

nature



**CARBON DIOXIDE
CAPTURE**

Cagey ZIFs for
emissions control

'FLIPPING' ENZYMES

Action mechanism of
HIV's Achilles' heel

SCIENCE AND MUSIC

First in a new series

THE PLATYPUS GENOME

Sequence analysis reveals clues
to early mammalian evolution

NATUREJOBS
Up Canada way



Genome analysis of the platypus reveals unique signatures of evolution

A list of authors and their affiliations appears at the end of the paper

We present a draft genome sequence of the platypus, *Ornithorhynchus anatinus*. This monotreme exhibits a fascinating combination of reptilian and mammalian characters. For example, platypuses have a coat of fur adapted to an aquatic lifestyle; platypus females lactate, yet lay eggs; and males are equipped with venom similar to that of reptiles. Analysis of the first monotreme genome aligned these features with genetic innovations. We find that reptile and platypus venom proteins have been co-opted independently from the same gene families; milk protein genes are conserved despite platypuses laying eggs; and immune gene family expansions are directly related to platypus biology. Expansions of protein, non-protein-coding RNA and microRNA families, as well as repeat elements, are identified. Sequencing of this genome now provides a valuable resource for deep mammalian comparative analyses, as well as for monotreme biology and conservation.

The platypus (*Ornithorhynchus anatinus*) has always elicited excitement and controversy in the zoological world¹. Some initially considered it to be a true mammal despite its duck-bill and webbed feet. The platypus was placed with the echidnas into a new taxon called the Monotremata (meaning 'single hole' because of their common external opening for urogenital and digestive systems). Traditionally, the Monotremata are considered to belong to the mammalian subclass Prototheria, which diverged from the therapsid line that led to the Theria and subsequently split into the marsupials (Marsupialia) and eutherians (Placentalia). The divergence of monotremes and therians falls into the large gap in the amniote phylogeny between the eutherian radiation about 90 million years (Myr) ago and the divergence of mammals from the sauropsid lineage around 315 Myr ago (Fig. 1). Estimates of the monotreme–theria divergence time range between 160 and 210 Myr ago; here we will use 166 Myr ago, recently estimated from fossil and molecular data².

The most extraordinary and controversial aspect of platypus biology was initially whether or not they lay eggs like birds and reptiles. In 1884, William Caldwell's concise telegram to the British Association announced "Monotremes oviparous, ovum meroblastic", not holoblastic as in the other two mammalian groups^{3,4}. The egg is laid in an earthen nesting burrow after about 21 days and hatches 11 days later^{5,6}. For about 4 months, when most organ systems differentiate, the young depend on milk sucked directly from the abdominal skin, as females lack nipples. Platypus milk changes in protein composition during lactation (as it does in marsupials, but not in most eutherians⁵). The anatomy of the monotreme reproductive system reflects its reptilian origins, but shows features typical of mammals⁷, as well as unique specialized characteristics. Spermatozoa are filiform, like those of birds and reptiles, but, uniquely among amniotes, form bundles of 100 during passage through the epididymis. Chromosomes are arranged in defined order in sperm⁸ as they are in therians, but not birds⁹. The testes synthesize testosterone and dihydrotestosterone, as in therians, but there is no scrotum and testes are abdominal¹⁰.

Other special features of the platypus are its gastrointestinal system, neuroanatomy (electro-reception) and a venom delivery system, unique among mammals¹¹. Platypus is an obligate aquatic feeder that relies on its thick pelage to maintain its low (31–32 °C) body temperature during feeding in often icy waters. With its eyes, ears and nostrils closed while foraging underwater, it uses an

electro-sensory system in the bill to help locate aquatic invertebrates and other prey^{12,13}. Interestingly, adult monotremes lack teeth.

The platypus genome, as well as the animal, is an amalgam of ancestral reptilian and derived mammalian characteristics. The platypus karyotype comprises 52 chromosomes in both sexes^{14,15}, with a few large and many small chromosomes, reminiscent of reptilian macro- and microchromosomes. Platypuses have multiple sex chromosomes with some homology to the bird Z chromosome¹⁶. Males have five X and five Y chromosomes, which form a chain at meiosis and segregate into 5X and 5Y sperm^{17,18}. Sex determination and sex chromosome dosage compensation remain unclear.

Platypuses live in the waterways of eastern and southern Australia, including Tasmania. Its secretive lifestyle hampers understanding of its population dynamics and the social and family structure. Platypuses are still relatively common in the wild, but were recently reclassified as 'vulnerable' because of their reliance on an aquatic environment that is under stress from climate change and degradation by human activities. Water quality, erosion, destruction of habitat and food resources, and disease now threaten populations. Because the platypus has rarely bred in captivity and is the last of a long line of ornithorhynchid monotremes, their continued survival is of great importance. Here we describe the platypus genome sequence and compare it to the genomes of other mammals, and of the chicken.

Sequencing and assembly

All sequencing libraries were prepared from DNA of a single female platypus (Glennie; Glenrock Station, New South Wales, Australia) and were sequenced using established whole-genome shotgun (WGS) methods¹⁹. A draft assembly was produced from ~6× coverage of whole-genome plasmid, fosmid and bacterial artificial chromosome (BAC) reads (Supplementary Table 1) using the assembly program PCAP²⁰ (Supplementary Notes 1). A BAC-based physical map was developed in parallel with the sequence assembly and subsequently integrated with the WGS assembly to provide the primary means of scaffolding the assembly into larger ordered and oriented groupings (ultracontigs; Supplementary Notes 2 and 3 and Supplementary Table 2). Because there were no platypus linkage maps available, we used fluorescent *in situ* hybridization (FISH) to localize a subset of the sequence scaffolds to chromosomes following the agreed nomenclature²¹. Of the 1.84 gigabases (Gb) of assembled sequence, 437 megabases (Mb) were ordered and oriented along 20 of

the platypus chromosomes. We analysed numerous metrics of assembly quality (Supplementary Notes 4–11) and we conclude that despite the adverse contiguity, the existing platypus assembly, given its structural and nucleotide accuracy, provides a reasonable substrate for the analyses presented here.

Non-protein-coding genes

In general, the platypus genome contains fewer computationally predicted non-protein-coding (nc)RNAs (1,220 cases excluded high repetitive small nucleolar RNA (snoRNA) copies; see below) than do other mammalian species (for example, human with 4,421 Rfam hits), similar to observations in chicken¹⁹ (655 Rfam-based ncRNAs). This is probably because of the extensive retrotransposition of ncRNAs in therian mammals and the apparent lack of L1-mediated retrotransposition in chicken and platypus. The exception to this is the platypus family of snoRNAs, which is markedly expanded (~2,000 matches to the Rfam covariant models) compared to that for therian mammals (~200). snoRNAs are involved in RNA modifications, in particular of ribosomal RNA, and are often located in introns of protein-coding genes²². Our investigations revealed a novel short-interspersed-element (SINE)-like, snoRNA-related retrotransposon—which we have labelled snoRTEs—that has duplicated in platypus to ~40,000 full-length or truncated copies. It is retrotransposed by means of retrotransposon-like non-LTR (long terminal repeat) transposable elements (RTE) as opposed to the L1-mediated transposition mechanism in therians²³. We constructed a complementary DNA library of small, ncRNAs and identified 371 consensus sequences of small RNAs that included 166 snoRNAs²³ (Supplementary Table 3). Ninety-nine of these cloned snoRNAs are found in paralogous families, and 21 of them belong to the snoRTE class. The presence of both the structural requirements

known to be important in snoRNA function²⁴ and evidence of their expression are consistent with these snoRTE elements being functional in the platypus. Similar to other unrelated ncRNAs that have proliferated in therian mammals (for example, 7SL RNA-derived primate *Alu* elements, tRNA-derived rodent identifier (ID) elements), this recent SINE-like expansion is probably due to chance events. However, given the RNA modification activity of snoRNAs, and our increasing awareness of the cellular importance of RNA molecules, it might be that some of the retrotranspositionally duplicated RNAs were exapted into new functions in this species.

Other small RNAs. Overall, we found commonalities with small RNA (sRNA) pathways of other mammals, but also features that are unique to monotremes. Components of the RNA interference machinery are conserved in platypus, including elements of biogenesis pathways (Dicer and Drosha) and RNA-interference effector complexes (argonaute proteins; Supplementary Table 4). Of 20,924,799 platypus and echidna sRNA reads derived from liver, kidney, brain, lung, heart and testis, 67% could be assigned to known microRNA (miRNA) families. Established patterns of miRNA expression were generally recapitulated in monotremes.

To determine the conservation patterns of miRNAs in platypus, we identified platypus miRNAs sharing at least 16-nucleotide identity with miRNAs in eutherian mammals (mouse/human) and chicken. Although most conserved miRNAs were identified across these vertebrate lineages (137 miRNAs), 10 miRNAs were shared only with eutherians (mouse/human) and 4 only with chicken (Fig. 2a). miRNAs can be classified into families based on identity of the functional 'seed' region at position 2–8 of the mature miRNA strand. We identified miRNA families that were shared between platypus and eutherians but not chicken (40 families), or between platypus and chicken but not eutherians (8 families), suggesting that for some miRNAs only

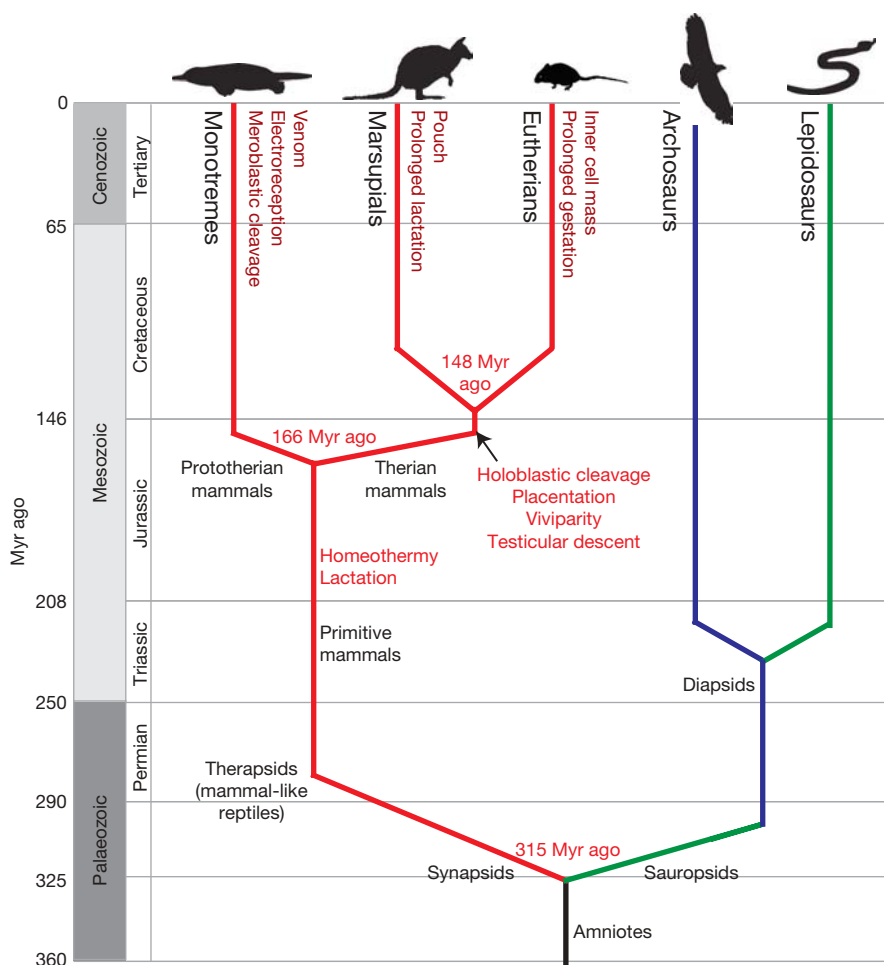


Figure 1 | Emergence of traits along the mammalian lineage. Amniotes split into the sauropsids (leading to birds and reptiles) and synapsids (leading to mammal-like reptiles). These small early mammals developed hair, homeothermy and lactation (red lines). Monotremes diverged from the therian mammal lineage ~166 Myr ago² and developed a unique suite of characters (dark-red text). Therian mammals with common characters split into marsupials and eutherians around 148 Myr ago² (dark-red text). Geological eras and periods with relative times (Myr ago) are indicated on the left. Mammal lineages are in red; diapsid reptiles, shown as archosaurs (birds, crocodilians and dinosaurs), are in blue; and lepidosaurs (snakes, lizards and relatives) are in green.

the seed region may have been selectively conserved (Fig. 2a). Conserved miRNAs tended to be more robustly expressed in the platypus tissues analysed than lineage-restricted miRNAs (Fig. 2b).

To identify miRNAs unique to monotremes we used a heuristic search that identifies miRNA candidates in deep-sequencing data sets²⁵. This method predicted 183 novel miRNAs in platypus and echidna (Fig. 2a). Notably, 92 of these lay in 9 large clusters, on platypus chromosome X1 and contigs 1754, 7160, 7359, 8388, 11344, 22847, 198872 and 191065. Physical mapping confirmed that at least five of these contigs are linked to the long arm of chromosome X1 (ref. 25). These abundantly expressed clusters were sequenced almost exclusively from platypus and echidna testis (Fig. 2b). The expansion of this unique miRNA class and its expression domain suggest possible roles in monotreme reproductive biology²⁵.

Piwi-interacting RNAs (piRNAs) associate with a germline-expressed clade of argonaute proteins, known as Piwi²⁶, and have a role in transposon silencing and genome methylation²⁶. Monotreme piRNAs bear strong structural similarity to those in eutherians. They are ~29 nucleotides in length and arise from large testis-specific genomic clusters with distinct genomic strand asymmetry, often with a typical 'bidirectional' organization. We identified 50 major platypus piRNA clusters as well as numerous smaller clusters²⁵. In contrast to piRNAs in mouse, platypus piRNAs are repeat-rich and bear strong signatures of active transposon defence.

Gene evolution

We set out to define the protein-coding gene content of platypus to illuminate both the specific biology of the monotreme clade and for comparisons to eutherians and marsupials, or to chicken, the representative sauropsid. Protein-coding genes were predicted using the established Ensembl pipeline²⁷ suitably modified for platypus (Supplementary Notes 14), with a greater emphasis placed on similarity matches to mammalian genes. Overall this resulted in 18,527 protein-coding genes being predicted from the current platypus assembly. The number of platypus protein-coding genes thus is similar to estimates (18,600–20,800) for human and opossum^{28,29}.

We were interested first in identifying platypus genes that contribute most to core biological functions that are conserved across the mammals. These will typically be 'simple' 1:1 orthologues, genes that have remained as single copies without duplication or deletion in platypus, in Eutheria (specifically, in dog, human and mouse) and in opossum, a representative marsupial. Subsequently, we considered genes that have been duplicated or deleted in the monotreme lineage, or that have been lost in eutherian and/or marsupial lineages. Such genes are proposed to contribute most to the lineage-specific biological functions that distinguish individual mammals³⁰. These

studies required the use of an outgroup species, here chicken, a representative of the sauropsids.

As expected, the majority of platypus genes (82%; 15,312 out of 18,596) have orthologues in these five other amniotes (Supplementary Table 5). The remaining 'orphan' genes are expected to primarily reflect rapidly evolving genes, for which no other homologues are discernible, erroneous predictions, and true lineage-specific genes that have been lost in each of the other five species under consideration. Simple 1:1 orthologues, which have been conserved without duplication, deletion or non-functionalization across the five mammalian species, were greatly enriched in housekeeping functions, such as metabolism, DNA replication and mRNA splicing (Supplementary Table 6).

We then identified evolutionary lineages that experienced the most stringent purifying selection. The mouse terminal lineage exhibited a significantly higher degree of purifying selection (the ratio of amino acid replacement to silent substitution rates, $d_N/d_S = 0.105$, $P < 0.001$) than dog, opossum and chicken terminal branches (values of 0.123–0.128); human and platypus terminal lineages showed significantly reduced purifying selection (both 0.132, $P < 0.03$). These values probably reflect the increased efficiency of purifying selection in populations of larger effective size, such as that of mouse³¹. We find that at least one nucleotide substitution has occurred, on average, in synonymous sites of platypus and human orthologues since their last common ancestor (Supplementary Notes 17 and Supplementary Fig. 1). This means that most neutral sequence cannot be aligned accurately between monotreme and eutherian genomes.

Next, we determined the genetic distance of echidna (*Tachyglossus aculeatus*) from platypus. The median d_S value of 0.125 for the orthologues of echidna and platypus, when compared to the value for the monotreme lineage, predicts that platypus and echidna last shared a common ancestor 21.2 Myr ago. Although similar to previous estimates³², this value seems to be at odds with fossil evidence, perhaps owing to relatively recent reductions of mutational rates in the monotreme lineage³³.

Monotreme biology

We next investigated whether the ancestral reptilian characters of monotremes are reflected in the set of genes that have been retained in platypus, sauropsids and other vertebrates from outside of the amniote clade (such as frogs and fish), but have been lost from eutherian and marsupial lineages (Fig. 1). These ancestral, sauropsid-like, characters of platypus include oviparity (egg laying) and the outward appearances of its spermatozoa and retina. Simultaneously, we sought genetic evidence within the platypus genome both for characteristics peculiar to monotremes, such as venom production and

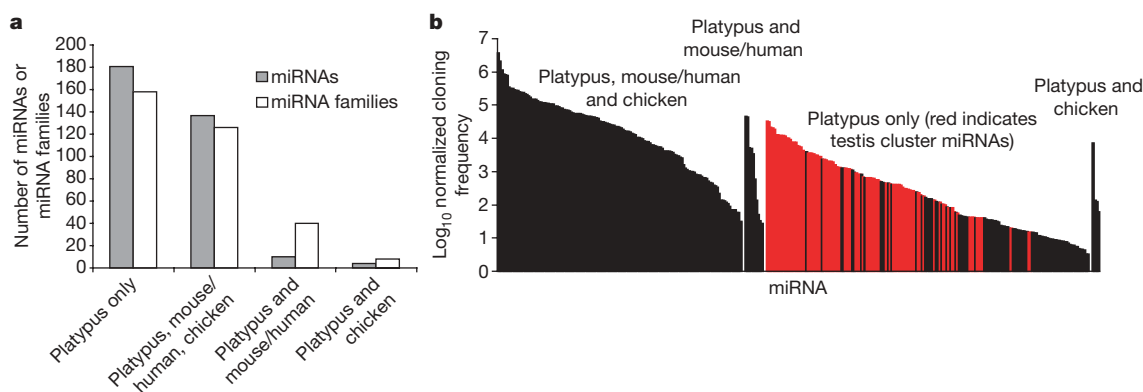


Figure 2 | Platypus miRNAs. **a**, Platypus has miRNAs shared with eutherians and chickens, and a set that is platypus-specific. miRNAs cloned from six platypus tissues were assigned to families based on seed conservation. Platypus miRNAs and families were divided into classes (indicated) based on their conservation patterns with eutherian mammals

(mouse/human) and with chicken. **b**, Expression of platypus miRNAs. The cloning frequency of each platypus mature miRNA sequenced more than once is represented by a vertical bar and clustered by conservation pattern. miRNAs from a set of monotreme-specific miRNA clusters that are expressed in testis are shaded in red.

electro-reception, and for characteristics unique to mammals, in particular lactation. By investigating platypus homologues of genes already known to be involved in specific physiological processes (see Methods), we highlight those platypus genes for which evolution exemplifies the ancestral or derived physiological characters of monotremes.

Chemoreception. The semi-aquatic platypus was expected to sense its terrestrial, but not aquatic, environment by detecting airborne odorants using olfactory receptors and vomeronasal receptors (types 1 and 2: V1Rs, V2Rs). Nevertheless large numbers of odorant receptor, V1R and V2R homologues (approximately 700, 950 and 80, respectively) are apparent in the platypus genome assembly, although for each family only a minority lack frame disruptions (approximately 333, 270 and 15, respectively)³⁴. Many of these platypus genes and pseudogenes are monophyletic, having arisen by duplication in the 166 Myr since the last common ancestor of monotremes and therians. Although mouse and rat genomes possess greater numbers of odorant receptors and V2Rs than the platypus genome^{35,36}, the platypus repertoire of V1Rs, showing undisrupted reading frames, is the largest yet seen, 50% more than for mouse (Fig. 3b). This is particularly noteworthy as the *Anolis carolinensis* lizard (sequence data used with the permission of the Broad Institute) and the chicken¹⁹ seem to possess no such receptors. The large expansion of the platypus V1R gene family might reflect sensory adaptations for pheromonal communication or, more generally, for the detection of water-soluble, non-volatile odorants, during underwater foraging.

The platypus odorant receptor gene repertoire is roughly one-half as large as those in other mammals³⁷. Nevertheless, platypus odorant receptors fall into class, family and subfamily structures that are well represented from across the mammals, with a few notable exceptions such as family 14 (Fig. 3a). Together with the finding that lizard contains only ~200 odorant receptor genes and pseudogenes, this indicates that the platypus olfactory repertoire is, as expected, more akin to other mammals than it is to sauropsids.

Eggs. Fertilization in the platypus exhibits both sauropsid and therian characteristics. Platypus ova are small (4 mm diameter) relative to comparably sized reptiles and birds, and eggs hatch at an early stage of development so that most growth of the embryo and infant is dependent on lactation, as in marsupials. Like all mammals and many other amniotes, when fertilization occurs the ovum is invested with a zona pellucida. The platypus genome encodes each of the four proteins of the human zona pellucida³⁸, as well as two ZPAX genes (Table 1) that previously were observed only in birds, amphibians and fish. The aspartyl-protease nothepsin is present in platypus, but has been lost from marsupial and eutherian genomes (Table 1). In zebrafish, this gene is specifically expressed in the liver of females under the action of oestrogens, and accumulates in the ovary³⁹. These are the same characteristics as of the vitellogenins, indicating that nothepsin may be involved in processing vitellogenin or other egg-yolk proteins. We find that platypus has retained a single vitellogenin gene and pseudogene, whereas sauropsids such as chicken have three and the viviparous marsupials and eutherians have none.

Spermatozoa. Orthologues of many of the eutherian sperm membrane proteins related to fertilization⁴⁰ are present in platypus (and marsupial) genomes. These include the genes for a number of putative zona pellucida receptors and proteins implicated in sperm-oolemma fusion. Testis-specific proteases, which in eutherians participate in degradation of the zona pellucida during fertilization, are all absent from the platypus genome assembly.

Monotreme spermatozoa undergo some post-testicular maturational changes, including the acquisition of progressive motility, loss of cytoplasmic droplets and aggregation of single spermatozoa into bundles during passage through the epididymis¹¹. Nevertheless, maturational changes in the sperm surface that are both unique and essential in other mammals for fertilization of the ovum have yet to be identified. Also, the epididymis of monotremes is not highly

adapted for sperm storage as in most marsupial and eutherian mammals. Consistent with these findings is the absence of platypus genes for the epididymal-specific proteins that have been implicated in sperm maturation and storage in other mammals. The most abundant secreted protein in the platypus epididymis is a lipocalin, the homologues of which are the most secreted proteins in the reptilian epididymis⁴¹. Notably, *ADAM7*, a protease that is secreted in the epididymis of eutherians, has an orthologue in the platypus. This is a bona fide protease with a characteristic Zn²⁺-coordinating sequence HExxH in the platypus, in the opossum and the tree shrew (*Tupaia belangeri*). However, loss of its proteolytic activity is predicted in eutherians⁴² owing to a single point mutation within its active site (E to Q).

Lactation and dentition. Lactation is an ancient reproductive trait whose origin predates the origin of mammals. It has been proposed that early lactation evolved as a water source to protect porous parchment-shelled eggs from desiccation during incubation⁴³ or as a protection against microbial infection. Parchment-shelled egg-laying monotremes also exhibit a more ancestral glandular mammary patch or areola without a nipple that may still possess roles in egg protection. However, in common with all mammals, the milk of monotremes has evolved beyond primitive egg protection into a true milk that is a rich secretion containing sugars, lipids and milk proteins with nutritional, anti-microbial and bioactive functions. In a reflection of this eutherian similarity platypus casein genes are tightly clustered together in the genome, as they are in other mammals, although platypus contains a recently duplicated β -casein gene (Supplementary Fig. 2).

Mammalian casein genes are thought to have originally arisen by duplication of either enamel or ameloblastin⁴⁴, both of which are tooth enamel matrix protein genes that are located adjacent to the casein gene cluster in eutherians and, we find, also in platypus. Adult platypuses, as well as echidnas, lack teeth but the conservation of these enamel protein genes is consistent with the presence of teeth and enamel in the juvenile, as well as the fossil platypuses⁴⁵.

Venom. Only a handful of mammals are venomous, but the male platypus is unique among them in delivering its poison not via a bite but from hind-leg spurs. Despite the obvious difficulties in obtaining

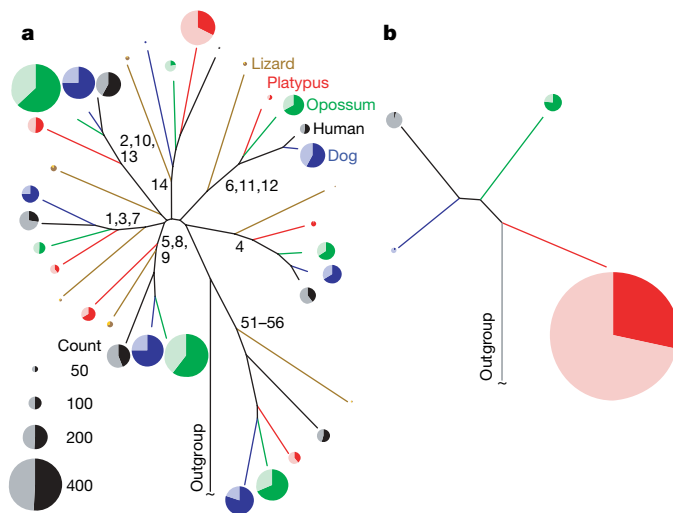


Figure 3 | The platypus chemosensory receptor gene repertoire. **a, b,** The platypus genome contains only few olfactory receptor genes from olfactory receptor families that are greatly expanded among therians (three other mammals and a reptile shown), but many genes in olfactory receptor family 14 (**a**), and relatively numerous vomeronasal type 1 (V1R) receptors (**b**). These schematic phylogenetic trees show relative family sizes and pseudogene contents of different gene families (enumerated beside internal branches) and the V1R repertoire in platypus. Pie charts illustrate the proportions of intact genes (heavily shaded) versus disrupted pseudogenes (lightly shaded).

samples, it is now known that platypus venom is a cocktail of at least 19 different substances⁴⁶ including defensin-like peptides (vDLPs), C-type natriuretic peptide (vCNP) and nerve growth factor (vNGF). When analysed phylogenetically and mapped to the platypus genome assembly, these sequences are revealed to have arisen from local duplications of genes possessing very different functions (Fig. 4). Notably, duplications in each of the β -defensin, C-type natriuretic peptide and nerve growth factor gene families have also occurred independently in reptiles during the evolution of their venom⁴⁷. Convergent evolution has thus clearly occurred during the independent evolution of reptilian and monotreme venom⁴⁸.

Immunity. Although the major organs of the monotreme immune system are similar to those of other mammals⁴⁹, the repertoire of immunity molecules shows some important differences from those of other mammals. In particular, the platypus genome contains at least 214 natural killer receptor genes (Supplementary Notes 18) within the natural killer complex, a far larger number than for human (15 genes⁵⁰), rat (45 genes⁵⁰) or opossum (9 genes⁵¹).

Both platypus and opossum genomes contain gene expansions in the cathelicidin antimicrobial peptide gene family (Supplementary Fig. 3). Among eutherians, primates and rodents have a single cathelicidin gene^{52,53}, whereas sheep and cows have numerous genes that have been duplicated only recently⁵⁴. The expanded repertoire of cathelicidin genes in both marsupials and monotremes may arm their immunologically naive young with a diverse arsenal of innate immune responses. In eutherians, with their increases in length of gestation and advances in development *in utero* of their immune systems, the diversity of antimicrobial peptide genes may have become less critical. The platypus genome also contains an expansion in the macrophage differentiation antigen *CD163* gene family (Supplementary Notes 18).

Genome landscape

First, we analyse the phylogenetic position of platypus and confirm that marsupials and eutherians are more closely related than either is to monotremes (Supplementary Notes 19). We then describe platypus chromosomes and observe some properties of platypus interspersed and tandem repeats. We also discuss a potential relationship between interspersed repeats and genomic imprinting and investigate how the extremely high G + C fraction in platypus affects the strong association seen in eutherians between CpG islands and gene promoters.

Platypus chromosomes. Platypus chromosomes provide clues to the relationship between mammal and reptile chromosomes, and to the origins of mammal sex chromosomes and dosage compensation. Our analysis provides further insight with the following findings: the 52 platypus chromosomes show no correlation between the position of orthologous genes on the small platypus chromosomes and chicken microchromosomes; for the unique 5X chromosomes of platypus we reveal considerable sequence alignment similarity to chicken Z and no orthologous gene alignments to human X, implying that the platypus X chromosome evolved directly from a bird-like ancestral reptilian system⁵⁵; and the genes on the five platypus X chromosomes appear to be partially dosage compensated (Supplementary Fig. 5), perhaps parallel to the incomplete dosage compensation recently described in birds⁵⁶.

Repeat elements. About one-half of the platypus genome consists of interspersed repeats derived from transposable elements. The most abundant and still active repeats are (severely truncated) copies of the 5-kb long-interspersed-element (LINE2) and its non-autonomous SINE-companion mammalian-wide interspersed repeat (MIR, Mon-1 in monotremes) that became extinct in marsupials and in eutherians 60–100 Myr ago. We estimate that there are 1.9 and 2.75 million copies of LINE2 and MIR/Mon-1, respectively, in the 2.3-Gb platypus genome. DNA transposons and LTR retroelements are quite

Table 1 | Platypus genes that have been lost from the eutherian lineage

Description	Platypus Ensembl gene	Proposed function
Retinal guanylate cyclase activator 1A	ENSOANG00000012043	In zebrafish, expressed in retina
Enoyl-CoA hydratase/isomerase	ENSOANG00000012890	Involved in fatty acid metabolism
Ferric reductase/cytochrome b561	ENSOANG00000019725	Absorption of dietary iron
Nothepsin, aspartic proteinase	ENSOANG00000005955	Processes egg-yolk proteins
Glutamine synthetase	ENSOANG00000008089	Role in nitrogen metabolism
Vitellogenin II	Contig 10010	Major egg-yolk protein
Cytochrome P450, CYP2-like	ENSOANG00000004537	Toxin degradation
ATP6AP1 paralogue	ENSOANG00000004825	Retinal pigmentation
Organic solute transporter alpha (2 genes)	Ultracontig 462, Contig 159089	Bile acid transport
Neuropeptide Y7 receptor	ENSOANG00000014966	Regulator of food intake
Melatonin receptor 1C	ENSOANG00000011638	Circadian rhythm regulation
Epidermal differentiation-specific proteins (3 genes)	ENSOANG00000005335, ENSOANG00000003767, ENSOANG00000013512	Neural and epidermal differentiation
TRPV7/TRPV8 transient receptor potential cation channels	ENSOANG00000015080, ENSOANG00000015083	Novel epithelial calcium channels
Shortwave-sensitive-2 (SWS2) opsin gene	Ultracontig 401	Cone visual pigment
Opsin 5 paralogue	ENSOANG00000009478	Light-sensitive receptor
Indigoidine synthase A	Contig 29616	Pigmentation
ZPAX, egg envelope glycoprotein	ENSOANG00000007840, ENSOANG00000002187	Egg envelope protein
Galanin receptor	ENSOANG00000020606	Neuropeptide receptor
Kainate-binding protein	ENSOANG00000007006	Glutamate receptor
Anti-dorsalizing morphogenetic protein	ENSOANG00000002980	Patterning of the body axis during gastrulation
Retinal genes (2 genes)	ENSOANG00000001054, ENSOANG00000004065	Unknown function
Uteroglobin-like secretoglobins (3 genes)	ENSOANG00000020019, ENSOANG00000022350, ENSOANG00000021122	Unknown function
Testis homeobox C14-like proteins (>2 genes)	ENSOANG00000020069, ENSOANG00000022694	Unknown function
Parvalbumin	ENSOANG00000000764	Muscle function
Slc7a2-prov protein	ENSOANG00000009602	Cationic amino acid transporter
Cystine/glutamate transporter	ENSOANG00000005615	Amino acid transporter
SOUL protein	ENSOANG000000013998	Retina and pineal gland haem protein, oxygen sensing
Twin-pore potassium channel Talk-1-like	ENSOANG00000011839	Potassium channel
Alpha-aspartyl dipeptidase	ENSOANG00000009001	Unknown function
Monovalent cation/H ⁺ antiporter	ENSOANG00000012961	Unknown function; conserved in other metazoa and in yeast

Sequences without Ensembl nomenclature are found in Supplementary Information.

rare in platypus, but there are thousands of copies of an ancient gypsy-class LTR element (all LTR elements previously identified in mammals, birds, or reptiles belong to the retrovirus clade). Overall, the frequency of interspersed repeats (over 2 repeats per kb) is higher than in any previously characterized metazoan genome. Population analysis using LINE2/Mon-1 elements distinguished the Tasmanian population from three other mainland clusters (Supplementary Fig. 4a, b), in good agreement with tree-based analysis, physical proximity and previous knowledge of platypus population relationships⁵⁷.

Cluster analysis of all LINE2 copies revealed a phylogenetic relationship lacking branches, as if a single-locus, fast-evolving gene has steadily spread an exceptional number of pseudogenes over time (Supplementary Fig. 6). This 'master gene' appearance is, to a lesser degree, also observed for LINE1 in eutherians⁵⁸, but not to the same extent for MIR/Mon-1 or other retrotransposons in mammals. The phylogeny of LINE2 and Mon-1 was also supported by a genome-wide transposition-in-transposition (TinT) analysis⁵⁹ (Supplementary Tables 7 and 8). LINE2 density is similar on all chromosomes (Supplementary Fig. 7); it does not correlate with chromosome length (and recombination rate) as the CR1 LINE density does in the chicken genome¹⁹, nor is it higher on sex chromosomes than on autosomes, as LINE1 density is in eutherians (which has led to postulations on a function in dosage compensation)⁶⁰.

We compared microsatellites in the platypus genome with those of representative vertebrates (Supplementary Notes 22). The mean microsatellite coverage of platypus genomic sequences assembled into chromosomes is $2.67 \pm 0.34\%$; significantly lower than all other mammalian genomes sequenced so far and most similar to that observed in chicken (Supplementary Fig. 8). Microsatellites are on average shorter in platypus than in other genomes (Supplementary Table 9), but microsatellite coverage surpasses chicken owing to very long tri- and tetranucleotide repeats (Supplementary Fig. 9). The platypus has a higher proportion of microsatellites with high A+T content, in comparison to the other vertebrates examined, an abundance distribution that has more in common with reptiles than with mammals (Supplementary Fig. 10).

Genomic imprinting. Genomic imprinting is an epigenetic phenomenon that results in monoallelic gene expression. In the vertebrates, imprinting seems to have evolved recently and has only been

confirmed in marsupials and eutherian mammals^{61,62}. The autosomal localization of some imprinted orthologues in platypus is known⁶³. However, we examined the conservation of synteny and the distribution of retrotransposed elements in all orthologous eutherian-imprinted clustered and non-clustered genes in the platypus genome. A representative cluster is shown in Fig. 5 (see also Supplementary Fig. 12).

Clusters that became imprinted in therians (with the exception of the Prader–Willi–Angelman locus⁶⁴) have not been assembled recently and reside in ancient syntenic mammalian groups, although some regions have expanded by mechanisms such as gene duplication or transposition. There were significantly fewer LTR and DNA elements across all platypus orthologous regions relative to eutherian imprinted genes ($P < 0.04$ and 0.04 , respectively), whereas there was a significant increase in the sequences masked by SINEs ($P < 0.03$). The chicken had fewer total repeats and no SINEs or rRNAs. Comparison of all regions in the platypus with the orthologous regions in opossum, mouse, dog and human demonstrates that accumulation of LTR, DNA elements, and simple and low complexity repeats coincides with, and may be a driving force in, the acquisition of imprinting in these regions in therian mammals.

The CpG fraction. The eutherian and chicken genomes generally average around 41% G+C content, although many intervals differ substantially from the average, particularly in humans (Supplementary Notes 23). In contrast, the platypus genome averages 45.5% G+C content and rarely deviates far from the average. The opossum genome averages only 38% G+C content and also has a narrow distribution (Supplementary Fig. 13). The source of the elevated G+C fraction in platypus remains unclear. It is explained only in part by monotreme interspersed repeat elements, as platypus DNA outside of known interspersed repeats is 44.7% G+C. Furthermore, tandem repeats of short DNA motifs (microsatellites) in platypus show an A+T bias, as with other mammals. Recombination-driven biased gene conversion may be a factor, in agreement with what has been shown for eutherians⁶⁵ and marsupials⁶⁶. This is suggested by the observation that the six platypus chromosomes where the currently mapped DNA sequence averages over 45% G+C content (that is, 17, 20, 15, 14, 10 and 11 in order of decreasing G+C fraction) are among the 10 shortest (Supplementary Fig. 14), because short chromosomes have a higher recombination rate⁶⁷. However, a direct test

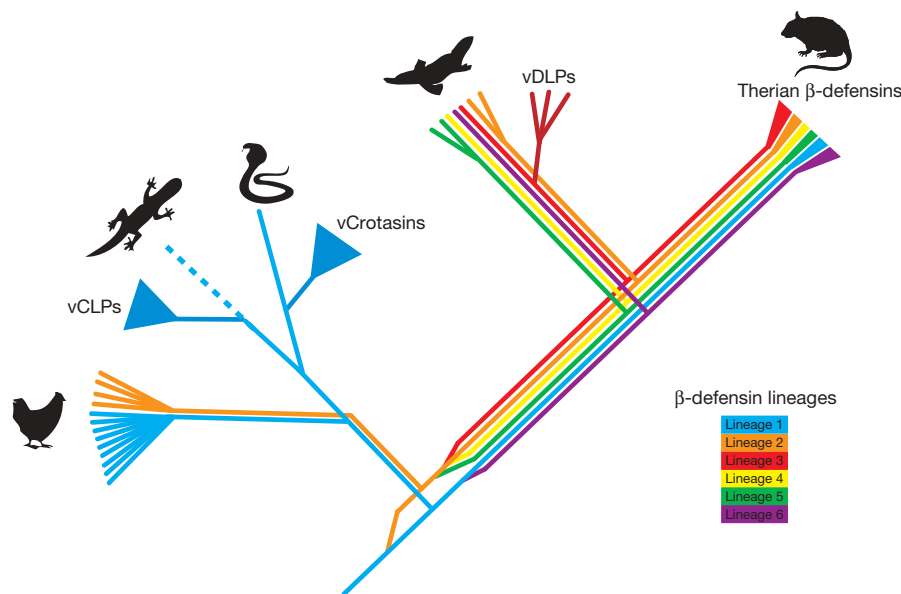


Figure 4 | The evolution of β -defensin peptides in platypus venom gland. The diagram illustrates separate gene duplications in different parts of the phylogeny for platypus venom defensin-like peptides (vDLPs), for lizard

venom crotamine-like peptides (vCLPs) and for snake venom crotamines. These venom proteins have thus been co-opted from pre-existing non-toxin homologues independently in platypus and in lizards and snakes⁴⁸.

is currently lacking because platypus recombination rates have not been measured. A further examination of the CpG fraction, that associated with promoter elements, is found in Supplementary Notes 24 and Supplementary Fig. 15.

Conclusions

The egg-laying platypus is a remarkable species with many biological features unique among mammals. Our sequencing of the platypus genome now enables us to compare its sequence characteristics and organization with those of birds and therian mammals in order to address the questions of platypus biology and to date the emergence of mammalian traits. We report here that sequence characteristics of the platypus genome show features of reptiles as well as mammals.

Platypus contains a largely standard repertoire of non-protein-coding, ncRNAs, except for the snoRNAs, which exhibit a marked expansion associated with at least one retrotransposed subfamily. Some of these retrotransposed snoRNAs are expressed and thus may have functional roles. The platypus has fully elaborated piRNA and miRNA pathways, the latter including many monotreme-specific miRNAs and miRNAs that are shared with either mammals or chickens. Many functional assessments of these novel miRNAs remain to be carried out and will surely add to our knowledge of mammalian miRNA evolution.

The 18,527 protein-coding genes predicted from the platypus assembly fall within the range for therian genomes. Of particular interest are families of genes involved in biology that links

monotremes to reptiles, such as egg-laying, vision and envenomation, as well as mammal-specific characters such as lactation, characters shared with marsupials such as antibacterial proteins, and platypus-specific characters such as venom delivery and underwater foraging. For instance, anatomical adaptations for chemoreception during underwater foraging are reflected in an unusually large repertoire of vomeronasal type 1 receptor genes. However, the repertoire of milk protein genes is typically mammalian, and the arrangement of milk protein genes seems to have been preserved since the last common ancestor of monotremes and therian mammals.

Since its initial description, the platypus has stood out as a species with a blend of reptilian and mammalian features, which is a characteristic that penetrates to the level of the genome sequence. The density and distribution of repetitive sequence, for example, reflects this fact. The high frequency of interspersed repeats in the platypus genome, although typical for mammalian genomes, is in contrast with the observed mean microsatellite coverage, which appears more reptilian. Additionally, the correlation of parent-of-origin-specific expression patterns in regions of reduced interspersed repeats in the platypus suggests that the evolution of imprinting in therians is linked to the accumulation of repetitive elements.

We find that the mixture of reptilian, mammalian and unique characteristics of the platypus genome provides many clues to the function and evolution of all mammalian genomes. The wealth of new findings and confirmation of existing knowledge immediately evident from the release of these data promise that the availability of

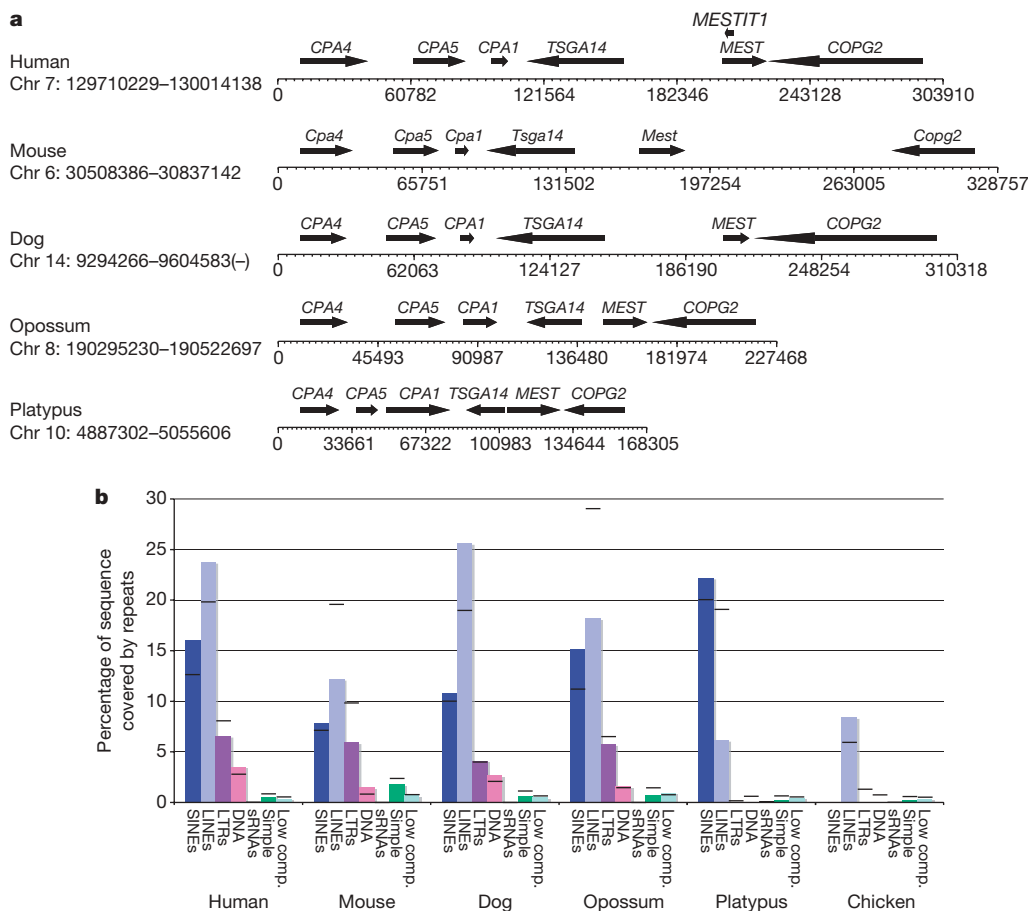


Figure 5 | Comparative mammalian analysis for a representative eutherian imprinted gene cluster (*PEG1/MEST*). a, The gene arrangement is conserved between mammals. However, non-coding regions are expanded in therians. Arrows indicate genes and the direction of transcription; the scale shows base pairs. **b**, Summary of repeat distribution for the *PEG1/MEST*

cluster. Histograms represent the sequence (%) masked by each repeat element within the *MEST* cluster; black bars represent repeat distribution across the entire genome. With the exception of SINES, platypus has fewer repeats of LINES, LTRs, DNA and simple repeats (Simple) than eutherian mammals. Low comp., low complexity; sRNAs, small RNAs.

the platypus genome sequence will provide the critically needed background to inspire rapid advances in other investigations of mammalian biology and evolution.

METHODS SUMMARY

Tissue resources. Tissue was obtained from animals captured at the Upper Barnard River, New South Wales, Australia, during breeding season (AEEC permit number R.CG.07.03 to F. Grützner; Environment ACT permit number LI 2002 270 to J. A. M. Graves; NPWS permit number A193 to R. C. Jones; AEC permit number S-49-2006 to F. Grützner).

Sequence assembly. A total of 26.9 million reads was assembled using the PCAP software²⁰. Attempts were made to assign the largest contiguous blocks of sequence to chromosomes using standard FISH techniques.

Non-coding RNAs. We used the established Rfam pipeline⁶⁸ and *de novo* sequencing to detect non-protein-coding RNAs (ncRNAs). Cloning, sequencing and annotation of sRNAs from platypus, echidna and chicken as well as miRNA sequences are described in ref. 25.

Genes. Protein-coding and non-protein-coding genes were computed using a modified version of the Ensembl pipeline (Supplementary Notes 14). Gene orthology assignment followed a procedure implemented previously⁶⁹. Orthology rate estimation was performed with PAML⁷⁰ using the model of ref. 71. In all cases, codon frequencies were estimated from the nucleotide composition at each codon position (F3X4 model).

Genome landscape. Pairwise alignments between human and dog, mouse, opossum, platypus and chicken were projected from whole-genome alignments of 28 species (<http://genome.cse.ucsc.edu/>). These alignments were the basis for phylogeny, chromosome synteny, interspersed repeats, imprinting and CpG fraction analyses.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 September 2007; accepted 25 March 2008.

- Home, E. A. A description of the anatomy of *Ornithorynchus paradoxus*. *Phil. Trans. R. Soc. Lond. B* **92**, 67–84 (1802).
- Biniinda-Emonds, O. R. P. et al. The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007).
- Caldwell, H. The embryology of Monotremata and Marsupialia. *Phil. Trans. R. Soc. Lond. B* **178**, 463–486 (1887).
- Griffiths, M. *Echidnas* (Pergamon, Oxford, 1968).
- Griffiths, M. *The Biology of the Monotremes* (Academic, New York, 1978).
- Renfree, M. B. Monotreme and marsupial reproduction. *Reprod. Fertil. Dev.* **7**, 1003–1020 (1995).
- Renfree, M. B. *Mammal Phylogeny* (Springer, New York, 1993).
- Watson, J. M., Meyne, J. & Graves, J. A. M. Ordered tandem arrangement of chromosomes in the sperm heads of monotreme mammals. *Proc. Natl Acad. Sci. USA* **93**, 10200–10205 (1996).
- Greaves, I. K., Rens, W., Ferguson-Smith, M. A., Griffin, D. & Graves, J. A. M. Conservation of chromosome arrangement and position of the X in mammalian sperm suggests functional significance. *Chromosome Res.* **11**, 503–512 (2003).
- Temple-Smith, P. & Grant, T. Uncertain breeding: a short history of reproduction in monotremes. *Reprod. Fertil. Dev.* **13**, 487–497 (2001).
- Temple-Smith, P. D. *Seasonal Breeding Biology of the Platypus*, *Ornithorynchus anatinus* (Shaw 199), with Special Reference to Male. PhD thesis, Australian National University (1973).
- Scheich, H., Langner, G., Tidemann, C., Coles, R. B. & Guppy, A. Electoreception and electrolocation in platypus. *Nature* **319**, 401–402 (1986).
- Pettigrew, J. D. Electoreception in monotremes. *J. Exp. Biol.* **202**, 1447–1454 (1999).
- Bick, Y. A. E. & Sharman, G. B. The chromosomes of the platypus (*Ornithorynchus: Monotremata*). *Cytobios* **14**, 17–28 (1975).
- Wrigley, J. M. & Graves, J. A. M. Two monotreme cell lines, derived from female platypuses (*Ornithorynchus anatinus*; Monotremata, Mammalia). *In Vitro* **20**, 321–328 (1984).
- El-Mogharbel, N. et al. *DMRT* gene cluster analysis in the platypus: new insights into genomic organization and regulatory regions. *Genomics* **89**, 10–21 (2007).
- Rens, W. et al. Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3Y3X4Y4X5Y5 male sex chromosome constitution. *Proc. Natl Acad. Sci. USA* **101**, 16257–16261 (2004).
- 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).
- International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
- Huang, X. et al. Application of a superword array in genome assembly. *Nucleic Acids Res.* **34**, 201–205 (2006).
- McMillan, D. et al. Characterizing the chromosomes of the platypus (*Ornithorynchus anatinus*). *Chromosome Res.* **15**, 961–974 (2007).
- Maxwell, E. S. & Fournier, M. J. The small nucleolar RNAs. *Annu. Rev. Biochem.* **64**, 897–934 (1995).
- Schmitz, J. et al. Retroposed SNOfall—A mammalian-wide comparison of platypus snoRNAs. *Genome Res.* doi:10.1101/gr.7177908 (in the press).
- Zemann, A., op de Bekke, A., Kieffmann, M., Brosius, J. & Schmitz, J. Evolution of small nucleolar RNAs in nematodes. *Nucleic Acids Res.* **34**, 2676–2685 (2006).
- Murchison, E. P. et al. Conservation of small RNA pathways in platypus. *Genome Res.* doi:10.1101/gr.73056.107 (in the press).
- Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi/piRNA pathway provides adaptive defense for the transposon arms race. *Science* **318**, 761–764 (2007).
- Curwen, V. et al. The Ensembl automatic gene annotation system. *Genome Res.* **14**, 942–950 (2004).
- Goodstadt, L. & Ponting, C. P. Phylogenetic reconstruction of orthology, paralogy, and conserved synteny for dog and human. *PLoS Comput. Biol.* **2**, e133 (2006).
- Goodstadt, L., Heger, A., Webber, C. & Ponting, C. P. An analysis of the gene complement of a marsupial, *Monodelphis domestica*: evolution of lineage-specific genes and giant chromosomes. *Genome Res.* **17**, 969–981 (2007).
- Emes, R. D., Goodstadt, L., Winter, E. E. & Ponting, C. P. Comparison of the genomes of human and mouse lays the foundation of genome zoology. *Hum. Mol. Genet.* **12**, 701–709 (2003).
- Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98 (1973).
- Kirsch, J. A. & Mayer, G. C. The platypus is not a rodent: DNA hybridization, annote phylogeny and the palimpsest theory. *Phil. Trans. R. Soc. Lond. B* **353**, 1221–1237 (1998).
- Rowe, T., Rich, T. H., Vickers-Rich, P., Springer, M. & Woodburne, M. O. The oldest platypus and its bearing on divergence timing on platypus and echidna clades. *Proc. Natl Acad. Sci. USA* **105**, 1238–1242 (2008).
- Grus, W. E., Shi, P. & Zhang, J. Largest vertebrate vomeronasal type 1 receptor (V1R) gene repertoire in the semi-aquatic platypus. *Mol. Biol. Evol.* **24**, 2153–2157 (2007).
- Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
- Aloni, R., Olender, T. & Lancet, D. Ancient genomic architecture for mammalian olfactory clusters. *Genome Biol.* **7**, R88 (2006).
- Jovine, L., Qi, H., Williams, Z., Litscher, E. S. & Wassarman, P. M. Features that affect secretion and assembly of zona pellucida glycoproteins during mammalian oogenesis. *Soc. Reprod. Fertil.* **63** (suppl.), 187–201 (2007).
- Riggio, M., Scudiero, R., Filosa, S. & Parisi, E. Sex- and tissue-specific expression of aspartic proteinases in *Danio rerio* (zebrafish). *Gene* **260**, 67–75 (2000).
- Nixon, B., Aitken, R. J. & McLaughlin, E. A. New insights into the molecular mechanisms of sperm-egg interaction. *Cell. Mol. Life Sci.* **64**, 1805–1823 (2007).
- Morel, L., Dufaure, J. P. & Depeiges, A. The lipocalin sperm coating lizard epididymal secretory protein family: mRNA structural analysis and sequential expression during the annual cycle of the lizard, *Lacerta vivipara*. *J. Mol. Endocrinol.* **24**, 127–133 (2000).
- Schlomann, U. et al. The metalloprotease disintegrin ADAM8. Processing by autocatalysis is required for proteolytic activity and cell adhesion. *J. Biol. Chem.* **277**, 48210–48219 (2002).
- Ofedal, O. The origin of lactation as a water source for parchment-shelled eggs. *J. Mammary Gland Biol. Neoplasia* **7**, 253–266 (2002).
- Kawasaki, K. & Weiss, K. M. Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. *Proc. Natl Acad. Sci. USA* **100**, 4060–4065 (2003).
- Lester, K. S., Boyde, A., Gilkeson, C. & Archer, M. Marsupial and monotreme enamel structure. *Scanning Microsc.* **1**, 401–420 (1987).
- de Plater, G., Martin, R. L. & Milburn, P. J. A pharmacological and biochemical investigation of the venom from the platypus (*Ornithorhynchus anatinus*). *Toxicol.* **33**, 157–169 (1995).
- Fry, B. G. From genome to “venome”: molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res.* **15**, 403–420 (2005).
- Whittington, C. et al. Defensins and the convergent evolution of platypus and reptile venom genes. *Genome Res.* doi:10.1101/gr.7149808 (in the press).
- Diener, E. & Ealey, E. H. Immune system in a monotreme: studies on the Australian echidna (*Tachyglossus aculeatus*). *Nature* **208**, 950–953 (1965).
- Kelley, J., Walter, L. & Trowsdale, J. Comparative genomics of natural killer cell receptor gene clusters. *PLoS Genet.* **1**, 129–139 (2005).
- Belov, K. et al. Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. *Genome Res.* **17**, 982–991 (2007).
- Durr, U. H., Sudheendra, U. S. & Ramamoorthy, A. LL-37, the only human member of the cathelicidin family of antimicrobial peptides. *Biochim. Biophys. Acta* **1758**, 1408–1425 (2006).
- Bals, R. & Wilson, J. M. Cathelicidins—a family of multifunctional antimicrobial peptides. *Cell. Mol. Life Sci.* **60**, 711–720 (2003).

54. Zanetti, M. Cathelicidins, multifunctional peptides of the innate immunity. *J. Leukoc. Biol.* **75**, 39–48 (2004).
55. Veyrunes, F. *et al.* Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res.* doi:10.1101/gr.7101908 (in the press).
56. Itoh, Y. *et al.* Dosage compensation is less effective in birds than in mammals. *J. Biol.* **6**, 2 (2007).
57. Gemmell, N. J. *et al.* Determining platypus relationships. *Aust. J. Zool.* **43**, 283–291 (1995).
58. Smit, A. F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657–663 (1999).
59. Kriegs, J. O. *et al.* Waves of genomic hitchhikers shed light on the evolution of gamebirds (Aves: Galliformes). *BMC Evol. Biol.* **7**, 190 (2007).
60. Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
61. Killian, J. K. *et al.* M6P/IGF2R imprinting evolution in mammals. *Mol. Cell* **5**, 707–716 (2000).
62. Suzuki, S. *et al.* Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. *PLoS Genet.* **3**, e55 (2007).
63. Edwards, C. A. *et al.* The evolution of imprinting: chromosomal mapping of orthologues of mammalian imprinted domains in monotreme and marsupial mammals. *BMC Evol. Biol.* **7**, 157 (2007).
64. Rapkins, R. W. *et al.* The Prader-Willi/Angelman imprinted domain was assembled recently from non-imprinted components. *PLoS Genet.* **2**, 1–10 (2006).
65. Meunier, J. & Duret, L. Recombination drives the evolution of GC-content in the human genome. *Mol. Biol. Evol.* **21**, 984–990 (2004).
66. Mikkelsen, T. S. *et al.* Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447**, 167–177 (2007).
67. Coop, G. & Przeworski, M. An evolutionary view of human recombination. *Nature Rev. Genet.* **8**, 23–34 (2007).
68. Griffiths-Jones, S. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124 (2005).
69. Heger, A. & Ponting, C. P. Evolutionary rate analyses of orthologues and paralogues from twelve *Drosophila* genomes. *Genome Res.* **17**, 1837–1849 (2007).
70. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
71. Goldman, N. & Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The sequencing of platypus was funded by the National Human Genome Research Institute (NHGRI). This research was supported by grant HG002238 from the NHGRI (W.M.), NGFN (0313358A; to J.S. and J.B.), the DFG (SCHM 1469; to J.S. and J.B.), National Science Foundation BCS-0218338 (M.A.B.) and EPS-0346411 (M.A.B.), National Institutes of Health RO1 GM59290 (M.A.B.), National Institutes of Health RO1HG02385 (E.E.E.), Australian Research Council (F.G.), UK Medical Research Council (C.P.P. and A.H.), Ministry of Science-Spain (X.S.P. and C.L.-O.) and the State of Louisiana Board of Regents Support Fund (M.A.B.). We thank T. Grant, S. Akiyama, P. Temple-Smith, R. Whittington and the Queensland Museum for platypus sample collection and DNA, and Macquarie Generation and Glenrock station for providing access and facilities during sampling. Approval to collect animals was granted by the New South Wales National Parks and Wildlife Services, New South Wales. Funding support for some platypus samples was provided by Australian Research Council and W.V. Scott Foundation. We thank M. Shelton, I. Elton and the Healesville Sanctuary for platypus pictures. We thank L. Duret for assistance on genome landscape analysis; G. Shaw for use of the silhouettes on Fig. 1; and Z.-X. Luo, M. Archer and R. Beck for advice on the Fig. 1 phylogeny. We acknowledge the approved use of the green anole lizard sequence data provided by the Broad Institute. Resources for exploring the sequence and annotation data are available on browser displays available at UCSC (<http://genome.ucsc.edu>), Ensembl (<http://www.ensembl.org>) and the NCBI (<http://www.ncbi.nlm.nih.gov>).

Author Information The *Ornithorhynchus anatinus* whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the project accession AAPN000000000. The version described in this paper is the first version, AAPN010000000. The SNPs have been deposited in the dbSNP database (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) with Submitter Method IDs PLATYPUS-ASSEMBLY_SNPS_200801 and PLATYPUS-READS_SNPS_200801. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all

readers at www.nature.com/nature. Correspondence and requests for materials should be addressed to W.C.W. (wwarren@wustl.edu) or R.K.W. (rwilson@wustl.edu).

Wesley C. Warren¹, LaDeana W. Hillier¹, Jennifer A. Marshall Graves², Ewan Birney³, Chris P. Ponting⁴, Frank Grützner⁵, Katherine Belov⁶, Webb Miller⁷, Laura Clarke⁸, Asif T. Chinwalla¹, Shiaw-Pyng Yang¹, Andreas Heger⁴, Devin P. Locke¹, Pat Miethke², Paul D. Waters², Frédéric Veyrunes^{2,9}, Lucinda Fulton¹, Bob Fulton¹, Tina Graves¹, John Wallis¹, Xose S. Puente¹⁰, Carlos López-Otin¹⁰, Gonzalo R. Ordóñez¹⁰, Evan E. Eichler¹¹, Lin Chen¹¹, Ze Cheng¹¹, Janine E. Deakin², Amber Alsop², Katherine Thompson², Patrick Kirby², Anthony T. Papenfuss¹², Matthew J. Wakefield¹², Tsviya Olender¹³, Doron Lancet¹³, Gavin A. Huttley¹⁴, Arian F. A. Smit¹⁵, Andrew Pask¹⁶, Peter Temple-Smith^{16,17}, Mark A. Batzer¹⁸, Jerilyn A. Walker¹⁸, Miriam K. Konkel¹⁸, Robert S. Harris⁷, Camilla M. Whittington⁶, Emily S. W. Wong⁶, Neil J. Gemmell¹⁹, Emmanuel Buschiazio¹⁹, Iris M. Vargas Jentzsch¹⁹, Angelika Merkel¹⁹, Juergen Schmitz²⁰, Anja Zemann²⁰, Gennady Churakov²⁰, Jan Ole Kriegs²⁰, Juergen Brosius²⁰, Elizabeth P. Murchison²¹, Ravi Sachidanandam²¹, Carly Smith²¹, Gregory J. Hannon²¹, Enkhjargal Tsend-Ayush⁵, Daniel McMillan², Rosalind Attenborough², Willem Rens², Malcolm Ferguson-Smith⁹, Christophe M. Lefèvre^{22,23}, Julie A. Sharp²³, Kevin R. Nicholas²³, David A. Ray²⁴, Michael Kube²⁵, Richard Reinhardt²⁵, Thomas H. Pringle²⁶, James Taylor²⁷, Russell C. Jones²⁸, Brett Nixon²⁸, Jean-Louis Dacheux²⁹, Hitoshi Niwa³⁰, Yoko Sekita³⁰, Xiaohu Huang³¹, Alexander Stark³², Pouya Kheradpour³², Manolis Kellis³², Paul Flicek³, Yuan Chen³, Caleb Webber⁴, Ross Hardison⁷, Joanne Nelson¹, Kym Hallsworth-Pepin¹, Kim Delehaunty¹, Chris Markovic¹, Pat Minx¹, Yucheng Feng¹, Colin Kremitzki¹, Makedonka Mitreva¹, Jarret Glasscock¹, Todd Wylie¹, Patricia Wohldmann¹, Prathapan Thiru¹, Michael N. Nhan¹, Craig S. Pohl¹, Scott M. Smith¹, Shunfeng Hou¹, Marilyn B. Renfree¹⁶, Elaine R. Mardis¹ & Richard K. Wilson¹

¹Genome Sequencing Center, Washington University School of Medicine, Campus Box 8501, 4444 Forest Park Avenue, St Louis, Missouri 63108, USA. ²Australian National University, Canberra, Australian Capital Territory 0200, Australia. ³EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. ⁴MRC Functional Genetics Unit, University of Oxford, Department of Human Physiology, Anatomy and Genetics, South Parks Road, Oxford OX1 3QX, UK. ⁵Discipline of Genetics, School of Molecular & Biomedical Science, The University of Adelaide, 5005 South Australia, Australia. ⁶Faculty of Veterinary Science, The University of Sydney, Sydney, New South Wales 2006, Australia. ⁷Center for Comparative Genomics and Bioinformatics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ⁸Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ⁹Cambridge University, Department of Veterinary Medicine, Madingley Road, Cambridge CB3 0ES, UK. ¹⁰Instituto Universitario de Oncología, Departamento de Bioquímica y Biología Molecular, Universidad de Oviedo, 33006-Oviedo, Spain. ¹¹Department of Genome Sciences, Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. ¹²The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3050, Australia. ¹³Crown Human Genome Center, Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. ¹⁴John Curtin School of Medical Research, The Australian National University, Canberra, Australian Capital Territory 0200, Australia. ¹⁵Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103-8904, USA. ¹⁶Department of Zoology, The University of Melbourne, Victoria 3010, Australia. ¹⁷Monash Institute of Medical Research, 27–31 Wright Street, Clayton, Victoria 3168, Australia. ¹⁸Department of Biological Sciences, Center for Bio-Modular Multi-Scale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, Louisiana 70803, USA. ¹⁹School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand. ²⁰Institute of Experimental Pathology, University of Muenster, Von-Esmarch-Strasse 56, D-48149 Muenster, Germany. ²¹Cold Spring Harbor Laboratory, Howard Hughes Medical Institute, Cold Spring Harbor, New York 11724, USA. ²²Victorian Bioinformatics Consortium, Monash University, Clayton, Victoria 3080, Australia. ²³CRC for Innovative Dairy Products, Department of Zoology, University of Melbourne, Victoria 3010, Australia. ²⁴Department of Biology, West Virginia University, Morgantown, West Virginia 26505, USA. ²⁵MPI Molecular Genetics, D-14195 Berlin-Dahlem, Ihnestrasse 73, Germany. ²⁶Sperling Foundation, Eugene, Oregon 97405, USA. ²⁷Courant Institute, New York University, New York, New York 10012, USA. ²⁸Discipline of Biological Sciences, School of Environmental and Life Sciences, University of Newcastle, New South Wales 2308, Australia. ²⁹UMR INRA-CNRS 6073, Physiologie de la Reproduction et des Comportements, Nouzilly 37380, France. ³⁰Laboratory for Pluripotent Cell Studies, RIKEN Center for Developmental Biology (CDB), 2-2-3 Minatogima-minamimachi, Chuo-ku, Kobe 6500047, Japan. ³¹Department of Computer Science, Iowa State University, 226 Atanasoff Hall, Ames, Iowa 50011, USA. ³²The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02139, USA.

METHODS

Sequence assembly. A total of 26.9 million reads was assembled using the PCAP software²⁰. Assembly quality assessment accounted for read depth, chimaeric reads, repeat content, cloning bias, G+C content and heterozygosity (Supplementary Notes 4–11). We identified a total of ~1.2 million single nucleotide polymorphisms (SNPs) within the 1.84-Gb sequenced female platypus genome using two independent analyses, SSAHA2 (SSAHA: a fast search method for large DNA databases⁷²) and PCAP output²⁰ (Supplementary Notes 11).

Non-coding RNAs. snoRNA annotation is as described in ref. 23. miRNAs sharing a heptamer at nucleotide position 2–8 were defined as a family. Homology with mouse/human miRNAs was based on annotated miRNAs in Rfam (<http://microrna.sanger.ac.uk/sequences/index.shtml>). piRNA sequences have been submitted to GEO (<http://www.ncbi.nlm.nih.gov/geo/>). miRNA total cloning frequency was normalized across tissue libraries by scaling cloning frequency per library by a factor representing total number of miRNA reads per library.

Genes. Orthologue groups were selected based on whether they contained genes predicted only from the platypus, and not from the chicken, opossum, dog, mouse or human genome assemblies (Supplementary Notes 15–17). Other groups were selected where the number of in-paralogous platypus genes exceeded the numbers of the other (chicken, opossum, dog, mouse and human) terminal lineages. Some of these groups represent erroneous gene predictions where, for example, protein-coding sequence predictions represented instead transposed element or highly repetitive sequence, or overlapped, on the reverse strand, other well-established coding sequence. Such instances were discarded. Lineage-specific gene loss was detected by inspection of BLASTZ alignment chains and nets at the UCSC Genome Browser (<http://genome.cse.ucsc.edu/>); by the interrogation of all known cDNA, EST and protein sequences held in GenBank using BLAST; and by attempting to predict orthologous genes within genomic intervals flanked by syntenic anchors.

Genome landscape. To establish phylogeny we extended the basic data sampling approach described previously⁷³ to protein-coding genes, and used established techniques to analyse protein-coding indels⁷⁴ and retrotransposon insertions⁷⁵ (Supplementary Notes 19).

The population structure of 90 platypuses from different regions in Australia was determined using Structure software v2.1 (ref. 76) using genotypes of 57 polymorphic Mon-1 and LINE2 loci. Five thousand replications were examined (Supplementary Notes 21).

Microsatellites were identified across the platypus genome (ornAna1) combining two programs: Tandem Repeat Finder (TRF)⁷⁷ and Sputnik⁷⁸ (Supplementary Notes 22).

For the imprinting cluster of *PEG1/MEST*, comparative maps were compiled from Vega annotations for the mouse and human, and Ensembl gene builds for other species. Multiple alignments of each region for repeat distribution analyses were constructed using MLAGAN⁷⁹ with translated anchoring.

We examined genomic assemblies for human (hg18), mouse (musMus8), dog (canFam2), opossum (monDom4), platypus (ornAna1) and chicken (galGal3), downloaded from the UCSC Genome Browser (<http://genome.ucsc.edu>), and computed the fraction of G+C nucleotides in each non-overlapping 10,000-bp window free of ambiguous bases. Bases in repeats were not distinguished and were counted along with non-repeat bases. For platypus all assembled sequence was analysed; for the other species only bases assigned to chromosomes were used.

72. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. *Genome Res.* 11, 1725–1729 (2001).

73. Huttley, G. A., Wakefield, M. J. & Easteal, S. Rates of genome evolution and branching order from whole genome analysis. *Mol. Biol. Evol.* 24, 1722–1730 (2007).

74. Murphy, W. J., Pringle, T. H., Crider, T. A., Springer, M. S. & Miller, W. Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res.* 17, 413–421 (2007).

75. Kriegs, J. O. et al. Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol.* 4, e91 (2006).

76. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959 (2000).

77. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580 (1999).

78. La Rota, M., Kantety, R. V., Yu, J. K. & Sorrells, M. E. Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics* 6, 23–29 (2005).

79. Brudno, M. et al. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* 13, 721–731 (2003).