146) as previously described<sup>88</sup> with the following exceptions. Slides were denatured at 70 °C for 3 min in 70% formamide in 2× SSC, 1 mg DNA probe was used per slide, pre-annealing of repetitive DNA sequences was done at 37 °C for 30–60 min. Detection of biotin-labelled probes was done using Rhodamine Avidin D (Vector Laboratories, A-2002), goat Biotinylated anti-avidin D (Vector Laboratories, BA-0300) and Rhodamine Avidin D. Slides were blocked in 4 × SSC, 1% BSA fraction V, for 30 min at 37 °C. Rhodamine Avidin D and Biotinylated anti-avidin D and the second Rhodamine Avidin D were diluted in 4 × SSC, 1% BSA fraction V and were incubated on slides for 45 min at 37 °C, after each step washes were done in 4 × SSC, 4 × SSC, 0.1% triton, 4 × SSC at room temperature for 10 min each. Slides were mounted in VECTASHIELD with DAPI (Vector Laboratories, H-1200). Sample size was determined according to ref. <sup>89</sup>, but was limited by material availability. Images were captured on a Nikon Ti Microscope using NIS-Elements AR 4.20.00 software and processed with ImageJ (v.2.0.0). Fisher's exact test was performed with matrix containing mean of associated and non-associated cells from the three replicates. No blinding nor randomization was performed.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

### Data availability

The platypus whole-genome shotgun project has been deposited at GenBank (project accessions PRJNA489114 and PRJNA489115), CNSA (<https://db.cngb.org/cnsa/>) of CNGBdb (accession CNP0000130) and GenomeArk ([https://vgp.github.io/genomeark-curated-assembly/Ornithorhynchus\\_anatinus/](https://vgp.github.io/genomeark-curated-assembly/Ornithorhynchus_anatinus/)). The echidna whole-genome shotgun project has been deposited at GenBank (project accession PRJNA576333), CNSA of CNGBdb (accession CNP0000697) and GenomeArk at ([https://vgp.github.io/genomeark/Tachyglossus\\_aculeatus/](https://vgp.github.io/genomeark/Tachyglossus_aculeatus/)). Echidna RNA-sequencing data have been deposited at GenBank (project accession PRJNA591380) and CNSA of CNGBdb (accession CNP0000779). Public database used in this study include: NCBI (<https://www.ncbi.nlm.nih.gov/>), Ensembl (release 87) (<http://dec2016.archive.ensembl.org/index.html>), Uniprot (<https://www.uniprot.org/>) and Repbase (<https://www.girinst.org/repbase/>). Accession codes of genes are available in Supplementary Tables 31, 33, 37, 49, 51.

## Code availability

In-house-generated scripts used in this study are shared on GitHub (<https://github.com/ZhangLabSZ/MonotremeGenome>).

43. Cardoso-Moreira, M. et al. Gene expression across mammalian organ development. *Nature* **571**, 505–509 (2019).
44. Kajitani, R. et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014).
45. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
46. Birney, E., Clamp, M. & Durbin, R. Genewise and genomewise. *Genome Res.* **14**, 988–995 (2004).
47. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
48. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
49. Jain, C., Koren, S., Diltthey, A., Phillippy, A. M. & Aluru, S. A fast adaptive algorithm for computing whole-genome homology maps. *Bioinformatics* **34**, 1748–1756 (2018).
50. Rens, W. et al. The multiple sex chromosomes of platypus and echidna are not completely identical and several share homology with the avian Z. *Genome Biol.* **8**, R243 (2007).
51. Bao, W., Kojima, K. K. & Kohany, O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
52. Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **5**, 4.10.1–4.10.14 (2004).
53. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
54. Yates, A. et al. Ensembl 2016. *Nucleic Acids Res.* **44**, D710–D716 (2016).
55. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
56. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).
57. Brawand, D. et al. The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
58. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
59. Perteza, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
60. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
61. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
62. Bickhart, D. M. et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.* **49**, 643–650 (2017).
63. Harris, R. S. *Improved Pairwise Alignment of Genomic DNA*. PhD thesis, Pennsylvania State Univ. (2007).
64. Zhang, G. et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
65. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
66. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
67. Benton, M. J. et al. Constraints on the timescale of animal evolutionary history. *Palaeontol. Electronica* **18**, 1–106 (2015).
68. Blanchette, M. et al. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).
69. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).
70. Li, L., Stoeckert, C. J., Jr & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
71. Han, M. V., Thomas, G. W., Lugo-Martinez, J. & Hahn, M. W. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997 (2013).
72. Seki, R. et al. Functional roles of Aves class-specific cis-regulatory elements on macroevolution of bird-specific features. *Nat. Commun.* **8**, 14229 (2017).
73. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).
74. Ma, J. et al. Reconstructing contiguous regions of an ancestral genome. *Genome Res.* **16**, 1557–1565 (2006).
75. Jones, B. R., Rajaraman, A., Tannier, E. & Chauve, C. ANGES: reconstructing ANcestral GEnomeS maps. *Bioinformatics* **28**, 2388–2390 (2012).
76. Deakin, J. E. et al. Reconstruction of the ancestral marsupial karyotype from comparative gene maps. *BMC Evol. Biol.* **13**, 258 (2013).
77. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
78. Tesler, G. GRIMM: genome rearrangements web server. *Bioinformatics* **18**, 492–493 (2002).
79. Kim, J. et al. Reconstruction and evolutionary history of eutherian chromosomes. *Proc. Natl Acad. Sci. USA* **114**, E5379–E5388 (2017).
80. Löytynoja, A. Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* **1079**, 155–170 (2014).
81. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
82. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
83. Yanai, I. et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).
84. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
85. Ramirez, F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
86. Jolma, A. et al. DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).
87. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Syst. Mol. Biol.* **2**, 28–36 (1994).
88. Tsend-Ayush, E. et al. Higher-order genome organization in platypus and chicken sperm and repositioning of sex chromosomes during mammalian evolution. *Chromosoma* **118**, 53–69 (2009).
89. Ling, J. Q. et al. CTCF mediates interchromosomal colocalization between *Igf2/H19* and *Wsb1/Nfi*. *Science* **312**, 269–272 (2006).
90. Parra, Z. E. et al. Comparative genomic analysis and evolution of the T cell receptor loci in the opossum *Monodelphis domestica*. *BMC Genomics* **9**, 111 (2008).
91. Van Laere, A. S., Coppieters, W. & Georges, M. Characterization of the bovine pseudoautosomal boundary: documenting the evolutionary history of mammalian sex chromosomes. *Genome Res.* **18**, 1884–1895 (2008).

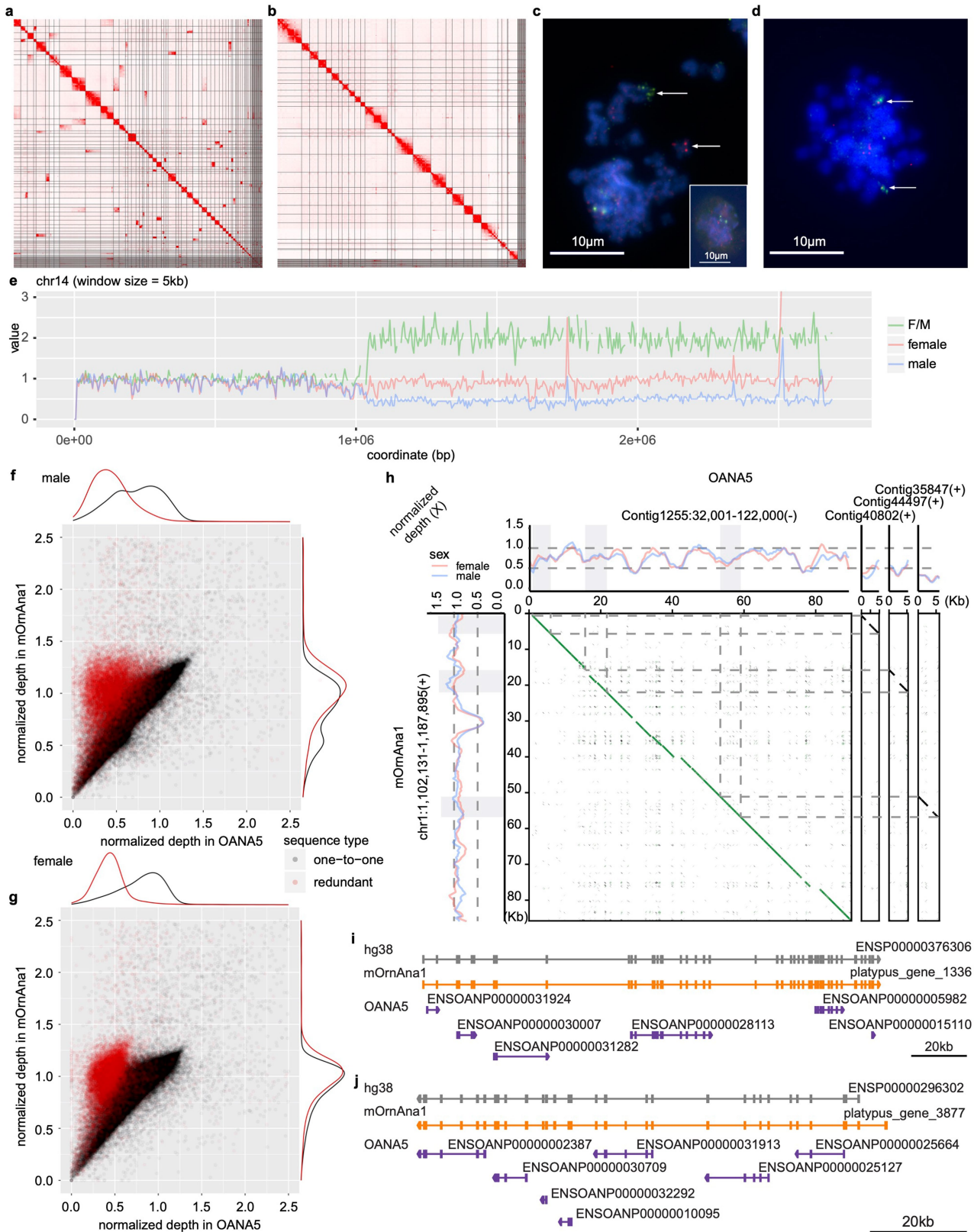
**Acknowledgements** We thank members of BGI-Shenzhen, China National GeneBank and VGP, and P. Baybayan, R. Hall and J. Howard for help carrying out the sequencing of the platypus and echidna genomes, M. Asahara for discussion, and D. Charlesworth for comments. Work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31020000), the National Key R&D Program of China (MOST) grant 2018YFC1406901, International Partnership Program of Chinese Academy of Sciences (152453KYSB20170002), Carlsberg foundation (CF16-0663) and Villum Foundation (25900) to G.Z. Q.Z. is supported by the National Natural Science Foundation of China (31722050, 31671319 and 32061130208), Natural Science Foundation of Zhejiang Province (LD19CJ90001), European Research Council Starting Grant (grant agreement 677696) and start-up funds from Zhejiang University, F.G., L.S.-W. and T.D. are supported by Australian Research Council (FT160100267, DP170104907 and DP110105396). M.B.R., J.C.F. and S.D.J. are supported by the Australian Research Council (LP160101728). We acknowledge the Kyoto University Research Administration Office for support and Human Genome Center, the Institute of Medical Science, the University of Tokyo for the super-computing resource for supporting T.H.'s research facilities. T.H. was financed by JSPS KAKENHI grant numbers 16K18630 and 19K16241 and the Sasakawa Scientific Research Grant from the Japan Science Society. The echidna RNA-sequencing analysis was supported by H.K.'s grant from the European Research Council (615253, OntoTransEvol). This work was supported by Guangdong Provincial Academician Workstation of BGI Synthetic Genomics No. 2017B090904014 (H.Y.), Robert and Rosabel Osborne Endowment, Howard Hughes Medical Institute (E.D.J.), Rockefeller University start-up funds (E.D.J.), Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health (A.R. and A.M.P.), Korea Health Technology R&D Project through the Korea Health Industry Development Institute HI17C2098 (A.R.). This work used the computational resources of BGI-Shenzhen and the NIH HPC Biowulf cluster (<https://hpc.nih.gov>). Animal icons are from <https://www.flaticon.com/> (made by Freepik) and <http://phylopic.org/>.

**Author contributions** G.Z., Q.Z. and F.G. initiated and designed the monotreme genome project with early input from J.G. G.Z., F.G., Q.L., T.D., B.H., J.M., O.F., E.D.J., O.A.R., H.K., P.D., J.K., T.P., H.Y. and J. Wang coordinated the project and were involved in the collection, extraction and sequencing of samples. A.M.P., A.R., G.Z., Y.Z., K.H., Y.S., J. Wood, Q.L., Q.Z., F.G. and L.S.-W. performed genome assembling, evaluation and chromosome assignment. G.Z., Y.Z., Z.S., Y.G. and F.G. performed the annotation and mammalian macro-evolutionary analysis including divergence time estimation, gene family and MSHCE analysis. G.Z., Y.Z., H.A.L., J.D. and Q.Z. performed ancestral karyotype analysis. Q.Z., J.L., Z.Z., F.G., F.P., G.Z., Y.Z., L.S.-W. and Y.G. performed analysis to sex chromosome and FISH validation. K.B., E.P., Y.C., Y.Z., D.S., Z.S., G.Z. and F.G. performed the immune gene analysis. Y.Z., J.C.F., M.B.R., Z.S. and G.Z. performed the tooth-related gene analysis. N.B., F.G. and Y.Z. performed the digestive gene analysis and PCR validation. T.H., H.S., M.N. and F.G. performed the chemosensory gene analysis. L.S.-W., F.G., Y.Z. and Z.S. performed the haemoglobin-degradation gene analysis. M.B.R., J.C.F. and S.J. performed the reproductive gene analysis. G.Z., Q.Z., F.G., Y.Z., L.S.-W., J.L., M.B.R., J.C.F., K.B., E.P., Y.C., D.S., N.B., F.P., T.H., H.A.L., J.D., A.M.P., A.R., K.H., J. Wood, O.F. and J.G. wrote the manuscript.

**Competing interests** J.K. is an employee of Pacific Biosciences, a company that develops single-molecule sequencing technologies.

**Additional information**  
**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-03039-0>.

**Correspondence and requests for materials** should be addressed to F.G., Q.Z. or G.Z.  
**Peer review information** Nature thanks Janine Deakin, Rebecca Johnson and Hugues Roest Colluis for their contribution to the peer review of this work.  
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

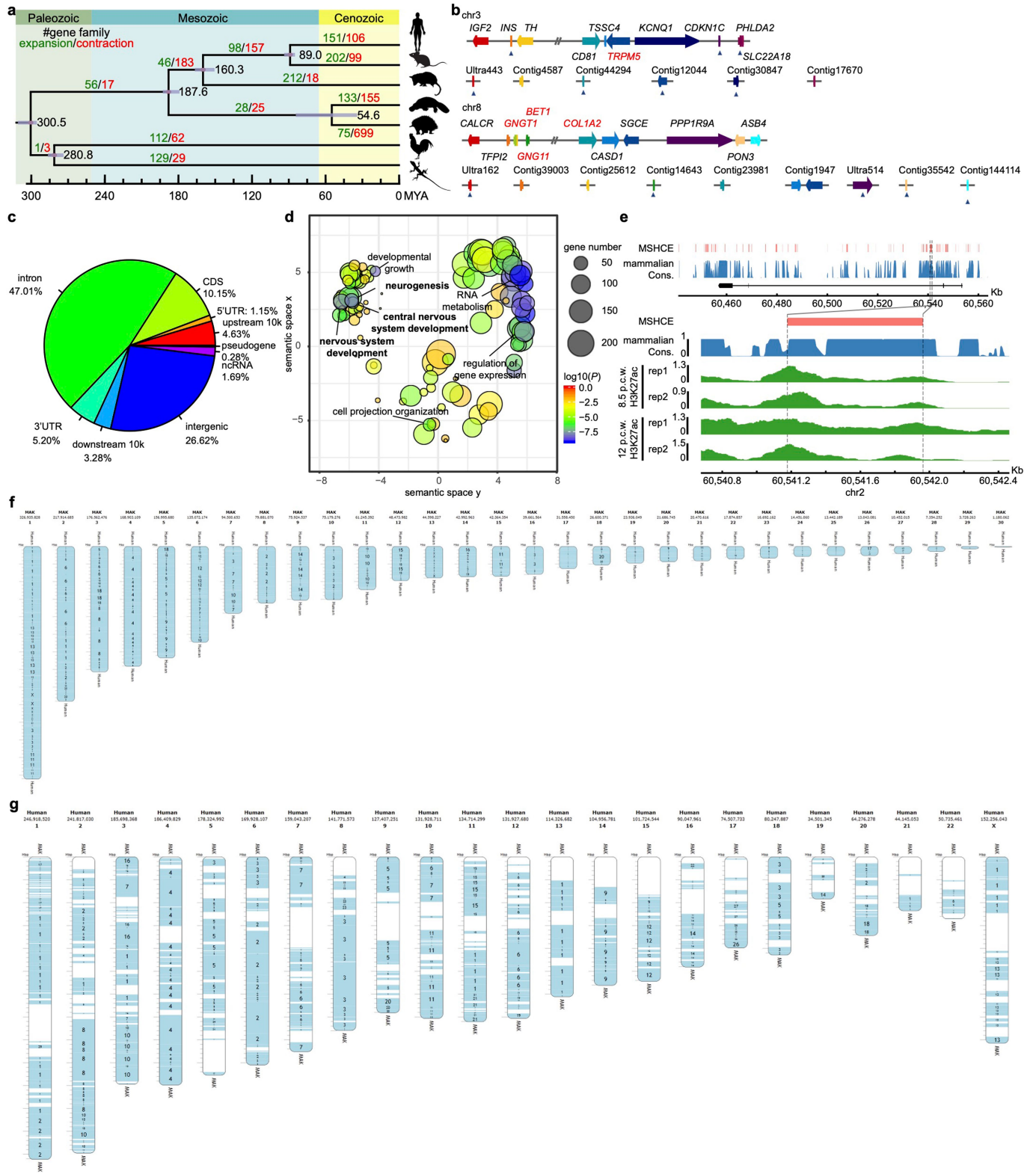


Extended Data Fig. 1 | See next page for caption.

# Article

**Extended Data Fig. 1 | Platypus genome assembly and evaluation. a, b,** Hi-C two-dimensional juicebox maps of mOrnAna1 before (**a**) and after (**b**) manual assembly curation. The grey lines depict scaffold boundaries. The off-diagonal matches between scaffolds indicate potential missed joins, whereas 'empty' areas within scaffold boundaries indicate misjoins. The gEVAL-supported manual assembly curation led to a notably improved arrangement with >96% of the assembly sequence inside chromosome-scale scaffolds. **c, d,** The Super\_Scaffold\_40 was misassigned to chromosome 15 in OANAS but FISH on metaphase spreads from platypus fibroblasts map it to chromosome 13. **c,** Co-hybridization of the BAC of chromosome 15 (green, top arrow) and Super\_Scaffold\_40 probe (red, bottom arrow) showing an absence of co-localization (14 nuclei scored, 2 independent experiments). Inset, interphase example. **d,** Co-hybridization of the BAC of chromosome 13 (green) and Super\_Scaffold\_40 probe (red) showing co-localization (arrows, 40 nuclei scored, 5 independent experiments). Scale bars, 10  $\mu\text{m}$ . **e,** An example of scaffold chromosome misassignment in OANAS. Female-to-male (F/M) depth ratio, normalized female depth and normalized male depth along OANAS chromosome 14 in 5-kb non-overlapping windows. Depth ratio, normalized female depth and normalized male depth all suggest that OANAS chromosome 14 should be an X-borne rather than autosomal sequence. **f, g,** Normalized depth distribution of redundant sequences and one-to-one sequences in male (**f**) and female (**g**). Redundant sequences (red) in OANAS are probably assembly artefacts due to heterozygotes of the sequenced individual of OANAS, and are therefore featured with  $0.5\times$  normalized depth in OANAS but  $1\times$  normalized

depth in mOrnAna1 in both male and female. One-to-one sequences in OANAS (black) have  $1\times$  normalized depth in both OANAS and mOrnAna1 in reads that are mapped from both sexes as expected. Each dot represents one mapping region between OANAS and mOrnAna1 by Mashmap, and the normalized depth values of each dot are calculated as the mean depth across the mapping region in OANAS and mOrnAna1. The small peak in one-to-one sequence density plot in the male indicates candidate X-linked sequences. **h,** Example redundant sequences Contig40802, Contig44497 and Contig35847 in OANAS that could be interpreted as false duplications. Dot plot is generated between the target region of mOrnAna1 chromosome 1 and OANAS Contig1255, Contig40802, Contig44497 and Contig35847 by FlexiDot. Candidate redundant sequences are those mapped to the same region in mOrnAna1 chromosome 1, highlighted by dashed lines in the dot plot and grey in the normalized depth plot. Normalized male and female read depths along each sequence are calculated in 500-bp windows, and plotted along each sequence. Although the normalized depth is always around 1 in the region of mOrnAna1 chromosome 1, normalized depth drops half in Contig40802, Contig44497, Contig35847 and the aligned regions in Contig1255, indicating that Contig40802, Contig44497 and Contig35847 are probably redundant sequences in OANAS. **i, j,** Examples of gene annotation artefacts in OANAS: *CIT* (**i**) and *PBRMI* (**j**) have been fragmented into multiple small artificial genes in OANAS (purple) but have now been fully recovered in mOrnAna1 (orange). Orthologous human genes (grey) are also shown to indicate that the mOrnAna1 rather than OANAS annotation has a similar gene structure to that of the human genes.

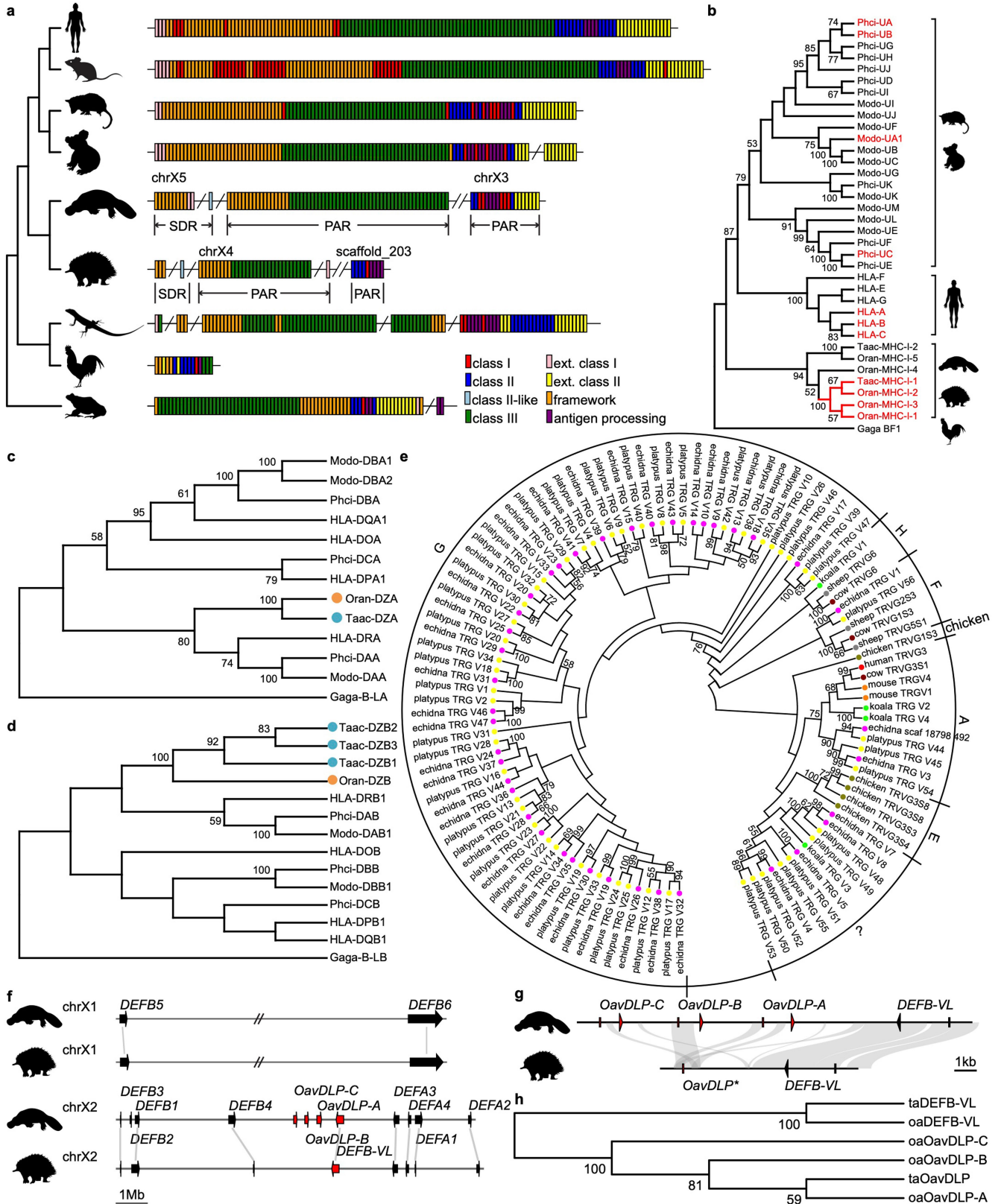


Extended Data Fig. 2 | See next page for caption.

# Article

**Extended Data Fig. 2 | Mammalian genome evolution. a**, Phylogenetic tree constructed using fourfold degenerate sites from 7,946 one-to-one orthologues among seven representative species (human, mouse, opossum, platypus, echidna, chicken and green lizard). The fossil time calibration of the nodes marked by circles were obtained from a previously published study<sup>67</sup>. The numbers of gene families that have undergone significant (Viterbi  $P < 0.05$ ) lineage-specific expansions (green) and contractions (red) are marked on each branch. Exact  $P$  values are available in Supplementary Table 29. No multiple-testing correction was applied. **b**, Examples of some imprinting gene clusters improved in mOrnAna1 compared to OANA5. The first line of each synteny plot represents mOrnAna1 and the second line represents OANA5. Names of genes that have been found to be imprinted in human and mouse are highlighted in black, and non-imprinting genes in red. Fragmented genes with alignment rate lower than 70% are marked by triangles. The double slash represents the intermediate region longer than 100 kb. **c**, Distribution of MSHCEs on genomic elements. **d**, Enriched GO terms in the top-300 MSHCE-associated genes.  $P$  values of enrichment are calculated using a  $\chi^2$  test, and FDRs are computed to

adjust for multiple testing. GO terms are clustered based on semantic similarity. GO terms related to nervous system development are highlighted in bold. **e**, A case of one MSHCE in *BCL11A* that overlaps with the enhancer signals inferred from H3K27ac ChIP-seq experiments at 8.5 and 12 weeks after conception (p.c.w.). **f**, Evolution highway comparative chromosome browser visualization of reconstructed MACs at a 500-kb resolution. Blocks overlaid on each MAC represent human syntenic fragments. Numbers within blocks indicate the homologous human chromosome. **g**, Evolution highway comparative chromosome browser visualization of the human genome at a 500-kb resolution, with each block overlaid on each human chromosome representing putative chromosome fragments of the ancestral mammalian genome. Numbers within blocks depict the ancestral mammal chromosome numbers. Silhouettes of the human and opossum are from <https://www.flaticon.com/>. The silhouette of the platypus is created by S. Werning and is reproduced under a Creative Commons Attribution 3.0 Unported licence (<http://creativecommons.org/licenses/by/3.0/>).



Extended Data Fig. 3 | See next page for caption.

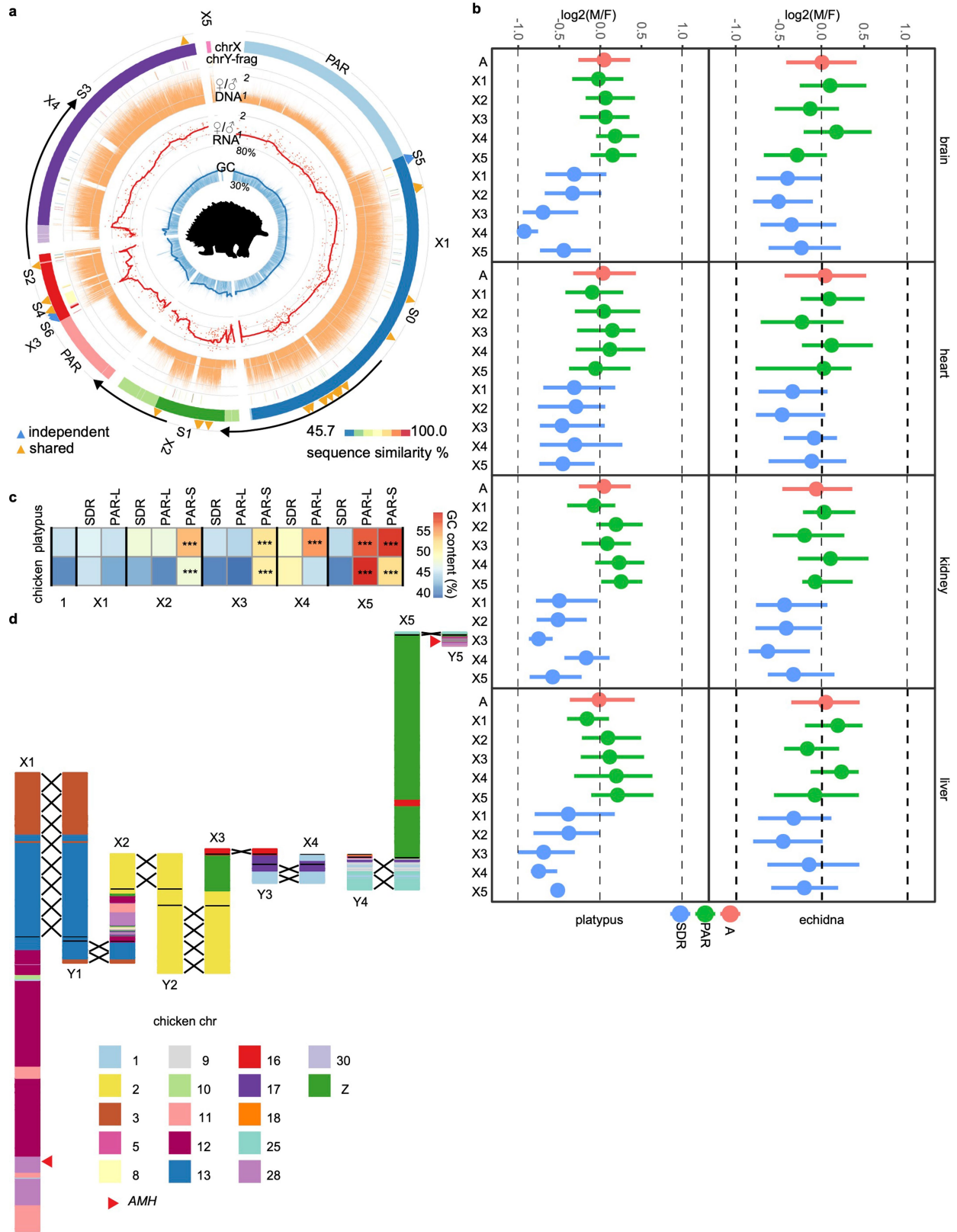


# Article

## Extended Data Fig. 3 | Evolution of immune gene family in monotremes.

**a**, MHC genes in platypus and echidna are located on two different chromosomes, but the classical class I and II genes involved in antigen presentation are located within a single cluster in each genome. **b**, Phylogenetic relationship of class I genes in representative mammals and chicken. Classical class I genes (red) in monotremes exhibit high similarity, which is rarely observed in other species. Only bootstrap values with >50% support are shown. **c, d**, Phylogenetic relationship of MHC class II alpha (**c**) and beta (**d**) genes. Genes with prefix 'HLA', 'Modo', 'Phci', 'Oran', 'Taac' and 'Gaga' indicate genes in human, opossum, koala, platypus, echidna and chicken, respectively. Only bootstrap values with >50% support are shown. **e**, Phylogenetic relationship among putative functional V<sub>γ</sub> sequences from platypus (yellow), echidna (purple), koala (green), mouse (orange), human (red), sheep (grey), cow (dark red) and chicken (dark yellow). Groups according to a previous study<sup>90</sup> are displayed around the outside of the tree, with the putative marsupial–monotreme-specific group denoted by a '?'. Only

bootstrap values with greater than 50% support are shown. **f**, Synteny conservation of beta-defensin genes in monotremes and loss of functional venom defensins in echidna. Venom defensins (OavDLP genes) and venom-like defensin (DEFB-VL genes) are shown in red. Only putative functional defensins are shown. **g**, Putative OavDLP loss in echidna. OavDLP genes and *DEFB-VL* each contain two exons (indicated by a box and triangle) in platypus. Both exons of platypus *DEFB-VL* can be mapped to echidna chromosome X2. A single platypus OavDLP exon can be mapped to echidna chromosome X2 while the second exons cannot. Grey links indicate platypus–echidna LASTZ alignment. **h**, Phylogenetic relationship of *DEFB-VL* and OavDLP genes suggested that ancestral monotremes had all three OavDLP genes but that echidna has lost the two of them (*OavDLP-B* and *OavDLP-C*). Branch length is not shown. ta, echidna; oa, platypus. Silhouettes of the human, opossum, koala and frog are from <https://www.flaticon.com/>. The silhouette of the platypus is created by S. Werning and is reproduced under a Creative Commons Attribution 3.0 Unported licence (<http://creativecommons.org/licenses/by/3.0/>).



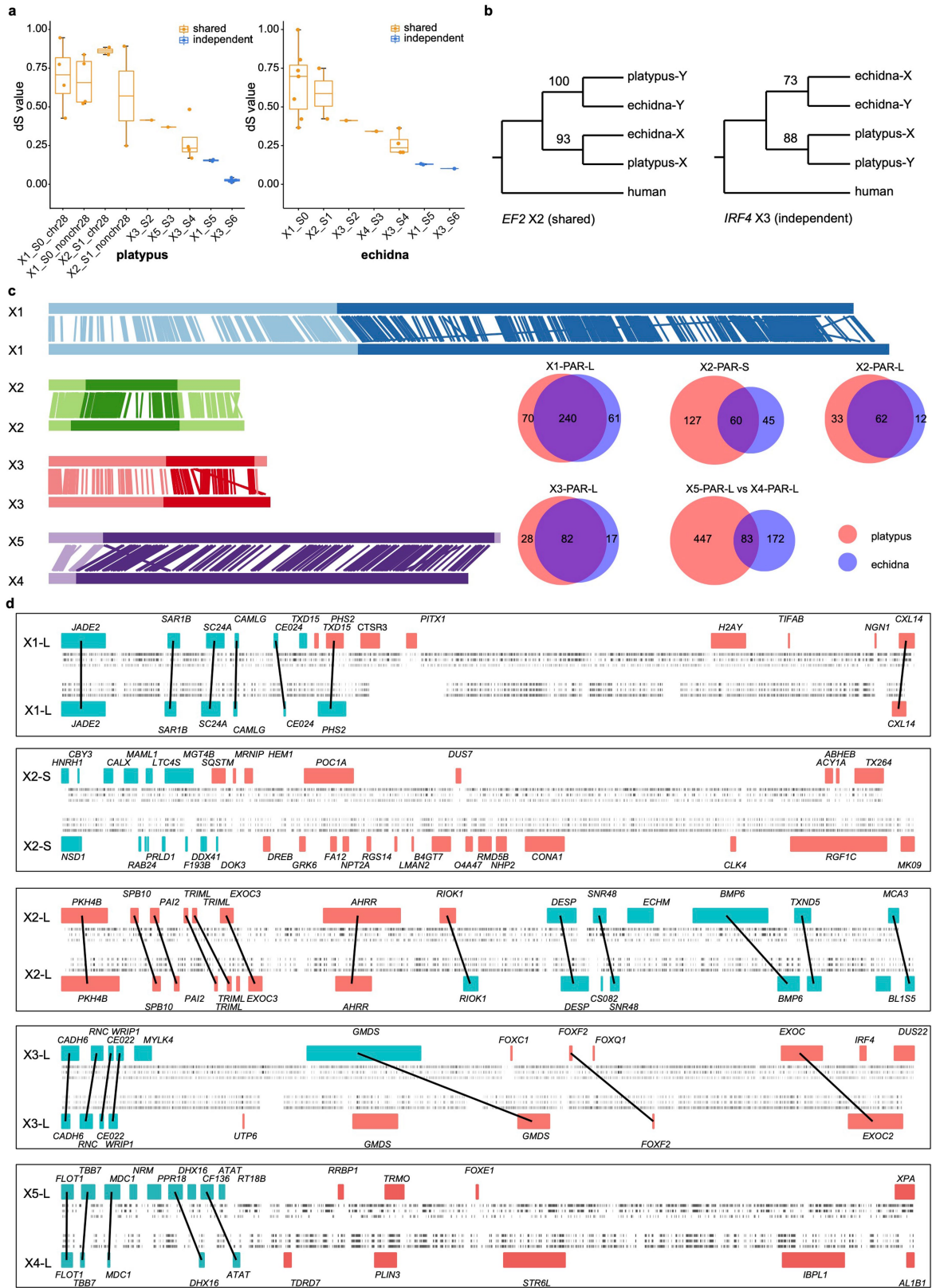
Extended Data Fig. 4 | See next page for caption.

# Article

## Extended Data Fig. 4 | Genomic composition of monotreme sex chromosomes.

**a**, Composition of the echidna sex chromosomes. The circos plot (from outer to inner rings) shows: X chromosomes with PARs shown as light colours and SDRs as dark colours; assembled Y chromosome fragments showing the colour-scaled sequence similarity levels with homologous X chromosomes; normalized F/M ratios of Illumina DNA-sequencing depth in non-overlapping 5-kb windows; F/M expression ratios (each red dot is one gene) of adult kidney and smoothed expression ratio trend; and GC content in non-overlapping 2-kb windows. In addition, Y-linked fragments with a similar level of sequence divergence from the X chromosome indicate a pattern of evolutionary strata. As expected, F/M DNA depth ratio is centred at 1 at PARs, but is around 2 at SDRs. Some PARs show significantly higher GC content than the regions that suppressed recombination between X and Y. **b**, Partial dosage compensation in monotremes. The four point range plots show  $\log_2$ -transformed values of the male-to-female (M/F) expression ratio in the brain, kidney, heart and liver of platypus and echidna. As expected,  $\log_2$ -transformed

values of the M/F expression ratio is close to 0 for genes on autosomes (A) and PARs, whereas for genes on SDRs, the expression is female-biased in all tissues, which suggests that monotremes have partial dosage compensation. Whiskers indicate the 25–75th percentiles and circles are the median value. **c**, Some PARs show significantly higher GC content than SDRs. For platypus, some PARs (X2-PAR-S, X3-PAR-S, X4-PAR-L, X5-PAR-S and X5-PAR-L (where -S is the shorter PAR of the chromosome and -L the longer PAR)) show significantly ( $P < 0.01$ ) higher GC content (1-kb non-overlapping windows) than the SDRs of the same chromosome, which are labelled as asterisks in the heat map. We also checked their orthologous sequences in chicken, as a proxy for the ancestral status before the chromosome became a sex chromosome, and found similarly higher GC content in the orthologous region of PARs than those of SDRs in chicken.  $***P < 0.01$  (all  $P < 2.2 \times 10^{-16}$ ), one-sided Wilcoxon rank-sum test. **d**, Atlas of orthologous chicken fragments along each platypus sex chromosome. The PARs between the platypus X and Y chromosomes are indicated by crosses. We also labelled the position of the putative sex-determining gene *AMH*.

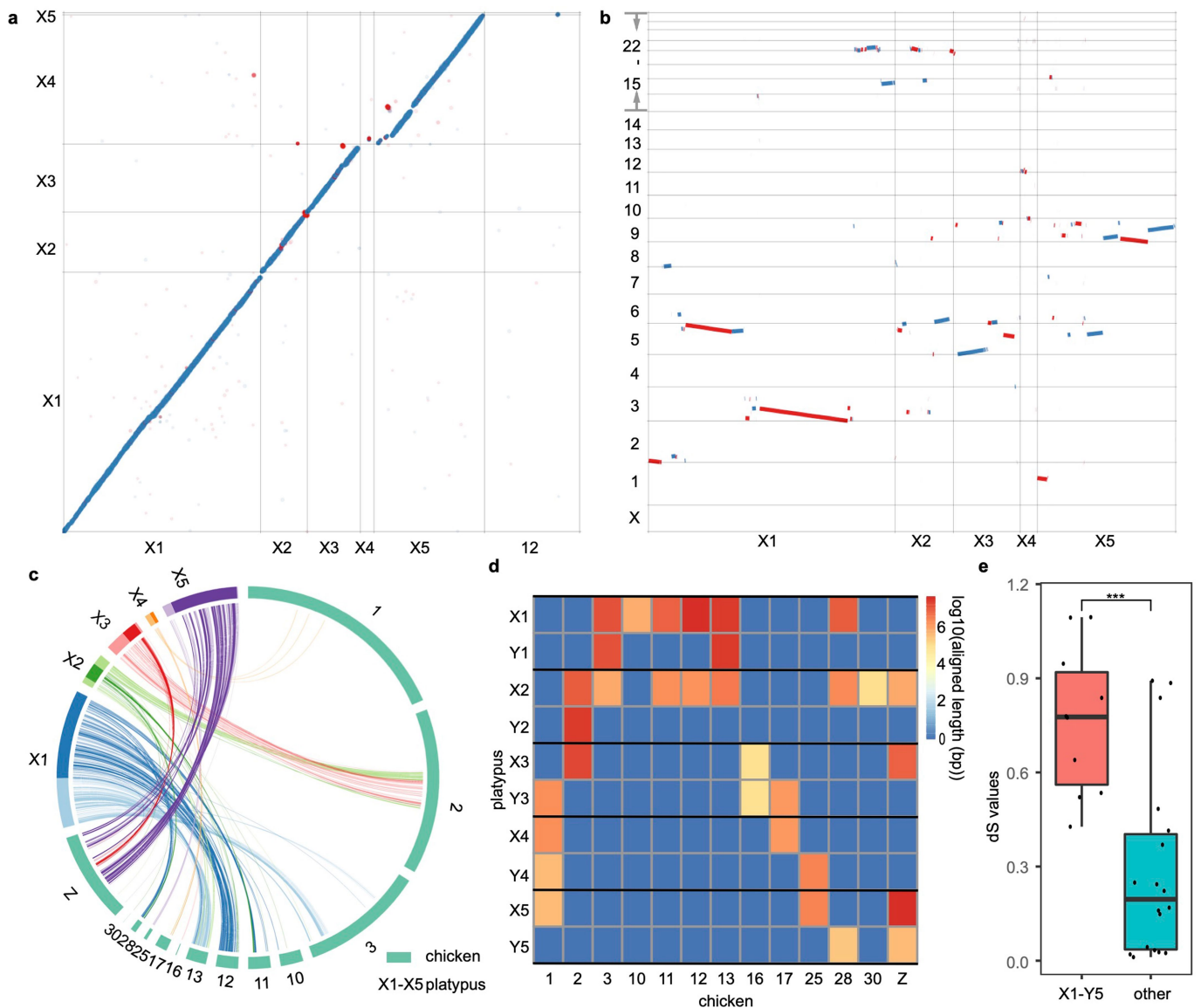


Extended Data Fig. 5 | See next page for caption.

# Article

**Extended Data Fig. 5 | Evolution of PARs after the platypus and echidna divergence.** **a**, The distribution of pairwise dS values of platypus and echidna sex chromosomes. In both platypus and echidna, gametologue pairs in the X1 S0 region (Fig. 2), which is largely homologous to chicken chromosome 28, have a higher dS value than those of any other sex-linked regions. This suggests that X1 S0 is the oldest evolutionary stratum. Therefore, we also show platypus genes of X2 with an orthologue on chicken chromosome 28 separately from others (X1\_S0\_chr28). Following the order of dS values of different chromosome regions, we inferred the time order of formation of evolutionary strata, called S0–S6. For platypus,  $n = 5, 5, 2, 2, 1, 1, 4, 2$  and 6 XY gametologue pairs are plotted, from left to right. For echidna,  $n = 7, 2, 1, 1, 4, 2$  and 1 XY gametologue pairs are plotted, from left to right. Box plots show median, quartiles (boxes) and range (whiskers). **b**, Phylogenetic tree examples of gametologues that evolved in the common ancestor of monotremes (*EF2* in X2) and independently in two monotreme species (*IRF4* in X3). **c**, Alignments of platypus and echidna X chromosomes (PAR, light colours; SDR, dark colours; the top chromosomes are from platypus and the bottom chromosomes are

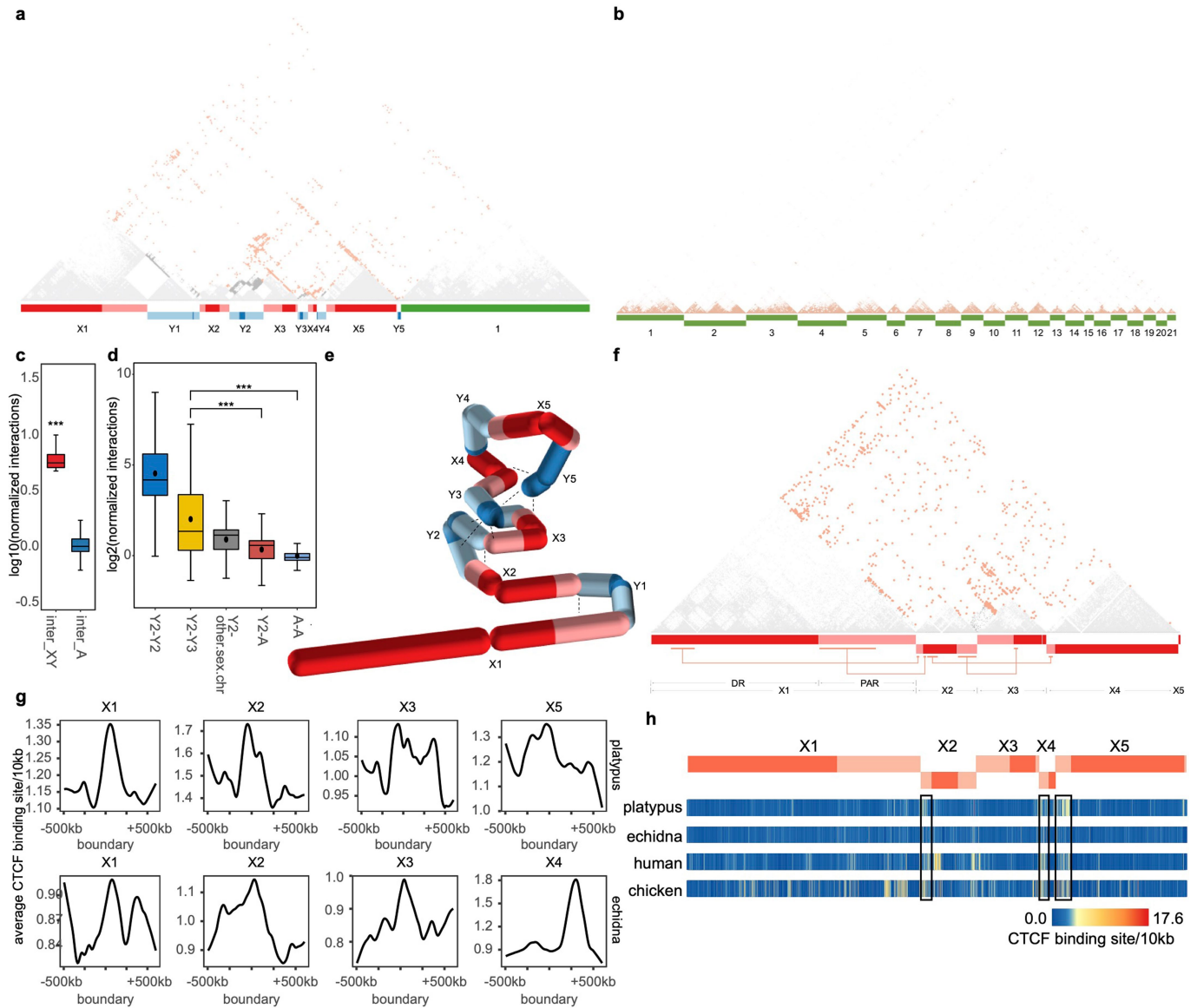
from echidna) were used to infer X2-PAR-S and X5-PAR-L of platypus evolved independently from echidna after their divergence, given their different lengths. This is supported by the Venn diagrams of PAR genes between platypus and echidna, in which most genes are not shared within independently evolved PARs. **d**, Alignments of PAR–SDR boundaries between platypus and echidna. Alignments of genes ( $\pm 1$  Mb around the boundaries) support independent evolution of X2-PAR-S and X5-PAR-L in platypus and echidna, as most of their genes are not homologous at the PAR–SDR boundaries (blue, PAR genes; red, SDR genes; platypus, top chromosome, echidna, bottom chromosome). We used lines to connect the genes of the two species, whenever they are orthologous to each other. For each X chromosome, we also labelled their repeat information. Six repeat tracks between each X pair are shown, from top to bottom: the overall repeat content of platypus; LINE/L2 elements of platypus; SINE/MIR elements of platypus; SINE/MIR elements of echidna; LINE/L2 elements of echidna and overall repeat content of echidna. We did not find obvious repeat enrichment at PAR–SDR boundaries, as shown previously in cow<sup>91</sup>.



**Extended Data Fig. 6 | Sex chromosome evolution in monotremes.**

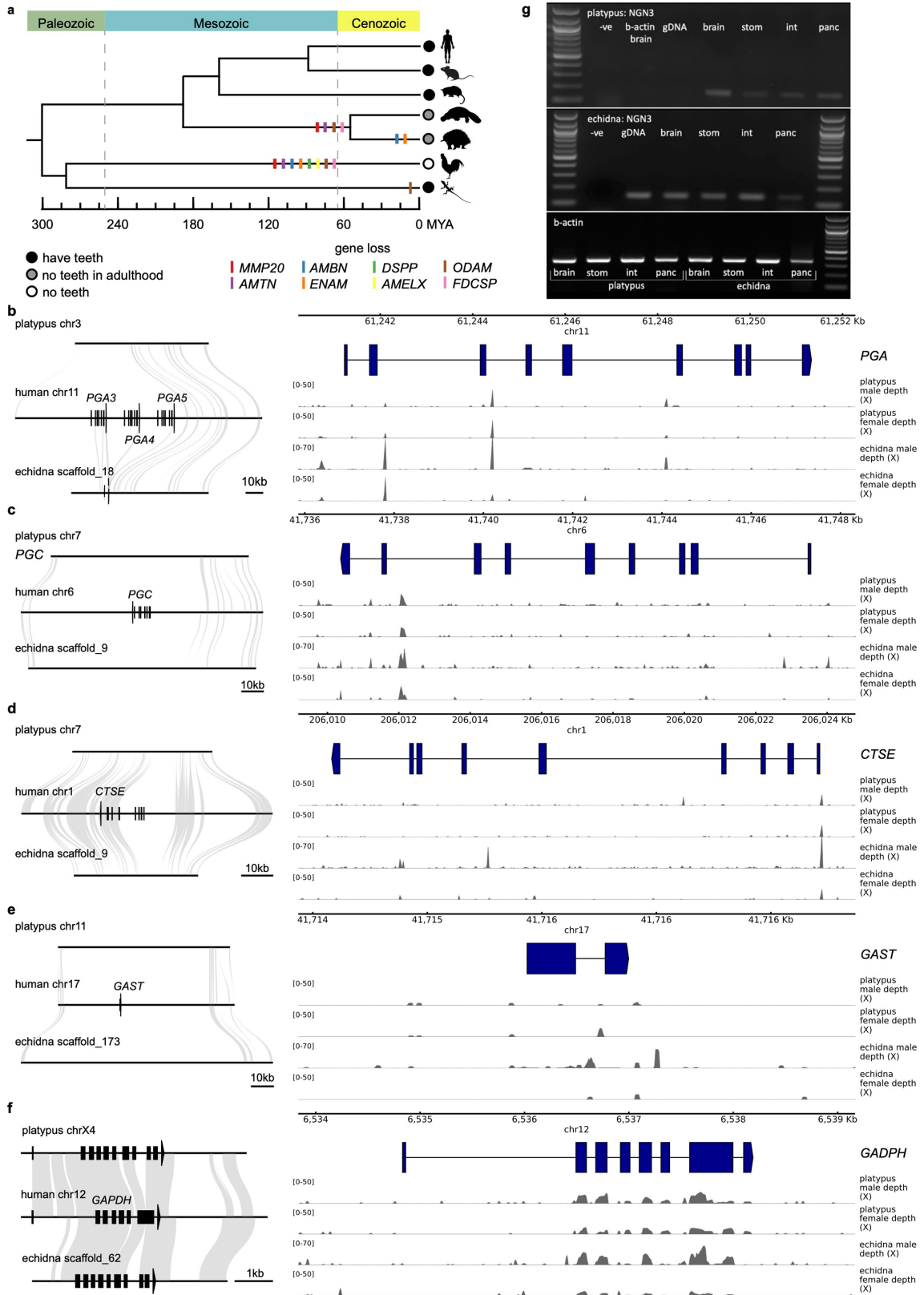
**a**, Mummerplot showing homology between platypus (x axis) and echidna (y axis) X chromosomes. Blue lines: forward alignment; red lines: reverse alignment. For echidna, X1, X2 and X3 are homologous to platypus X1, X2 and X3, respectively. Echidna X4 is homologous to platypus X5. And for echidna X5, it is not homologous to any platypus sex chromosome, and instead it is homologous to platypus chromosome 12. **b**, Homology between platypus X

chromosomes (x axis) and human chromosomes (y axis). **c**, Homologous relationships between platypus sex chromosomes and chicken. **d**, Alignment between platypus and chicken showing the alternating pairing pattern of the platypus sex chromosome chain. **e**, X/Y pairwise dS comparison between gametologues on X1-Y5 pair ( $n=18$ ) and other sex chromosome pairs ( $n=10$ ). Box plots show median, quartiles (boxes) and range (whiskers). \*\*\* $P < 0.001$  ( $P=0.0002954$ ), one-sided Wilcoxon rank-sum test.



**Extended Data Fig. 7 | Chromatin conformation of monotreme sex chromosomes.** **a**, Hi-C interactions between platypus sex chromosomes, with chromosome 1 shown as control. **a**, **b**, There are unexpected interchromosomal interactions (shown in red) between platypus sex chromosomes detected by Hi-C data (**a**), whereas most interactions are within the same chromosomes (shown in red in **b**) for the other chromosomes (**b**). **c**, The Hi-C interchromosomal interactions among platypus sex chromosome (inter\_XY,  $n = 2,711,100$ -kb windows) is significantly higher than that among autosomes (inter\_A,  $n = 14,342,930$  100-kb window). Box plots show median, quartiles (boxes) and range (whiskers).  $***P < 0.0001$  ( $P < 2.2 \times 10^{-16}$ ), one-sided Wilcoxon rank-sum test. **d**, The interaction strength is higher between Y2 and Y3 than the interaction strengths between Y2 and other chromosomes.  $n = 1,002, 228, 5,025, 67,313$  and  $6,904,867$  100-kb windows are shown in Y2-Y2, Y2-Y3, Y2-other.sex.chr, Y2-A and A-A, respectively. Box plots show median, quartiles

(boxes) and range (whiskers).  $***P < 0.0001$  ( $P < 2.2 \times 10^{-16}$ ), one-sided Wilcoxon rank-sum test. **e**, Inferred three-dimensional structure of the platypus sex chromosome system. X chromosomes are shown in red and Y chromosomes in blue, with PARs in light colour. Interchromosomal interactions inferred from Hi-C are shown by dashed lines. **f**, Hi-C interactions reveal unexpected interchromosomal interactions between the echidna sex chromosomes. **g**, Putative CTCF-binding sites are enriched at TAD boundaries in platypus and echidna sex chromosomes. For each X chromosome of platypus, we calculated their putative CTCF-binding-site density per 10 kb and plotted them along the  $\pm 500$  kb of TAD boundaries. Platypus X4 and echidna X5 are not shown because less than 10 TAD boundaries are detected. **h**, Putative CTCF-binding-site density plot showing its enrichment among the homologous regions of platypus, echidna, human and chicken.



Extended Data Fig. 8 | See next page for caption.



# Article

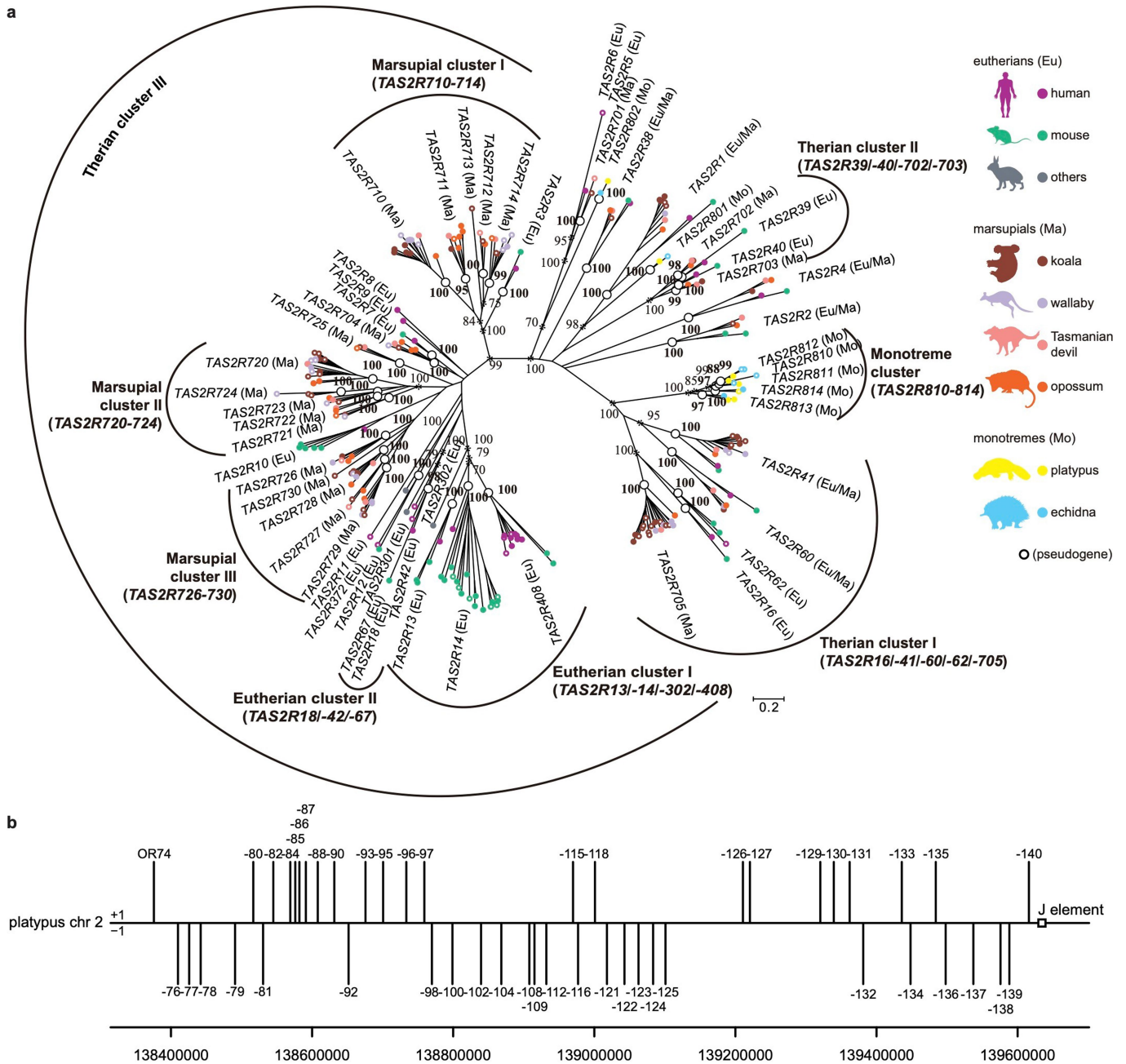
## Extended Data Fig. 8 | Loss of dietary-related genes in monotremes.

**a**, Tooth-related gene loss in representative mammals and reptiles.

**b–f**, Potential loss of digestion-related genes in both monotremes shown by whole-genome alignment and read mapping. In each panel there are three lines in the synteny plot, representing the orthologous region of the genes in platypus, human and echidna from top to bottom, respectively. Grey links indicate human–platypus and human–echidna LASTZ alignments. Each rectangle or triangle represents an exon. Fragmented genes are marked by dashed lines. Illumina reads of platypus and echidna are aligned to the platypus or human genome (Ensembl release 87) and the flanking region of each gene is visualized by pyGenomicTrack. *GAPDH* region is also plotted as a control.

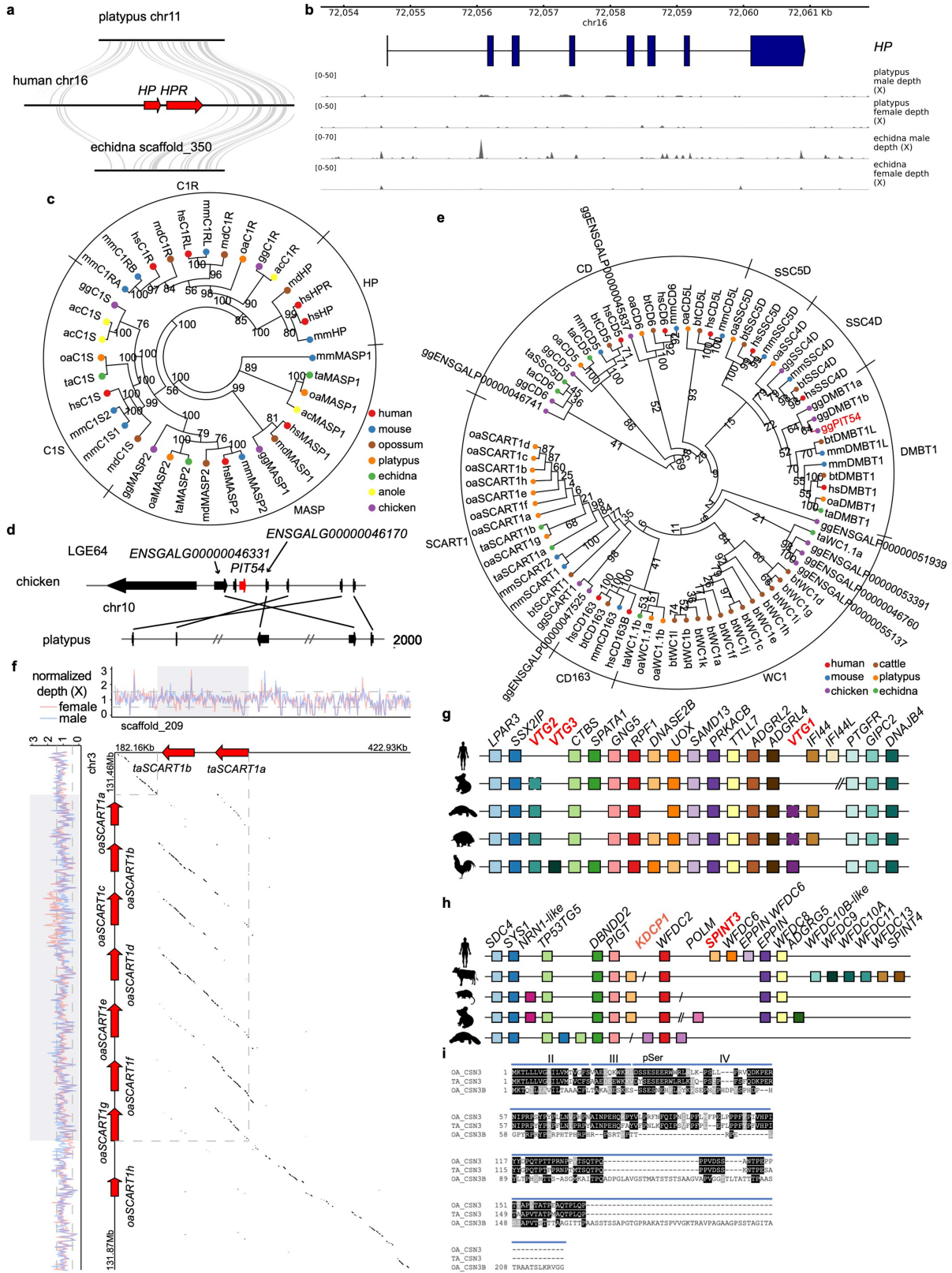
**g**, RT-PCR expression analysis shows expression of *NGN3* in brain, stomach, intestine and pancreas of both platypus and echidna. These results are similar

to other mammals. This, together with sequencing results, shows that *NGN3* in monotremes is present and is likely to be functioning normally. *NGN3*, *NGN3* primers;  $\beta$ -actin,  $\beta$ -actin primers; -ve, negative control, no template; gDNA, genomic DNA template; brain, brain cDNA template; stom, stomach cDNA template; int, intestine cDNA template; panc, pancreas cDNA template. Lanes 1 (top), 1, 8 (middle) and 9 (bottom) are a 100-bp DNA ladder: 1,517, 1,200, 1,000, 900, 800, 700, 600, 500/517, 400, 300, 200 and 100 bp. Expected sizes of PCR products for *NGN3* in platypus is 157 bp and for echidna 145 bp, and the PCR product for the  $\beta$ -actin genomic region is 597 bp and cDNA is 348 bp. Silhouettes of human and opossum are from <https://www.flaticon.com/>. The silhouette of the platypus is created by S. Werning and is reproduced under a Creative Commons Attribution 3.0 Unported licence (<http://creativecommons.org/licenses/by/3.0/>).



**Extended Data Fig. 9 | Taste-receptor evolution and olfactory-receptor organization in monotremes.** **a**, Maximum-likelihood mammalian-wide gene tree of the bitter taste receptors (TAS2R genes). There are 28 eutherian (Eu), 27 marsupial (Ma) and 7 monotreme-specific (Mo) orthologous gene groups (supported by  $\geq 95\%$  bootstrap values), where the nodes of orthologous gene group clades are indicated by white open circles. Bootstrap values of  $\geq 70\%$  in the nodes connecting orthologous gene group clades are indicated by asterisks. There are 3 therian (I, II and III), 2 eutherian (I and II), 3 marsupial (I, II and III) and one monotreme-specific clusters in which massive expansion events occurred in the common ancestor of each taxon after the split from its

previous ancestors. **b**, Genomic organization of the intact class I olfactory receptor (OR) cluster spanning over 1.2 Mb on platypus chromosome 2 (138,375,798–139,616,970 bp). The vertical lines indicate the 48 intact class I OR genes. The white open box indicates the J element, a presumable *cis*-regulatory element (enhancer) for the mammalian class I OR cluster (chromosome 2: 139,639,465–139,639,907 bp). Silhouettes of human, opossum and koala are from <https://www.flaticon.com/>. Silhouettes of the platypus and Tasmanian devil are created by S. Werning and are reproduced under a Creative Commons Attribution 3.0 Unported licence (<http://creativecommons.org/licenses/by/3.0/>).



Extended Data Fig. 10 | See next page for caption.

**Extended Data Fig. 10 | Genomic features related to haemoglobin clearance and reproduction in monotremes.** **a, b**, Confirmation of *HP* absence in monotremes by whole-genome alignment (**a**) and read mapping (**b**). Grey links indicate human–platypus and human–echidna LASTZ alignments. Illumina reads of platypus and echidna are aligned to the human genome (Ensembl release 87) and coding regions of *HP* are visualized by pyGenomicTrack. Limited coverage is found at the exons of *HP*, suggesting the absence of *HP* in monotremes. **c**, Phylogenetic tree of HP and related proteases across different species using the maximum-likelihood method. Node IDs are in format of 'species geneID'. Branch length is not shown here. **d**, Gene synteny plot of the *PITS4* region between chicken and platypus. Echidna is not shown in the figure as the flanking orthologues of *PITS4* are on different scaffolds, preventing us from determining the presence of the gene by synteny. **e**, Phylogenetic tree of members of the group B scavenger receptor cysteine-rich family across different species using the neighbour-joining method. Gene IDs are formatted as 'species geneID'. Branch length is not shown here. **f**, Confirmation of *SCART1* number difference by dot plot and mapping depth of *SCART1* orthologous regions between platypus and echidna. The region of the *SCART1* cluster in platypus is plotted along the x axis while the sequence of echidna is plotted along the y axis. Lines in dot plot are visualized according to LASTZ alignment between the two species. Normalized male and female read depths along each sequence is calculated in 500-bp windows, and plotted along each sequence. Normalized depth of both sexes, especially those in the shading region, is centred at 1 along both species, confirming the *SCART1* number difference between the two species is true and is not due to assembly issues. **g**, Synteny conservation of vitellogenin genes. Synteny conservation of the region

surrounding the vitellogenin (*VTG*) genes *VTG1*, *VTG2* and *VTG3*. Pseudogenes are marked by a dashed outline. Monotremes have pseudogene *VTG1*, functional *VTG2* and no *VTG3*; and there is a pseudogene *VTG2* in koala. Syntenic maps are shown for human (*Homo sapiens*), koala (*Phascogaleolemur cinereus*), chicken (*Gallus gallus*), platypus (*O. anatinus*) and echidna (*T. aculeatus*). Koala scaffold 1, NW\_018343984.1; koala scaffold 2, NW\_018344134.1. Gene distances are not to scale. **h**, Synteny conservation of regions containing *SPINT3*. Synteny conservation of the region surrounding serine peptidase inhibitor, Kunitz-type, 3 (*SPINT3*). No copy of *SPINT3* is detected in platypus but many of the other flanking genes in the region are conserved. Other members with a WFDC domain are detected including two Kunitz-domain members that did not align to any known gene (labelled *KDCPI*). Syntenic maps are reported for human (*H. sapiens*), cow (*B. taurus*), grey short-tailed opossum (*Monodelphis domestica*), koala (*P. cinereus*) and platypus (*O. anatinus*). Koala scaffold 1, NW\_018343967.1; koala scaffold 2, NW\_018344098.1. Gene distances are not to scale. **i**, Casein 3 (CSN3) protein sequence alignment in monotremes. All three CSN3 proteins identified in the monotremes have the classic five-exon structure of CSN3 with the untranslated exons I and IV (not shown), the signal peptide in exon II, a small exon III coding for 11 residues, a pSER cluster (S\*\*) at the 5' end of exon IV and a relatively large P/Q-rich exon IV. OA, *O. anatinus* (platypus); TA, *T. aculeatus* (short-beaked echidna). Silhouettes of human, opossum and koala are from <https://www.flaticon.com/>. The silhouette of the platypus is created by S. Werning and is reproduced under a Creative Commons Attribution 3.0 Unported licence (<http://creativecommons.org/licenses/by/3.0/>).

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

**Data collection** Fluorescence in situ hybridization (FISH) data were collected Nikon Ti Microscope using NIS-Elements software (supplied with a Nikon Eclipse-Ti inverted epifluorescent microscope) and processed with ImageJ.

**Data analysis** Common bioinformatic and statistical analysis software packages were used, including: smrtanalysis (v5.1.0.26412), FALCON (v5.1.1), FALCON-Unzip (v1.0.2), Scaff10X (v2.1 git 4.28.2018), BioNano Solve (v3.2.1\_04122018), SALSA (v2.0), Longranger (v2.2.2), FreeBayes (v1.2.0), BCFtools (v1.8), gEVAL (<https://vgp-geval.sanger.ac.uk>), Platanus (v1.2.1), BWA (v0.7.12), Samtools (v1.2), BLAST (v2.2.26), NCBI BLAST (v2.2.31), minimap2 (v2.13), Mashmap (v2.0), RepeatMasker (v4.0.6), trf (v4.07), ProteinMasker (v4.0.6), RepeatModeler (v1.0.8), GeneWise (v2.4.1), Augustus (v3.0.3), HISAT2 (v2.0.4), stringTie (v1.2.3), lprscan (v5.16-55.0), BEDTools (v2.26.0), LASTZ (v1.04.00), BUSCO (v3.0.2), RaxML (v8.2.4), PAML (v4.7), MULTIZ (v11.2), PHAST (v1.5), orthoMCL (v2.0.9), CAFE (v4.2), REVIGO (<http://revigo.irb.hr/>), inferCARs (2006-Jun-16), ANGES (v1.01), MCScanX (08-05-2012), GRIMM (v2.1), DESeq (v1.28.0), HiC-Pro (v2.10.0), ggplot2 (v3.2.1), HiCEXplorer (v3.0), MEME (v4.12.0), EMBOSS (v6.5.7), hmmer (v3.1b2), genscan (v1.0), MEGA (X, 7, v5.5.2), exonerate (v2.2.0), pyGenomeTracks (v2.1), MAFFT (v6.857b), MUSCLE (v3.8.31), TreeBest (v1.9.2), ggtree (v1.16.6), TOPCONS (v2.0), CD-HIT (v4.6.8), FATE (v2.7.0), GHOSTZ (v1.0.2), Clustal Omega (v1.2.4), Clustal W (v2.1), ImageJ (v2.0.0), NIS-Elements software (v4.20.00). Custom scripts are open source and available on GitHub at <https://github.com/ZhangLabSZ/MonotremeGenome>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Platypus whole genome shotgun project has been deposited at GenBank under the project accession PRJNA489114 and PRJNA489115, CNSA (<https://db.cngb.org/cnsa/>) of CNGBdb with accession CNP0000130 and GenomeArk at [https://vgp.github.io/genomeark-curated-assembly/Ornithorhynchus\\_anatinus/](https://vgp.github.io/genomeark-curated-assembly/Ornithorhynchus_anatinus/). Echidna whole genome shotgun project has been deposited at GenBank under the project accession PRJNA576333, CNSA of CNGBdb with accession CNP0000697 and GenomeArk at [https://vgp.github.io/genomeark/Tachyglossus\\_aculeatus/](https://vgp.github.io/genomeark/Tachyglossus_aculeatus/). Echidna RNA-seq has been deposited at GenBank under the project accession PRJNA591380 and CNSA of CNGBdb with accession CNP0000779.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The determination of sample size for genome sequencing is not applied in this study. Sample size for FISH experiments were determined according to the previously published work in the field (Ling et al. 2006), but was also limited by the availability of the material. Bioinformatic analyses were performed with all available data.
Data exclusions	None
Replication	FISH experiments showing that Super_Scaffold_40 did not co-localize with the chr15 BAC were carried out in two independent experiments and co-localization of Super_Scaffold_40 with the chr13 BAC was carried out in five independent experiments. Potential interchromosomal interactions indicated by Hi-C data were validated in three independent FISH experiments for each Y2-Y3, Y2-X1 and Y2-WSB1(chr17) pair.
Randomization	Randomization was not performed in this study. Only one cell line was used (established from a single individual) and all available data was collected in all experiments to fulfill the criteria for the sample size. In strata analysis gametologues were grouped according to the pairwise dS and phylogeny. Grouping in interchromosomal interaction analysis was based on sequence chromosome assignment.
Blinding	Our study was not an intervention study and therefore blinding was not required.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Fibroblast primary cell culture derived from the adult male platypus individual number 1 from NSW field expedition in New South Wales, year 2008.
---------------------	---

Authentication	None of the cell lines used were authenticated.
Mycoplasma contamination	Cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	No commonly misidentified lines were used.

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Study did not involve laboratory animals.
Wild animals	Study did not involve wild animals.
Field-collected samples	Platypus and echidna male individuals were collected during breeding season at the upper Barnard river area (New South Wales) and euthanised with an intraperitoneal injection of pentobarbitone sodium (Nembutal) at a dose of 0.1 mg/g body weight. Tissues were snap frozen in liquid N2 immediately. Emale01 was collected from Melbourne Zoo and provided by San Diego Zoo Global, with an estimated date of birth of 12/31/67.
Ethics oversight	Platypus and echidna samples (Pmale08, Pmale09, Emale12) were collected under the AEC permits S-49-2006, S-032-2008 and S-2011-146 at the Upper Barnard River (New South Wales, Australia) during the breeding season. Emale01 was collected under San Diego Zoo Global IACUC approval 18-024, and is vouchered at the San Diego Natural History Museum.

Note that full information on the approval of the study protocol must also be provided in the manuscript.