SCIENCE CHINA
Life Sciences

•RESEARCH PAPER•

October 2021 Vol.64 No.10: 1765–1780
https://doi.org/10.1007/s11427-020-1841-5

Genome-wide signatures of mammalian skin covering evolution

Peng Cao[1†*], Qinlong Dai[9,10†], Cao Deng[4†], Xiang Zhao[5†], Shishan Qin[4], Jian Yang[4], Ran Ju[7], Zhiwen Wang[5], Guoqing Lu[3], Xiaodong Gu[6], Zhisong Yang[8] & Lifeng Zhu[2,3*]

[1]*Key Laboratory of Drug Targets and Drug Leads for Degenerative Diseases, Affiliated Hospital of Integrated Traditional Chinese and Western Medicine, Nanjing University of Chinese Medicine, Nanjing 210023, China;*
[2]*College of Life Sciences, Nanjing Normal University, Nanjing 210046, China;*
[3]*University of Nebraska at Omaha, Omaha 68181, USA;*
[4]*DNA Stories Bioinformatics Center, Chengdu 610021, China;*
[5]*PubBio-Tech Services Corporation, Wuhan 430070, China;*
[6]*Sichuan Station of Wildlife survey and Management, Chengdu 610082, China;*
[7]*Nanjing Foreign Language School Xianlin Branch, Nanjing 210023, China;*
[8]*Key Laboratory of Southwest China Wildlife Resources Conservation (Ministry of Education), China West Normal University, Nanchong 637002, China;*
[9]*Shimian Research Center of Giant Panda Small Population Conservation and Rejuvenation, Yaan 625400, China;*
[10]*Liziping Natural Reserve, Yaan 625400, China*

Received August 15, 2020; accepted October 15, 2020; published online January 19, 2021

Animal body coverings provide protection and allow for adaptation to environmental pressures such as heat, ultraviolet radiation, water loss, and mechanical forces. Here, using a comparative genomics analysis of 39 mammal species spanning three skin covering types (hairless, scaly and spiny), we found some genes (e.g., *UVRAG*, *POLH*, and *XPC*) involved in skin inflammation, skin innate immunity, and ultraviolet radiation damage repair were under selection in hairless ocean mammals (e.g., whales and manatees). These signatures might be associated with a high risk of skin diseases from pathogens and ultraviolet radiation. Moreover, the genomes from three spiny mammal species shared convergent genomic regions (*EPHB2*, *EPHA4*, and *NIN*) and unique positively selected genes (*FZD6*, *INVS*, and *CDC42*) involved in skin cell polarity, which might be related to the development of spines. In scaly mammals, the shared convergent genomic regions (e.g., *FREM2*) were associated with the integrity of the skin epithelium and epidermal adhesion. This study identifies potential convergent genomic features among distantly related mammals with the same skin covering type.

comparative genomics, skin coverings, spiny mammals, hairless mammals, scaly mammals, convergent evolution

Citation: Cao, P., Dai, Q., Deng, C., Zhao, X., Qin, S., Yang, J., Ju, R., Wang, Z., Lu, G., Gu, X., et al. (2021). Genome-wide signatures of mammalian skin covering evolution. Sci China Life Sci 64, 1765–1780. https://doi.org/10.1007/s11427-020-1841-5

INTRODUCTION

Vertebrates have diverse, rigid body coverings (spines, scales, feathers, and fur). These body coverings provide protection and allow for adaptation to environmental pressures such as ultraviolet radiation, heat, water loss, and mechanical forces (Alibardi, 2003; Chuong and Homberger, 2003; Prum and Torres, 2004). Within vertebrates, mammals have skin that is covered by spines, scales, or fur or have no covering at all (i.e., hairless). What are the genome-wide signatures of mammalian skin covering evolution (spines, scales, and hairless)? And what is the genome-wide selective

†Contributed equally to this work
*Corresponding authors (Peng Cao emai: pcao79@yahoo.com; Lifeng Zhu, email: zhulf2020@126.com)

pressure under the different kinds of skin covering types?

One classic type of skin covering is hairless such as manatees (order Sirenia) and whales (order Cetartiodactyla). Lacking hair follicles is their typical feature. Comparative genomic analyses have revealed a substantial divergence of the naked mole-rat nuclear receptor corepressor hairless (*HR*) from ten other mammalian orthologues, and the presence of an amino acid substitution (cysteine to tryptophan at the 397 position) is associated with the hairless phenotype (Kim et al., 2011). Keratin-related gene families (e.g., keratin (*KRT*) and keratin-associated protein gene families (*KRTAP*)) essential for hair formation are contracted in the whale lineage (Yim et al., 2014). Hair follicles interact closely with the immune and neuroendocrine systems in the skin, supporting surveillance against pathogens and aiding sensory perception. Hair follicles also assist in wound healing and skin re-pigmentation (Yu et al., 2006). Cetaceans are affected by various cutaneous disorders caused by infectious agents, parasitic copepods and fishes and ultraviolet (UV) radiation (Bertulli et al., 2012; Martinez-Levasseur et al., 2011; van Bressem et al., 2007). One study found that 69% of 81 dead cetaceans in British waters had skin diseases, and the most common conditions were wounds and other traumatic injuries, viral infections (predominantly pox), and scars (Baker, 1992). Florida manatees also face a risk of infectious skin disease (Leistra et al., 2003; Marsh et al., 2011). The effects of UV radiation are unlikely to be negligible, particularly for species such as cetaceans in which anatomical (e.g., lack of fur, feathers or keratinized plates) or life-history constraints (e.g., obligate air-breathing, lactation and socialization at the sea surface) mean that it is impossible to avoid ongoing exposure to UV radiation (Acevedo-Whitehouse and Duffus, 2009; Geraci et al., 1986). Photographic and histological surveys of three seasonally sympatric whales revealed that UV radiation could pose a threat to whale health (e.g., blistering) (Martinez-Levasseur et al., 2011). Skin lesions commonly associated with severe sun damage in humans are widespread, and individuals with fewer melanocytes have more lesions and fewer apoptotic cells (Martinez-Levasseur et al., 2011). Moderate to severe skin burns have been reported in Amazonian manatees (*Trichechus inunguis* Natterer, 1883) after extended exposure to direct sunlight when out of the water (Rosas, 1994). Marine mammals from different orders share several phenotypic traits (e.g., limbs and bone density) adapted to the aquatic environment. Comparative genomic analyses have shown that convergent amino acid substitutions are widespread throughout the genomes of marine mammals and that a subset of these substitutions was in genes evolving under positive selection and putatively associated with a marine phenotype (Foote et al., 2015). Here, we ask the question: do distantly related hairless species share some common genomic features related to skin immunity and UV burn response.

Moreover, there are two other special types of skin coverings: scales and spines. For instance, pangolin (order Pholidota) skin is characterized by large, overlapping keratinized scales (Meyer et al., 2013). Four candidate *KRT* genes (*KRT36*, *KRT75*, *KRT82*, and *KRTAP3-1*) are under positive selection in the genomes of Malayan (*Manis javanica*) and Chinese (*Manis pentadactyla*) pangolins (Choo et al., 2016). Differences in the *KRTAP* gene repertoire and gene expression and distinct rates of gene conversion/recombination, pseudogenization, and positive selection are likely responsible for micro- and macro-phenotypic hair diversification among mammals (Khan et al., 2014). The outer shell of the nine-banded armadillo (*Dasypus novemcinctus*, order Cingulata) is composed of ossified dermal scutes covered by non-overlapping, keratinized epidermal scales connected by flexible bands of skin. The underside of the body and medial leg surfaces have no armor and are instead covered by tough skin and a layer of coarse hair (Feldhamer, 2007). The skull of pangolins is smoothly conical, lacking the ridges and crests found on most mammalian crania (Barlow, 1984). Pangolins (order Pholidota) and armadillos (order Edentata) have no close taxonomic relationship (belonging to different orders) but have similar types of defensive covering: external shields and protective armor not found in any other mammal. Given their shared habits, morphology and diet, do armadillos and pangolins also share selective pressures (potentially convergent features in their genomes) associated with specialized skeletal development and scale-covered skin? In addition, spiny skin coverings are found in hedgehogs (orders Eulipotyphla and Afrosoricida) and the distantly related porcupines (order Rodentia) (Chernova and Kuznetsov, 2001). Spines are a defense against potential predators because they can puncture the skin and inflict pain. Many comparative descriptions of the structural features of hairs and spines have been performed, primarily in taxonomic studies of rodents (Hoey et al., 2004; Po-Chedley and Shadle, 1955), but the question remains: do distantly related spiny mammals have common adaptive genomic features related to spine development?

Here, we sequenced the genomes of porcupine (*Hystrix brachyurus, Hbra*) and Philippine pangolin (*Manis culionensis, Mcul*) using a whole-genome shotgun strategy and the Illumina HiSeq 2000 platform. Therefore, to explore genome-wide signatures of mammalian skin covering among scaly, spiny, and hairless mammals, we used comparative genomics analyses across 39 mammal species (Figure 1). We focused on two main factors: (i) the high selective pressure in the mammalian genome in response to the environment (e.g., UV light), especially in mammals with hairless skin, and (ii) the potentially convergent genomic features related to hair formation among mammals with each type of skin covering. Finally, the aim of this study was to identify potential con-

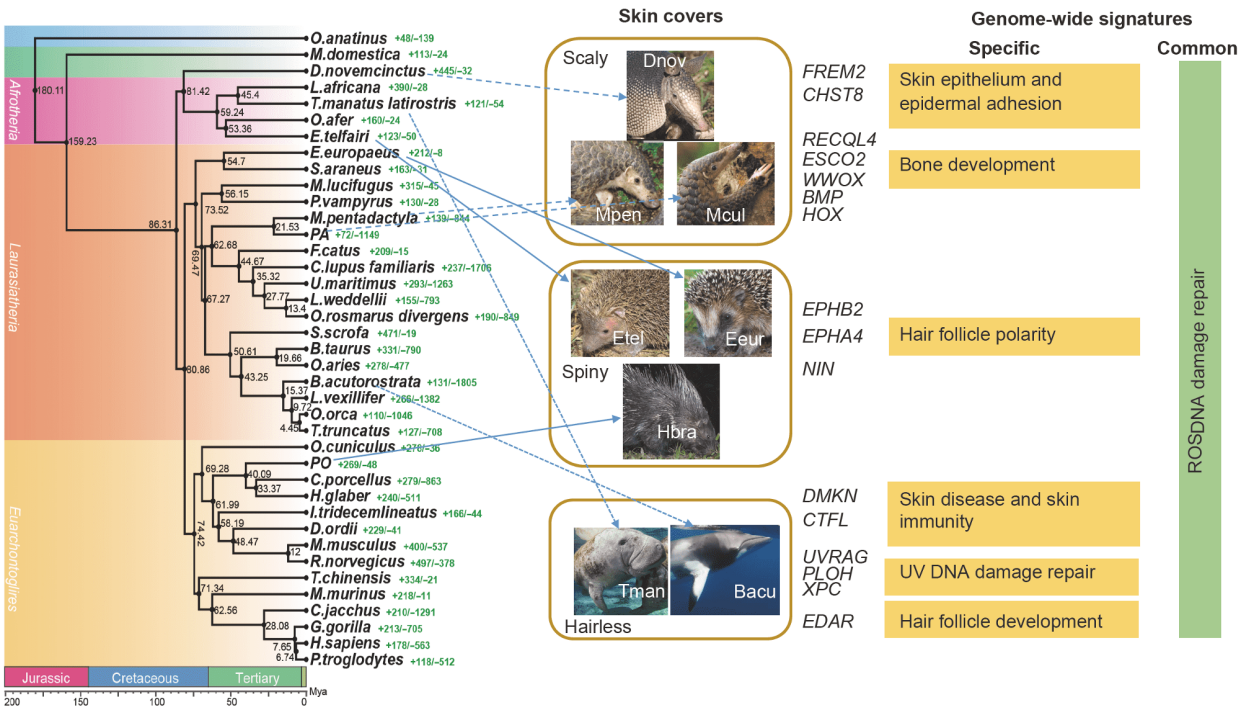

Figure 1 Species tree and divergence times of 39 mammalian species. The green numbers after each tip are the number of significantly expanded (+) and contracted (−) gene families. Dnov, *Dasypus novemcinctus*; Mpen, *Manis pentadactyla*; Mcul (PA), *Manis culionensis*; Eeur; *Erinaceus europaeus*; Etel, *Echinops telfairi*; Hbra (PO), *Hystrix brachyurus*; Tman, *Trichechus manatus latirostris*; Bacu, *Balaenoptera acutorostrata.* The photo information was in the Table S1 in Supporting Information.

vergent genomic features among distantly related mammals with the same skin covering type.

RESULTS

Genome assembly and annotation

To explore the genome-wide signatures of the evolution of mammalian skin coverings among scaly, spiny, and hairless mammals, we used comparative genomics analyses across 39 mammal species (Figure 1), in which, including the genomes of porcupine (*Hystrix brachyurus*, Hbra) and Philippine pangolin (*Manis culionensis*, Mcul) sequenced in this study. We generated ~352.36 gigabases (Gb) and 106.53 Gb of genomic sequence with 137.11-fold coverage and 44.57-fold coverage for porcupine and Philippine pangolin, respectively (Table 1; Figure S1 and Table S2 in Supporting Information). The assembled porcupine genome was ~2.33 Gb with a scaffold N50 value of 2.90 Mb, and approximately 90% of the genome sequence was contained in the 1,113 longest scaffolds (>234 kb), with the largest spanning 15.59 Mb (Table S3 in Supporting Information). The assembled Philippine pangolin genome was ~2.19 Gb with a scaffold N50 value of 13.55 kb (Table S4 in Supporting Information). The assembly quality of the Philippine pangolin genome was much lower than those of published mammalian genomes, and this might be due to DNA degradation and lack of mate-pair library sequencing data. Despite its low quality, this genome assembly captured more than 91.63% of the estimated 2.39 Gb genome size and 72.30% of core eukaryotic genes. Combining *de novo*- and homology-based strategies, we predicted 21,697 and 15,700 protein-coding genes in porcupine (Hbra) and Philippine pangolin (Mcul), respectively. Evolutionary relationships between species reflect their origins and phylogenetic relationships and are of great importance for downstream analyses. Using an orthologous set of 860 single-copy coding genes, our phylogenetic analysis and molecular evolutionary clock analysis with time constraints based on fossil records (Figure 1) revealed that Pholidota was closely related to Carnivora, and these two orders diverged ~62.68 Myr (million years) ago.

Positively selected genes and rapidly evolving genes among mammals with each skin covering type

Firstly, we evaluated the common genome-wide features for skin development. In total, 69 common positively selected genes (PSGs) were identified among eight species (Table S5 in Supporting Information; hairless: *Balaenoptera acutorostrata* (Bacu) and *Trichechus manatus latirostris* (Tman); spiny: *Echinops telfairi* (Etel), *Erinaceus europaeus* (Eeur), and Hbra; scaly: Mcul, *Manis pentadactyla* (Mpen), and

Table 1 Assembled mitogenomes from confiscated pangolin samples

MITObim version	Reference			Assemblies	
	Group	GenBank	Species	Name	Length
1.9.1	Asian	NC_036433	*Manis crassicaudata*	PAmt_Ref-Manis_crassicaudata	16,712
1.9.1	Asian	NC_036434	*Manis culionensis*	PAmt_Ref-Manis_culionensis	16,712
1.8	Asian	NC_016008	*Manis javanica*	PA_MT	16,712
1.9.1	Asian	NC_026781	*Manis javanica*	PAmt_Ref-Manis_javanica	16,712
1.9.1	Asian	KT445978	*Manis pentadactyla*	PAmt_Ref-Manis_pentadactyla	16,662
1.9.1	African	NC_026780	*Manis tricuspis*	PAmt_Ref-Manis_tricuspis	16,687

Dasypus novemcinctus (Dnov)) (Table S6 in Supporting Information). Seven common rapidly evolving genes (REGs) were identified among these eight species (Table S7 in Supporting Information). The common PSGs and REGs among the eight species were then functionally annotated using human homologs. Four genes, including *MAP3K1*, *PAX8*, *GUCY1B3*, and *VPS41*, were present in both lists (Figure 2; Tables S6 and S7 in Supporting Information). To explore the regulatory roles of these common PSGs and REGs in the context of protein-protein interaction networks, we combined the PSGs and REGs and mapped them to the STRING database (Figure 2). Several genes with a high degree of interaction were identified: *MAP3K1*, *DNMT1*, *MRPS9*, *CHFR*, *RBBP5*, *SIRT5*, and *BRCA2*. Notably, mouse genetics has revealed that the kinase *MAP3K1* is important in embryogenesis, keratinocyte migration, T cell cytokine production, and B cell antibody production (Nelson et al., 2015). *DNMT1* plays an essential role in maintaining the progenitor state in constantly replenished somatic tissues, such as mammalian epidermis (Sen et al., 2010). *BRCA2* is essential for repairing damaged DNA (Yoshida and Miki, 2004). All of these genes were under positive selection or had undergone rapid evolution in mammals with the three skin covering types examined.

Then, we evaluated the specific genome-wide features for each type of skin coverings. The unique PSGs in each skin covering type were further investigated. In hairless mammals (the manatee (Tman) and whale (Bacu)), 12 unique PSGs were identified, including *DMKN*, *CTFL*, *BRD9*, and *FAM57A* (Table S8 in Supporting Information). *DMKN* is highly expressed in epidermal stratification and differentiation during mouse development (Bazzi et al., 2007) and is upregulated in inflammatory diseases (Naso et al., 2007). In mice, dermokine-β/γ-expressed keratinocytes are increased in models of contact hypersensitivity, UV-irradiated skin injury, and wound healing (Hasegawa et al., 2013). *CTCF* positively regulates the homeostatic pool and the efficient migration of Langerhans cells required for modulating the functional immune network of the skin (Kim et al., 2015). These consensus genomic signatures likely reflect adaptation in response to pathogens or high skin disease pressure due to the loss of protection conferred by skin coverings.

In scaly mammals (Mcul, Mpen, and Dnov), we identified 31 unique PSGs, including *MLXIPL*, *ESCO2*, *RGL2*, *LONP1*, *EIF3A*, *SLC2AL10*, *RECQL4*, *VWA5B2*, *FN1*, and *AAAS* (Table S9 in Supporting Information). For example, *ESCO2* regulates *CX43* expression during skeletal regeneration in the zebrafish fin, and *CX43* mutations cause a short-fin phenotype in zebrafish and oculodentodigital dysplasia in humans (Banerji et al., 2016). A mutation in *SLC2A10* leads to arterial tortuosity syndrome, which is characterized by skin laxity and joint hypermobility (Segade, 2010). These unique PSGs, which are involved in putative normal erythropoiesis and skeletal development, revealed potential selection pressure along with scale development. Moreover, significantly over-represented GO terms (topGO results with P-value≤0.01) among the PSGs in Mcul were mainly enriched in the cellular process (BF), cellular developmental process (BF), macromolecular complex (CC), intracellular organelle (CC), chaperone binding (MF), and inositol bisphosphate phosphatase activity (MF) (Figure S2 and Table S10 in Supporting Information). One kind of macromolecular complex is protein, including KRT (found in hair and nails) (Lai-Cheong and McGrath, 2009). Here, the significantly over-represented GO terms among these PSGs in Mcul may be associated with their scaly keratinization.

In spiny mammals (Etel, Eeur, and Hbra), we identified six unique PSGs, including *RQCD1*, *GNPTAB*, *LCP2*, *FHAD1*, *WDR70*, and *TRMT13* (Table S11 in Supporting Information). *RQCD1* can regulate the activity of Akt, which is the downstream signal of the epidermal growth factor receptor (Ajiro et al., 2010). The connection between these PSGs and spine development was not clear. However, we identified 70 significantly over-represented GO terms among these PSGs in Hbra (topGO results with P-value≤0.01) (Table S12 in Supporting Information). REVIGO results of these GO terms in the cellular component category also included macromolecular complex and intracellular membrane-bound organelle (Figure S3 in Supporting Information), which might be related to spine keratinization.

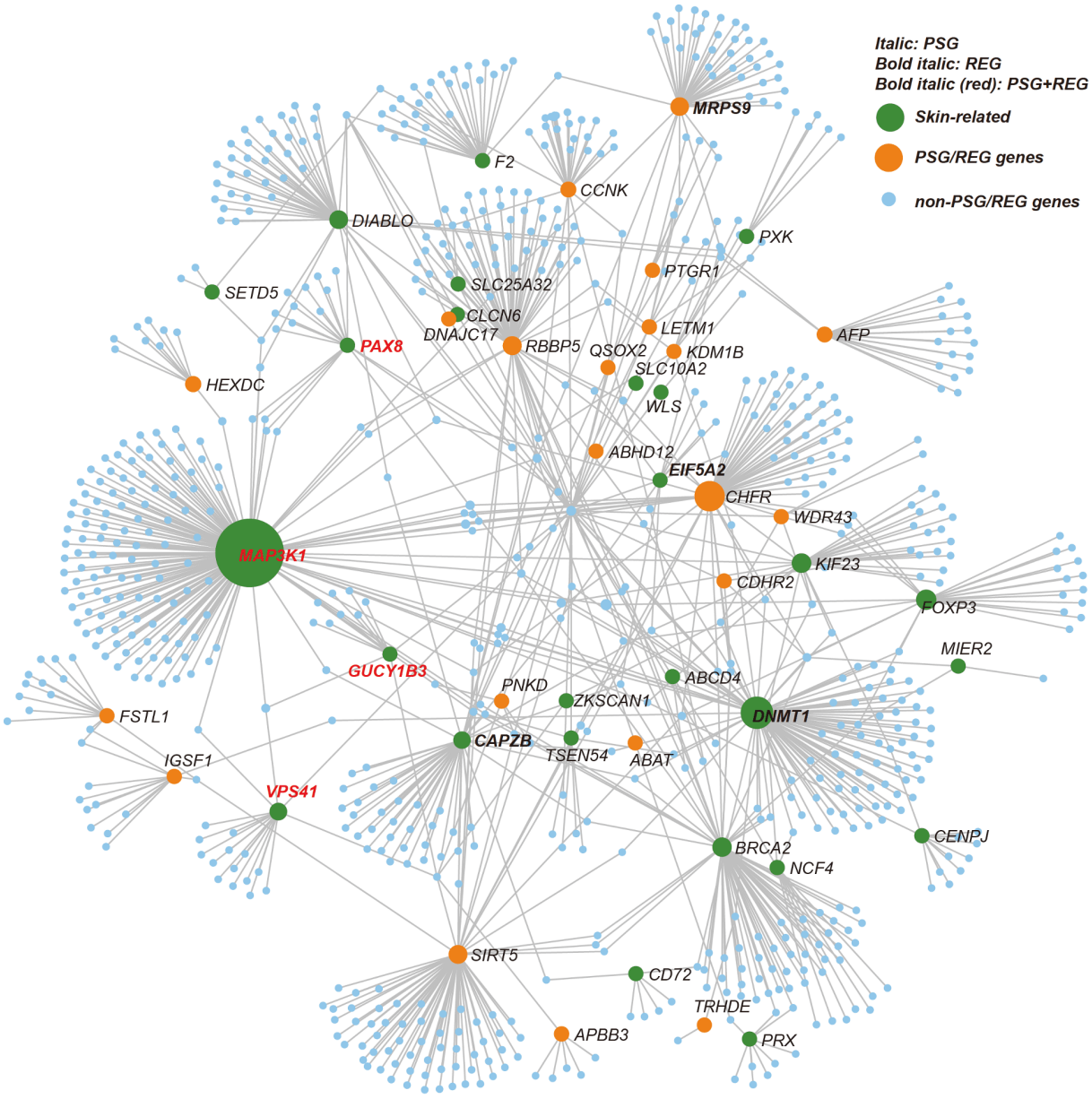

Figure 2 Protein-protein interaction networks of positively selected genes and rapidly evolving genes.

Convergently evolving genomic regions among mammals with each skin covering type

In hairless mammals (Tman and Bacu), 63 putative convergently evolving genomic regions (CEGRs) (Figure 3A) were identified (Table S13 in Supporting Information). Twenty-nine genes were further identified in these putative 63 CEGRs, including *UVRAG*, *CFAP44*, *PCDH15*, *WASF2*, *KCNK5*, and *CFL1*. For example, *UVRAG* plays an integral role in UV-induced DNA damage repair (Yang et al., 2016). The mutation in *WASF2* is associated with Wiskott-Aldrich syndrome, which is characterized by eczema (an inflammatory skin disorder), abnormal immune system function (immune deficiency), and a reduced ability to form blood clots (Chereau et al., 2005; Orange et al., 2004). In addition, the pathogenic variants in *PCDH15* result in hearing loss and Usher Syndrome Type IF (Benrebeh et al., 2016). CFAP44 is highly expressed in skin and testis, and mutations in this gene may result in male infertility and flagellum defects in Trypanosoma and humans (Coutton et al., 2018). *KCNK5* may play an important role in renal potassium transport (Kirkegaard et al., 2013). *CFL1* is essential in cytokinesis, endocytosis, and in the development of all embryonic tissues (Zhu et al., 2007). To some extent, the putative CEGR functions (e.g., *UVRAG* and *WASF2*) in these two unrelated hairless ocean mammals might reflect con-

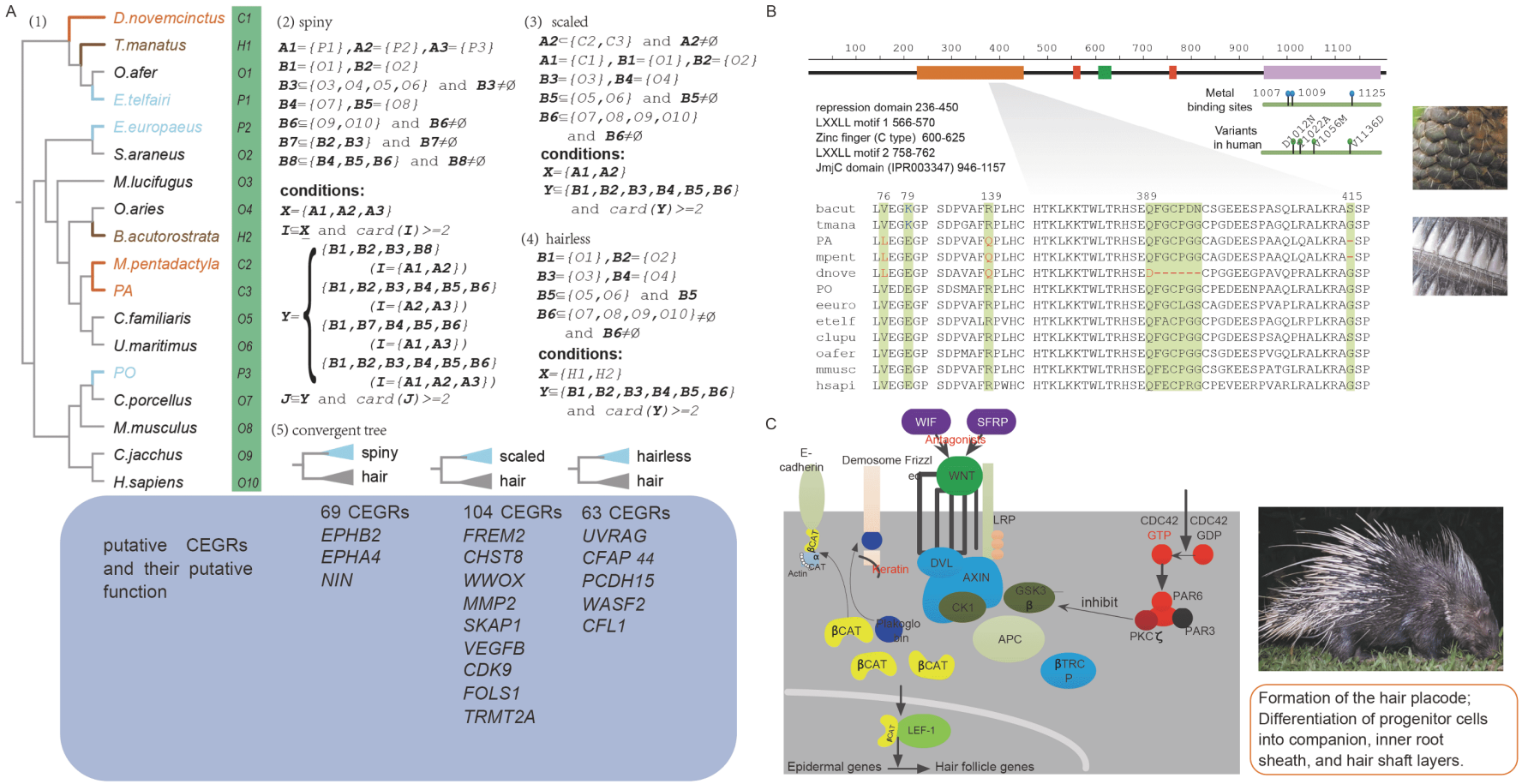

Figure 3 Convergently evolving genomic regions. A, The methodology used for CEGR identification. Genomic alignment sequences for each group of interest and background group (skin covering group) were extracted from 18-way multiple genome alignments of humans and 17 other species (1). The extracted blocks were kept if they followed the conditions required of spiny (4), scaled (3), and hairless (2) animals. Blocks were identified as CEGR if the tree was a convergent tree (5). B, Amino acid variation in *HR* genes. C, Cdc42 regulates hair follicle differentiation and cell-cell contacts in keratinocytes by modulating the turnover of β-catenin and plakoglobin (redrawn from Wu et al. (2006)). PA, *Manis culionensis*. PO, *Hystrix brachyurus*. The photo information was in the Table S1 in Supporting Information.

vergent evolution under high risk of UV damage due to the loss of the protection conferred by fur.

In scaly mammals (Mcul, Mpen, Dnov), 104 putative CEGRs were identified (Table S14 in Supporting Information), and 38 genes were further annotated in these putative CEGRs, such as *FREM2*, *CHST8*, *WWOX*, *MMP2*, *SKAP1*, *VEGFB*, *CDK9*, *FOLS1*, and *TRMT2A*. For instance, *FREM2* may be required for maintenance of the integrity of the skin epithelium and epidermal adhesion (Jadeja et al., 2005). Moreover, three BMP genes (*BMP3*, *BMP4*, and *BMP15*) associated with the formation of bone and cartilage (Bandyopadhyay et al., 2006; Hogan, 1996) were under positive selection in the pangolin. The number of PSGs in the HOX family was largest in scaly mammals (7 *HOX* genes, *HOXD3*, *HOXD10*, *HOXB1*, *HOXC8*, *HOXA2*, *HOXA3*, and *SHOX2*) compared to hairless (3 *HOX* genes, *HOXB3*, *HOXB9*, and *HOXA3*) and spiny mammals (2 *HOX* genes, *HOXB3* and *SHOX2*). *HOXD3*, *HOXD10*, and *FGF8* (PSGs in scaly mammals) play important roles in limb development (Capecchi, 1997; Carpenter et al., 1997; Lewandoski et al., 2000). Mutations in *HOXA2* (PSGs in pangolin) or *TBX10* (PSGs in armadillo) in mice have led to cleft palates and cranial skeletal defects (Bush et al., 2004; Gendron-Maguire et al., 1993). Thus, the genes identified in the CEGRs and some unique PSGs play important roles in the integrity of the skin epithelium, epidermal adhesion, bone development, cell proliferation, cell differentiation, and cell transformation. These findings suggested that armadillos and pangolins shared similar selective pressures in response to the development of specific parts of the skeleton (e.g., skull and limbs) along with their scaly skin coverings.

In spiny mammals (Etel, Eeur, Hbra), 69 putative CEGRs were identified (Table S15 in Supporting Information), and 28 genes were further annotated in these putative CEGRs, such as *EPHB2*, *EPHA4*, and *NIN*. *EPHB2* is a strong candidate for the derived head crest phenotype shared by numerous pigeon breeds (Shapiro et al., 2013). The crest originates early in development through the localized molecular reversal of feather bud polarity (Shapiro et al., 2013). Expression of the polarity marker *EPHA4* was assayed at an earlier developmental stage to test whether feather placodes, the ectodermal thickenings that give rise to feather buds, were also reversed (Shapiro et al., 2013).

Convergent amino acid substitutions among mammals with each skin covering type

In hairless mammals (Tman and Bacu), 97 convergent amino acid substitutions (CAAS) in 92 genes were identified (Tables S16 and S17 in Supporting Information). Interestingly, *EDAR* was identified as a CAAS gene. *EDAR* signaling is involved in the control of cell fate decisions in the embryonic epidermis and the regulation of cell differentiation programs in hair follicles (Botchkarev and Fessing, 2005). Loss or gain of *EDAR* signaling affects the initiation of several hair follicle types (guard and zig-zag hair follicles) and hair shaft formation (Blume-Peytavi et al., 2008; Botchkarev and Fessing, 2005). *EDAR* was also under positive selection in the minke whale, indicating a possible association with hair follicle development. Three CAAS genes (*MDC1*, *NEIL1*, and *POLH*) are associated with the repair of DNA damage due to reactive oxygen species during times of environmental stress (e.g., UV, X-rays, and chemical agents) (Jung et al., 2010; Stewart et al., 2003; Zhou et al., 2013).

In spiny mammals (Eeur, Etel, and Hbra), 22 CAAS in 22 genes were identified (Tables S18 and S19 in Supporting Information). For example, *PCDH15* (a CAAS gene) is a member of the cadherin superfamily, and members of this family encode integral membrane proteins that mediate calcium-dependent cell-cell adhesion (Hulpiau and van Roy, 2009). *CDHR2* (a CAAS gene), one of the members of the cadherin-related family (Gul et al., 2017), was under positive selection in only these three spiny species, and gene families in calcium-dependent cell-cell adhesion were significantly expanded in Etel and Hbra. Adherens junctions play an important role in intercellular junctions in epithelia (e.g., the epidermis and follicular epithelium) and require classic transmembrane cadherins (Samuelov et al., 2015). Cadherin family members may play a role in human hair growth, cycling, and pigmentation (Samuelov et al., 2015). Considering that the significantly expanded gene functions in spiny mammals include cell adhesion molecule binding and cell communication, adherens junctions might be associated with spiny skin and hair development. In addition, *XRCC5* (a CAAS gene) is associated with X-ray DNA damage repair, and rare microsatellite polymorphisms in this gene are associated with cancer in patients with varying radiosensitivity (Pucci et al., 2012).

Evolution of genes involved in hair follicle development and cycling

HR was under positive selection in two pangolin species (Mcul and Mpen). The *HR* gene plays a critical role in the maintenance of hair growth in mammals (Panteleyev et al., 1999). The discovery that *HR* is a corepressor provides a molecular basis for specific hair loss syndromes in humans and mice (Panteleyev et al., 1998). Comparative genomic analysis revealed substantial divergence of the naked mole-rat (*Heterocephalus glaber*) *HR* gene from ten other mammalian orthologues, and the presence of an amino acid substitution (cysteine to tryptophan at the 397 position) is associated with the hairless phenotype (Kim et al., 2011). Although there was no such mutation in the 397 position, several putative consensus amino acids were identified (Figure 3B). In the *HR* repression domain, two consensus

amino acid replacements (L76/V, Q139R) and two deletions (390–395 in the armadillo, 415 in the pangolins) existed (Figure 3B). After further investigating the selective pressure on KRT genes, three type II cytokeratin candidate *KRT* genes (*KRT24*, *KRT78*, and *KRT80*) were determined to be under positive selection in the armadillo genome, and the keratin filament family was under significant expansion in the pangolins. Two type II cytokeratin candidate *KRT* genes (*KRT23* and *KRT78*) and one candidate keratinocyte-associated protein gene, *KRTCAP2,* were found to be under positive selection in the porcupine. One candidate type I cytokeratin, *KRT20,* was found to be under positive selection in hairless minke whales. However, *KRT20* is a major cellular protein of mature enterocytes and goblet cells and is found explicitly in the gastric and intestinal mucosa (Moll et al., 1990). These genomic signatures of related hair and skin development genes reflect the different skin coverings, especially those of scaly mammals, in this study. For example, in scaly mammals, the hairless gene and three KRT genes were under positive selection, and the keratin filament family was also under significant expansion. We speculated that these findings might be associated with their morphological phenotypes, including scales.

The CEGR results indicated that three candidate genes (*EPHB2*, *EPHA4*, and *NIN*) might be involved in hair polarity in spiny mammals. Recent studies have demonstrated that a number of genes (*FZD6*, *VANGL2*, *CELSR1*, *INVS*, *RAC1*, *CDC42*, and *FUZ*) involved in the planar cell polarity (PCP) signaling pathway also participate in hair follicle development (Chrostek et al., 2006; Dai et al., 2011; Devenport and Fuchs, 2008; Guo et al., 2004; Ravni et al., 2009; Simons et al., 2005; Song et al., 2010; Wang et al., 2010; Wu et al., 2006). Here, three (*FZD6*, *INVS*, and *CDC42*) of these seven genes were under strong positive selection in the porcupine. These PCP genes control the orientation or differentiation of developing hair follicles, underscoring the importance of the PCP signaling pathway in hair follicle patterning (Chen and Chuong, 2012). Differentiation of skin stem cells into hair follicles requires the inhibition of β-catenin degradation. In the absence of *CDC42* (Figure 3C), degradation of β-catenin is increased, corresponding to decreased phosphorylation of GSK3β at Ser 9 and increased phosphorylation of *AXIN*, which is known to be required for binding of β-catenin to the degradation machinery (Etienne-Manneville and Hall, 2003). *NIN* is involved in the formation of the centrosome matrix and interacts with GSK3β, implying that it may also be regulated by GSK3β phosphorylation signaling (Hong et al., 2000). Wnt signaling and stabilization of β-catenin are required for the formation of the hair placode and for the differentiation of progenitor cells into the companion, inner root sheath, and hair shaft layers (Haq et al., 2003). If β-catenin is lost after hair follicle morphogenesis, follicle progenitor cells change their fate decision from hair follicle to epidermal differentiation (Huelsken et al., 2001). Several secreted protein families, including *WIF* (WNT inhibitory protein) and *SFRPs* (secreted Frizzled-related proteins), antagonize or modulate WNT/β-catenin signaling (Jansson et al., 2015), and both of these candidate genes were also under positive selection in the porcupine. In addition, mutant mice with a keratinocyte-restricted deletion of the *CDC42* gene show progressive hair loss (Wu et al., 2006). *CDC42* mediates the regulation of β-catenin turnover in the skin, which is important for WNT/β-catenin signaling in keratinocytes and then for terminal differentiation of hair follicle progenitor cells *in vivo* (Wu et al., 2006). Here, the genomic signatures, including PSGs (*FRZB*, *CDC42*, *WIF*, and *SFRP*) in WNT/β-catenin signaling pathways, PSGs (*KRT23*, *KRT78*, *KRTCAP2*, and *CDHR2*), CEGRs (*EPHB2*, *EPHA4*, and *NIN*), CAAS (cadherin family gene) and significantly expanded gene families in calcium-dependent cell-cell adhesion (cadherin), might reflect the complex regulation of hair follicle development in spiny mammals.

DISCUSSION

Potential adaptive evolution in mammals in response to exposure to the external environment

UVA, the primary type of UV light that comes from the sun and tanning beds, creates melanin by-products that damage DNA, generating DNA derivatives called cyclobutane pyrimidine dimers (CPDs) for up to three hours after light exposure (Premi et al., 2015). CPDs are associated with melanoma (Premi et al., 2015). DNA repair (e.g., nucleotide excision repair (NER)) is a general mechanism for the repair of lesions. This process excises CPDs and synthesizes new DNA to replace the damaged strand. Hairless ocean mammals, such as whales and manatees, have a high risk of skin diseases (e.g., pathogens, sunburn) due to their lack of fur and life-history constraints (Acevedo-Whitehouse and Duffus, 2009; Geraci et al., 1986; Marsh et al., 2011; Martinez-Levasseur et al., 2011; van Bressem et al., 2009; Wilson et al., 1999). The unique genomic signatures involved in skin inflammation, skin innate immunity, and UV damage repair identified here were consistent with high skin disease pressure (Figure 4A). *UVRAG*, *POLH* and *XPC* play essential roles in UV-induced photo lesson repair (Di Lucca et al., 2009; Stary and Sarasin, 2001; Yang et al., 2016). Reactive oxygen species are formed as a natural byproduct of the healthy metabolism of oxygen and have important roles in cell signaling and homeostasis. However, during times of environmental stress (e.g., UV, X-rays, heat exposure, oxygen radicals, and other chemical agents), reactive oxygen species (ROS) levels can increase dramatically (Devasagayam et al., 2004). This may result in significant damage to molecules inside the cell (e.g., DNA or RNA) (Devasagayam

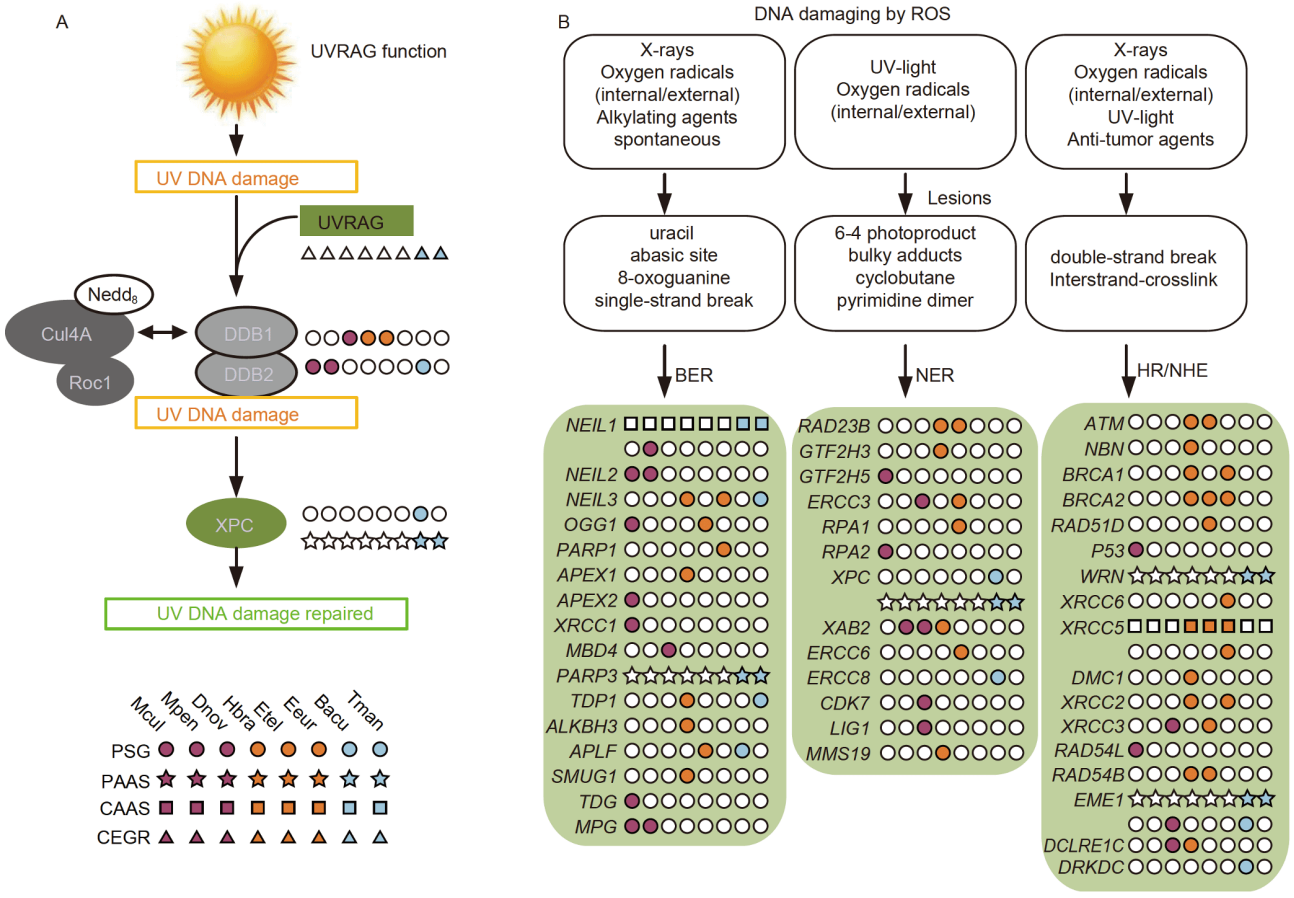

Figure 4 Genomic signatures of DNA damage repair in eight mammals. A, UVRAG functions in UV-induced DNA damage repair (redrawn from Yang et al. (2016)). B, DNA damage (caused by ROS) repair (redrawn from Bernstein and Bernstein 2015). Scaly: Dnov, *Dasypus novemcinctus*; Mpen, *Manis pentadactyla*; Mcul (PA), *Manis culionensis*. Spiny: Eeur; *Erinaceus europaeus*; Etel, *Echinops telfairi*; Hbra (PO), *Hystrix brachyurus*. Hairless: Tman, *Trichechus manatus latirostris*; Bacu, *Balaenoptera acutorostrata*.

et al., 2004). A high rate of oxidative damage to mammalian DNA has been demonstrated by measuring oxidized DNA bases excreted in urine after DNA repair (Bernstein and Bernstein, 2015; Hemnani and Parihar, 1998). The rate of oxidative DNA damage is directly related to the metabolic rate and inversely related to the life span of an organism (Bernstein and Bernstein, 2015; Hemnani and Parihar, 1998). At least 169 enzymes are either directly employed in DNA repair or influence DNA repair processes (Wood et al., 2001). Of these, 82 are directly employed in DNA repair processes (caused by ROS damage), including BER (base excision repair), NER, HRR (homologous recombination repair), and NHEJ (nonhomologous end joining) (Bernstein and Bernstein, 2015; Wood et al., 2001). Here, genomic signals in 44 genes were identified in these eight mammals compared to typical mammals with fur (Figure 4B), indicating putative selective pressure or adaptation. The substantial selection of DNA repair in these eight mammals is consistent with the fact that mammals with unusual skin coverings (especially hairless sea mammals) are less protected from sunlight than furry mammals.

Potentially convergent features in mammalian genomes related to skin coverings

Notably, no KRT genes, including those expressed in skin, were identified as a positive selection in hairless mammals. This may be another genomic signature associated with hairless skin. In contrast, several positively selected KRT genes or KRT gene family expansions were found in spiny or scaly mammals (porcupine, armadillo, and pangolin), consistent with previous research (Choo et al., 2016). The three spiny mammals shared convergent genomic regions involved in skin cell polarity. Several PSGs related to PCP in hair follicle development were identified in the porcupine. Wnt signaling and stabilization of β-catenin are required for hair follicle development, and genes associated with antagonists or agonists of the Wnt signaling pathway are under positive selection, indicating complex regulation in spine development. In the scaly mammals in this study, several unique PSGs (*BMP* and *HOX* families) and CEGRs might be associated with skeletal and bone development. Some PSGs might be required for maintenance of the integrity of the skin

epithelium, epidermal adhesion, and keratins. The armadillo and pangolin share similar anatomical and skeletal structure features (body shape, strong front legs, and long-tail) and are characterized by adaptations for digging. These findings might, therefore, reflect similar selective pressures in response to particular skeletal characteristics in addition to scaly skin.

Moreover, we compared the lists of genes involved in convergent evolution in the hairless whale and manatee with the results from several sea mammals in Foote's paper (Foote et al., 2015) and found seven overlapping genes, including *IRAK2*, *GCLC*, *SIAE*, *SERPINC1*, *DUSP27*, *M6PR*, and *S100A9*. Four (*GCLC*, *SERPINC1*, *M6PR*, and *S100A9*) were inferred to have evolved under positive selection in all three marine mammal lineages (including the hairless whales and manatees and the furry pinnipeds) and were putatively associated with marine phenotypes (e.g., bone density and diving) (Foote et al., 2015). However, we did not know whether these seven overlapping genes were associated with the hairlessness of whales and manatees in this study. Interestingly, GCLC expression is inducible in response to some forms of environmental exposure, such as toxic chemicals and oxidative stress (Franklin et al., 2009; Lu, 2009). Considering their ocean living conditions, the convergence and positive selection on this gene may reflect a selective pressure in sea mammals due to the direct exposure of their skin to seawater that may contain chemicals.

CONCLUSION

Here, we found the unique genomic signatures by using the comparative genomics analysis of 39 mammal species spanning three skin covering types (hairless, scaly, and spiny). First, we identified the selective pressure in some genes in hairless ocean mammals (e.g., whales and manatees, which might be associated with a high risk of skin diseases from pathogens and ultraviolet radiation. Second, we discovered that some CEGR regions and unique positively selected genes (involved in skin cell polarity) were shared in three spiny mammal species, which might be related to the development of spines. Third, in scaly mammals, the shared convergent genomic regions (e.g., *FREM2*) were associated with the integrity of the skin epithelium and epidermal adhesion. Finally, this study uncovered the potential convergent genomic features among distantly related mammals with the same skin covering type.

MATERIALS AND METHODS

Collection of samples

Before starting this study, we investigated the published genomes and found the genomes of the spiny and scaly mammals are rare. Thus, we planned to sequence the genomes from the mammals of these two types of skin coverings. A wild injured porcupine (Hbra, *Hystrix brachyura subcristata* based on morphological identification) died after being rescued near Baoxing Natural Reserve in Sichuan, and the body was sent to the Sichuan Station of Wildlife survey and Management. Tissues (muscles) from one pangolin (species and origin: unclear) were confiscated by Sichuan Station of Wildlife survey and Management from a local market. These tissue samples were sent to our research center for putative species identification using molecular markers.

DNA extraction and genome sequencing and assembly

Genomic DNA was extracted using Puregene Tissue Core Kit A (Qiagen, Germany). DNA from muscle from pangolin and porcupine was used for de novo sequencing. Libraries with different insert sizes were constructed at Majorbio (Shanghai, China), and the insert sizes of the libraries were 180 bp, 500 bp for pangolin, and 180 bp, 500 bp, 3 kb, 8 kb and 20 kb for porcupine. For short-insert libraries (180–500 bp), 6 μg of DNA was fragmented to the desired insert size, end-repaired and ligated to Illumina paired-end adaptors. Ligated fragments were size selected at 180 bp and 500 bp on agarose gels and purified by PCR amplification to yield the corresponding libraries. For long insert sizes (3, 8, and 10 kb) mate-pair library construction, 60 μg of genomic DNA was used; we circularized DNA, digested linear DNA, fragmented circularized DNA and purified biotinylated DNA and then performed adaptor ligation. The libraries were sequenced using a HiSeq2000 instrument.

Assembly was performed using the SOAP*denovo* (Li et al., 2010*)* with main parameters "-K 43 -d 1 -D 1 -R -F" for pangolin and "-K 45 -d 1 -D 1 -R -F" for porcupine. Gaps were then closed using GapCloser. Finally, the completeness of assembly was assessed by BUSCO v3 against the mammalian dataset (Simão et al., 2015) (Tables S20 and S21 in Supporting Information).

Species identification of the pangolin sample in this study

The mitochondrial genomes of the pangolin sample were reconstructed directly from genomic next-generation sequencing reads using the MITObim (version 1.8) (Hahn et al., 2013) with the reference mitogenome from *Manis javanica* (GenBank: NC_016008.1) (Table 1). To check the robustness of the assembly, we also assembled the mitogenome using the MITO version 1.9.1 (Hahn et al., 2013) with another five mitogenomes as the reference, respectively (NC_036433.1, *Manis crassicaudata*; NC_036434.1, *Manis culionensis*; NC_026781.1, *Manis javanica*; KT445978.1,

Manis pentadactyla; NC_026780.1, *Manis tricuspis*) (Table 1). Then, the "PA_MT" assembly was blasted against the other five assemblies, and the results showed that the assemblies are nearly identical (Table S22 in Supporting Information). Therefore, the "PA_MT" assembly was used as the final mitogenome of our pangolin sample.

To resolve the relationship of the pangolin sample with other pangolins, we first downloaded the mitochondria genomic DNA sequences of all available pangolin species and another seven species as outgroup (Table S23 in Supporting Information). Clustalw2 aligned whole mitochondria genomic sequences with default parameters, and then the conserved alignments were extracted by Gblocks (parameters: -t=d -b4=100 -b5=h) (Talavera and Castresana, 2007). The phylogenetic tree was constructed using the RAxML (parameters: -m GTRGAMMA -N 100 -f a -k -d -p 12345 -x 12345).

We obtained the final mitogenome of our pangolin sample as well as an additional five pangolin mitogenomes for reference purposes (Table 1). A Random Axelerated Maximum Likelihood phylogenetic tree (including all available pangolin species mitogenomes) showed that our pangolin sample was closely related to *Manis culionensis* (Figure S1 in Supporting Information).

Genome annotation

Transposable elements in the pangolin and porcupine genome were identified by a combination of homology-based and *de novo* approaches (Table S24 in Supporting Information)**.** Tandem repeats were identified using Tandem Repeat Finder (Parameters: 2 7 7 80 10 50 2000) (Benson, 1999). Interspersed repeats were characterized by homolog-based identification using RepeatMasker open-4.0.3 (Smit et al., 1996), and the repeat database, Repbase (Bao et al., 2015). Repeated proteins were masked using RepeatProteinMask with the transposable elements protein database. *De novo* identified interspersed repeats were annotated using RepeatModeler (Price et al., 2005), and LTR_FINDER (Xu and Wang, 2007) was used to identify the LTRs; these results were used to generate the *de novo* repeat libraries, and then RepeatMasker was run once more against the *de novo* libraries. All repeats identified in this manner were included in the total count of interspersed repeats (Table S25 in Supporting Information).

Gene structural annotation of a porcupine

The protein-coding genes were annotated following the use of a combination of homolog gene prediction and *de novo* gene prediction tools. For homolog gene prediction, the protein sequences from cow, goat, wild boar, mice, chimpanzee, and human were mapped to the genome of porcupine using tBLASTn (Boratyn et al., 2013). GeneWise (Birney et al., 2004) was used to predict the gene model based on the alignment results. *De novo* gene prediction was performed using GENSCAN (Burge and Karlin, 1997), AUGUSTUS (Stanke et al., 2006), and GLIMMERHMM (Majoros et al., 2004) based on the repeat-masked genome. Then, MAKER (Cantarel et al., 2008) was applied to integrate the predicted genes. Finally, manual integration was performed to construct the final gene set.

Gene structural annotation of Philippine pangolin (PA, Mcul) and Chinese pangolin

In order to increase the sample size of scaly mammals, we also used one published genome of Chinese pangolin (Mpen) and re-annotated in this study. *De novo* gene prediction and sequence homology-based predictions were used for gene prediction. Briefly, for *de novo* gene prediction, SNAP, GeneMark-ES, and Augustus were used to predict genes on transposable-elements-hard-masked genome sequences, and the high-quality data set for training these *ab initio* gene predictors were generated using CEGMA according to the manual of Augustus. For sequence homology-based gene prediction, proteins sequences from SwissProt vertebrate database and three organisms (human, cat, and dog from Ensembl release 78) were incorporated into MAKER2 to generate homology gene structures (Holt and Yandell, 2011). All predicted gene structures were integrated into consensus gene models using MAKER2.

Gene functional annotation

To determine the functional annotation of the gene models, we used a BLASTP search with an E-value$\leq 1\times10^{-5}$ was performed against protein databases, including NR (non-redundant protein sequences in NCBI), SwissProt, RefSeq, and Trembl. The resulting NR BLASTP hits were processed by BLAST2GO (Conesa et al., 2005) to retrieve associated gene ontology terms describing biological processes, molecular functions, and cellular components (E-value$\leq 1\times10^{-5}$). The motifs and domains of each gene model were predicted by InterProScan 4.8 against public protein databases, including ProDom, PRINTS, Pfam, SMART, PANTHER, PROSITE, and TIGR (Table S26 in Supporting Information).

Genome evolution

Identification of orthologous gene clusters

Genome and annotation data for the other 37 species were downloaded from Ensembl (release 78) and NCBI database (Table S27 in Supporting Information). The longest transcript was chosen to represent each gene, and gene models with protein length less than 50 amino acids in the genomes were removed. Then, these protein sets were pooled, and self-to-self BLASTP was conducted for all protein sequences

aforementioned with an *E*-value of 1×10^{-5} and hit with identity less than 30% and coverage less than 30% were removed. Then, based on filtered BLASTP results, orthologous groups were constructed by ORTHOMCL v2.0.9 (Table S28 in Supporting Information). CAFÉ was used to identify significantly expanded and contracted gene families (*P*-value<0.05) (De Bie et al., 2006).

Phylogenetic tree inference and estimation of divergence time

860 single-copy gene families were retrieved from OrthoMCL as described above and used for the following phylogenetic tree reconstruction steps. The families containing any sequences <200 residues were removed, the amino acid sequences in each family were aligned using MUSCLE v3.8.3138 with default parameters (Edgar, 2004), and the corresponding coding sequence alignments were back-translated from corresponding amino acid sequence alignments using PAL2NAL v14 (Suyama et al., 2006). Families were further filtered if the CDS alignment length remained lower than 30% or 600 bp, resulting in 277 alignments from gene families. The 277 remaining CDS alignments were extracted and assembled into one supergene for each species using custom-made Perl scripts, and the supergenes were used for the following phylogenetic tree reconstruction and species divergence time estimation steps. Phylogenetic trees were constructed using MrBayes v3.2.4 (Ronquist and Huelsenbeck, 2003) and PhyML with the GTR+I+G model (Guindon et al., 2005). Divergence times were estimated under a relaxed clock model with an approximate likelihood calculation algorithm implemented using the MCMCTREE program in PAML4.7 (Yang, 2007). The fossil calibration times are listed in Table S29 in Supporting Information. For MCMCTREE, two runs were performed to check convergence.

Genomic signatures of skin cover evolution

Identification of alpha-keratin and keratin-associated protein (KRTAP) genes

Alpha-keratin protein sequences of human, mouse, and dog from Ensembl release 78 were used as reference sequences. Alpha-keratin genes of 39 mammal species were then identified using maker2 pipeline (Table S30 in Supporting Information). To avoid the incorporation of other genes, we adopted two steps to filter out results. First, because the lengths of all the reference sequences are longer than 200 amino acids, we filtered sequences with lengths shorter than 200 amino acids. Then, a phylogenetic tree was constructed and used to filter out falsely identified genes (Figure S4 in Supporting Information). The filtered sequences were finally classified according to the phylogenetic tree. For the construction of the phylogenetic tree, all amino acid sequences were aligned using MUSCLE v3.8.3138 (Edgar, 2004) with default parameters, and the corresponding coding sequence alignments were back-translated from the corresponding amino acid sequence alignments using pal2nal v14. Then, trees were constructed by RAxML v8.2.7 (Stamatakis, 2006) with a GTR+I+G model. The embellishment of the tree was performed by iTOL. The identification and classification of keratin (KRTAP) genes have followed the method for alpha-keratin, except that we filtered sequences with lengths shorter than 30 amino acids (Table S31 and Figure S5 in Supporting Information).

Positively selected genes

Eight species (Mcul (Manis culionensis), Hbra (Hystrix brachyurus), Bacu (Balaenoptera acutorostrata), Dnov (Dasypus novemcinctus), Eeur (Erinaceus europaeus), Etel (Echinops telfairi), Mpen (Manis pentadactyla), Tman (Trichechus manatus latirostris)) with non-fair skin covers were selected as concerned species. For each concerned species, the subtree containing the concerned species (as foreground) and other species (as background) (Table S4 in Supporting Information) was extracted from the species tree containing all the 39 species, then the single-copy genes of selected species were obtained from orthologous gene clusters identified by OrthoMCL, and finally, the PSGs in the concerned species were identified (Table S5 in Supporting Information). During the identification of PSGs for each concerned species, we only used a limited number of background species rather than all the non-concerned species, as the fewer species we used, the more single-copy genes we will get, which ensures more genes could be subject to test. The subtrees were manually created according to their phylogeny.

For each single-copy gene, amino acid sequences were aligned using MUSCLE v3.8.3138 (Edgar, 2004) with default parameters, and corresponding coding sequence alignments were back-translated from the corresponding amino acid sequence alignments using pal2nal v14. The PSGs were identified using the branch-site model implanted by the CODEML from PAML4.7 (Yang, 2007). The concerned species was set as the foreground branch and others as background branches. Likelihood ratio tests were performed using the χ^2 statistic to calculate *P*-values, which were then corrected with Bonferroni methods. Genes that met the requirements of the corrected *P*-value<0.05 and with at least one positively selected site were selected as PSGs.

Common PSGs among eight concerned species were defined as a gene that was identified as PSG in at least four of eight concerning species. Common PSGs among eight concerned species were then annotated using human sequences (Table S6 in Supporting Information). Significantly over-represented GO terms among these PSGs in Philippine pangolin (PA, Mcul) and porcupine (Hbra) were identified using the topGO (http://www.bioconductor.org/packages/

release/bioc/html/topGO.html) package (BF, biological process; CC, cellular component; MF, molecular function) in R programming language (https://www.r-project.org/). Significantly over-represented GO terms were identified with *P*-values of ≤0.01 and were viewed using REViGO (http://revigo.irb.hr/).

Rapidly evolving genes

Data sets used for the identification of REGs were the same as those for the identification of PSGs. ω (Ka/Ks) ratios were calculated using the branch model of CODEML in PAML4.7 (Yang, 2007). The gene was assigned as a REG when the foreground ω (concerned species) was higher than the background ω (other species). Common REGs among eight concerned species were defined as a gene identified as a REG in at least four of eight concerning species. The common REGs among eight concerned species were then annotated using human sequences.

Convergently evolving genomic regions

To identify CEGRs in mammalian genomes with similar skin cover phenotypes 18 species were selected, including Mcul, Mpen, Hbra, Dnov, *Bacu*, Tman, Oaf (*Orycteropus afer*), Etel, Eeur, Clup (*Canis lupus familiaris)*, Umar (*Ursus maritimus)*, Oari (*Ovis aries*), Mluc (*Myotis lucifugus*), Sara (*Sorex araneus*), Cpor (*Cavia porcellus*), Mmus (*Mus musculus*), Cjac (*Callithrix jacchus*), and Hsap (*Homo sapiens*). These 18 species were further classified into three concerned groups with cover phenotypes and one background group. The three groups included an (i) spiny group containing porcupine, Etel, and Eeur; (ii) scaled group containing Mcul, *Mpen*, and Dnov; and (iii) hairless group containing Bacu and Tman. A background group is a hair group containing the remaining ten species.

The pairwise alignment between human and the other 17 species were performed by lastz (http://www.bx.psu.edu/miller_lab/). Lastz outputs in the axt format were chained by the axtChain program. The chained alignments were processed into nets with chainNet and netSyntenic (http://hgdownload.cse.ucsc.edu/admin/jksrc.zip). Best-chain alignments in axt format were extracted by the netToAxt program. These whole-genome alignments were prepared for multiz, which generates 18-way multiple genome alignments of human and other 17 species.

For the spiny group, genomic alignment sequences, including spiny group and background group (hair group), were extracted from an 18-way maf file using 1,000 bp sliding window size with 500 bp step size according to human chromosomes. The extracted block is filtered if it contains less than two concerned species or the convergent tree is the same as the species tree (Figure 3A). Each remaining extracted window was aligned with muscle, and a phylogenetic tree was constructed using RAxML with a GTR+I+G model (Stamatakis, 2006). The resultant trees were filtered with the following criteria: (i) species in the spiny group clustered into one clade; and (ii) topology of species from the background group was consistent with the species tree.

Genes crossing with CEGRs were identified using a rigorous criterion (overlap≥500 bp or ≥30% coverage) and loose criteria (overlap≥100 bp) respectively and annotated, which could be used to detect the functional influence of these CEGRs.

Convergent amino acid substitutions

Convergence can also occur at the protein sequence level. CAAS refer to changes from different ancestral amino acids to the same descendant amino acid along independent evolutionary lineages. However, this type only included sites where changes resulted in identical amino acid states in present-day species. These changes were hypothesized to be possible indicators of convergent evolution if adaptation to similar environments could be accomplished via multiple different amino acids at the same position.

To identify genes with convergent, we selected three groups of species. The three groups included three concerned skin cover phenotypes as in the analysis of CEGRs identification (Figure S6 in Supporting Information; Figure 3A). For the spiny group, each of the three spiny mammals, including Etel, Eeur and Hbra, the extant sequences at each position were compared to the ancestral sequence at the node corresponding to the most recent ancestor. Reconstruction of ancestral sequences was conducted for single-copy orthologs using the codeml program in PAMLv4.8 (Yang, 2007). For Etel, this node was the one shared with Oafe, for Eeur, this node was the one shared with Sara, and for Mcul, this node was the one shared with Cpor. Ancestral nodes are those at the roots of the black nodes in Figure S1 in Supporting Information. We identified amino acid positions for which changes were inferred to have occurred and further examined those positions that changed in more than one mammal. All three mammals could have shared these changes. Changes were also classified as four types: C, convergent amino acid substitutions; D, divergent amino acid substitutions, refer to changes from different ancestral amino acids to the different descendant amino acid along independent evolutionary lineages; and V, variants amino acid substitutions, refer to changes from the same ancestral amino acid to the different descendant amino acid along independent evolutionary lineages.

Data availability

The raw genome datasets are deposited in NCBI (PRJNA490788 and PRJNA490787). The full gene list involved in this manuscript could be found in Table S32 in Supporting Information.

Compliance and ethics *The author(s) declare that they have no conflict of interest.*

Acknowledgements *This work was supported by the National Natural Science Fund for Outstanding Youth Fund (31222009, 31272295, 31570489, 81622048 and 81473377), the Project of Quality Guarantee System of Chinese Herbal Medicines (201507002), Science Foundation for Distinguished Young Scholars of Jiangsu Province (BK20140049) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).*

References

Acevedo-Whitehouse, K., and Duffus, A.L.J. (2009). Effects of environmental change on wildlife health. Phil Trans R Soc B 364, 3429–3438.

Ajiro, M., Nishidate, T., Katagiri, T., and Nakamura, Y. (2010). Critical involvement of RQCD1 in the EGFR-Akt pathway in mammary carcinogenesis. Int J Oncol 37, 1085–1093.

Alibardi, L. (2003). Adaptation to the land: the skin of reptiles in comparison to that of amphibians and endotherm amniotes. J Exp Zool 298B, 12–41.

Baker, J. (1992). Skin disease in wild cetaceans from British waters. Aquat Mamm 18, 27–32.

Bandyopadhyay, A., Tsuji, K., Cox, K., Harfe, B.D., Rosen, V., and Tabin, C.J. (2006). Genetic analysis of the roles of BMP2, BMP4, and BMP7 in limb patterning and skeletogenesis. PLoS Genet 2, e216.

Banerji, R., Eble, D.M., Iovine, M.K., and Skibbens, R.V. (2016). Esco2 regulates *cx43* expression during skeletal regeneration in the zebrafish fin. Dev Dyn 245, 7–21.

Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. Mobile DNA 6, 11.

Barlow, J. (1984). Xenarthrans and pholidotes. In: Anderson, S., and Jones Jr, J.K., eds. Orders and Families of Recent Mammals of the World. New York: John Wiley & Sons. 219–239.

Bazzi, H., Fantauzzo, K.A., Richardson, G.D., Jahoda, C.A.B., and Christiano, A.M. (2007). Transcriptional profiling of developing mouse epidermis reveals novel patterns of coordinated gene expression. Dev Dyn 236, 961–970.

Benrebeh, I., Grati, M., Bonnet, C., Bouassida, W., Hadjamor, I., Ayadi, H., Ghorbel, A., Petit, C., and Masmoudi, S. (2016). Genetic analysis of Tunisian families with Usher syndrome type 1: toward improving early molecular diagnosis. Mol Vis 22, 827–835.

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

Bernstein, C., and Bernstein, H. (2015). Epigenetic reduction of DNA repair in progression to cancer. In: Chen, C., ed. Advances in DNA Repair. London: IntechOpen.

Bertulli, C., Cecchetti, A., van Bressem, M., and Van Waerebeek, K. (2012). Skin disorders in common minke whales and white-beaked dolphins off Iceland, a photographic assessment. J Marine Anim Ecol 5: 29–40.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and genomewise. Genome Res 14, 988–995.

Blume-Peytavi, U., Whiting, D.A., and Trüeb, R.M. (2008). Hair Growth and Disorders. Heidelberg: Springer.

Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezhuk, Y., et al. (2013). BLAST: a more efficient report with usability improvements. Nucleic Acids Res 41, W29–W33.

Botchkarev, V.A., and Fessing, M.Y. (2005). Edar signaling in the control of hair follicle development. J Invest Dermatol Sympos Proc 10, 247–251.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J Mol Biol 268, 78–94.

Bush, J.O., Lan, Y., and Jiang, R. (2004). The cleft lip and palate defects in Dancer mutant mice result from gain of function of the Tbx10 gene. Proc Natl Acad Sci USA 101, 7022–7027.

Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sanchez Alvarado, A., and Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18, 188–196.

Capecchi, M. (1997). *Hox* genes and mammalian development. Cold Spring Harb Symp Quant Biol, 62: 273–281.

Carpenter, E.M., Goddard, J.M., Davis, A.P., Nguyen, T.P., and Capecchi, M.R. (1997). Targeted disruption of Hoxd-10 affects mouse hindlimb development. Development 124, 4505–4514.

Chen, J., and Chuong, C.M. (2012). Patterning skin by planar cell polarity: the multi-talented hair designer. Exp Dermatol 21, 81–85.

Chereau, D., Kerff, F., Graceffa, P., Grabarek, Z., Langsetmo, K., and Dominguez, R. (2005). Actin-bound structures of Wiskott-Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly. Proc Natl Acad Sci USA 102, 16644–16649.

Chernova, O.F., and Kuznetsov, G.V. (2001). Structural features of spines in some rodents (Rodentia: Myomorpha, Hystricomorpha). Biol Bull Rus Acad Sci 28, 371–383.

Choo, S.W., Rayko, M., Tan, T.K., Hari, R., Komissarov, A., Wee, W.Y., Yurchenko, A.A., Kliver, S., Tamazian, G., Antunes, A., et al. (2016). Pangolin genomes and the evolution of mammalian scales and immunity. Genome Res 26, 1312–1322.

Chrostek, A., Wu, X., Quondamatteo, F., Hu, R., Sanecka, A., Niemann, C., Langbein, L., Haase, I., and Brakebusch, C. (2006). Rac1 is crucial for hair follicle integrity but is not essential for maintenance of the epidermis. Mol Cell Biol 26, 6957–6970.

Chuong, C.M., and Homberger, D.G. (2003). Development and evolution of the amniote integument: current landscape and future horizon. J Exp Zool 298B, 1–11.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676.

Coutton, C., Vargas, A.S., Amiri-Yekta, A., Kherraf, Z.E., Ben Mustapha, S.F., Le Tanno, P., Wambergue-Legrand, C., Karaouzène, T., Martinez, G., Crouzy, S., et al. (2018). Mutations in CFAP43 and CFAP44 cause male infertility and flagellum defects in Trypanosoma and human. Nat Commun 9, 686.

Dai, D., Zhu, H., Wlodarczyk, B., Zhang, L., Li, L., Li, A.G., Finnell, R.H., Roop, D.R., and Chen, J. (2011). Fuz controls the morphogenesis and differentiation of hair follicles through the formation of primary cilia. J Invest Dermatol 131, 302–310.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics 22, 1269–1271.

Devasagayam, T., Tilak, J., Boloor, K., Sane, K.S., Ghaskadbi, S.S., and Lele, R. (2004). Free radicals and antioxidants in human health: current status and future prospects. J Assoc Physicans India 52, 794–804.

Devenport, D., and Fuchs, E. (2008). Planar polarization in embryonic epidermis orchestrates global asymmetric morphogenesis of hair follicles. Nat Cell Biol 10, 1257–1268.

Di Lucca, J., Guedj, M., Lacapère, J.J., Fargnoli, M.C., Bourillon, A., Dieudé, P., Dupin, N., Wolkenstein, P., Aegerter, P., Saiag, P., et al. (2009). Variants of the xeroderma pigmentosum variant gene (*POLH*) are associated with melanoma risk. Eur J Cancer 45, 3228–3236.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792–1797.

Etienne-Manneville, S., and Hall, A. (2003). Cdc42 regulates GSK-3β and adenomatous polyposis coli to control cell polarity. Nature 421, 753–756.

Feldhamer, G.A. (2007). Mammalogy: Adaptation, Diversity, Ecology. Maryland: Johns Hopkins University Press.

Foote, A.D., Liu, Y., Thomas, G.W.C., Vinař, T., Alföldi, J., Deng, J., Dugan, S., van Elk, C.E., Hunter, M.E., Joshi, V., et al. (2015). Convergent evolution of the genomes of marine mammals. Nat Genet

47, 272–275.
Franklin, C.C., Backos, D.S., Mohar, I., White, C.C., Forman, H.J., and Kavanagh, T.J. (2009). Structure, function, and post-translational regulation of the catalytic and modifier subunits of glutamate cysteine ligase. Mol Aspects Med 30, 86–98.
Gendron-Maguire, M., Mallo, M., Zhang, M., and Gridley, T. (1993). *Hoxa-2* mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. Cell 75, 1317–1331.
Geraci, J., St Aubin, D., and Hicks, B. (1986). The epidermis of odontocetes: a view from within. In: Bryden, M.M., and Harrison, R., eds. Research on Dolphins. Oxford: Oxford University Press. 3–21.
Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005). PHYML Online—A web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33, W557–W559.
Gul, I.S., Hulpiau, P., Saeys, Y., and van Roy, F. (2017). Evolution and diversity of cadherins and catenins. Exp Cell Res 358, 3–9.
Guo, N., Hawkins, C., and Nathans, J. (2004). Frizzled6 controls hair patterning in mice. Proc Natl Acad Sci USA 101, 9277–9281.
Hahn, C., Bachmann, L., and Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Res 41, e129.
Haq, S., Michael, A., Andreucci, M., Bhattacharya, K., Dotto, P., Walters, B., Woodgett, J., Kilter, H., and Force, T. (2003). Stabilization of β-catenin by a Wnt-independent mechanism regulates cardiomyocyte growth. Proc Natl Acad Sci USA 100, 4610–4615.
Hasegawa, M., Higashi, K., Yokoyama, C., Yamamoto, F., Tachibana, T., Matsushita, T., Hamaguchi, Y., Saito, K., Fujimoto, M., and Takehara, K. (2013). Altered expression of dermokine in skin disorders. J Eur Acad Dermatol Venereol 27, 867–875.
Hemnani, T., and Parihar, M. (1998). Reactive oxygen species and oxidative DNA damage. Indian J Physiol Pharm 42, 440–452.
Hoey, K.A., Wise, R.R., and Adler, G.H. (2004). Ultrastructure of echimyid and murid rodent spines. J Zool 263, 307–315.
Hogan, B.L. (1996). Bone morphogenetic proteins: multifunctional regulators of vertebrate development. Genes Dev 10, 1580–1594.
Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinf 12, 491.
Hong, Y.R., Chen, C.H., Chang, J.H., Wang, S., Sy, W.D., Chou, C.K., and Howng, S.L. (2000). Cloning and characterization of a novel human ninein protein that interacts with the glycogen synthase kinase 3β. Biochim Biophys Acta 1492, 513–516.
Huelsken, J., Vogel, R., Erdmann, B., Cotsarelis, G., and Birchmeier, W. (2001). β-Catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. Cell 105, 533–545.
Hulpiau, P., and van Roy, F. (2009). Molecular evolution of the cadherin superfamily. Int J Biochem Cell Biol 41, 349–369.
Jadeja, S., Smyth, I., Pitera, J.E., Taylor, M.S., van Haelst, M., Bentley, E., McGregor, L., Hopkins, J., Chalepakis, G., Philip, N., et al. (2005). Identification of a new gene mutated in Fraser syndrome and mouse myelencephalic blebs. Nat Genet 37, 520–525.
Jansson, L., Kim, G.S., and Cheng, A.G. (2015). Making sense of Wnt signaling—linking hair cell regeneration to development. Front Cell Neurosci 9.
Jung, Y.S., Liu, G., and Chen, X. (2010). Pirh2 E3 ubiquitin ligase targets DNA polymerase eta for 20S proteasomal degradation. Mol Cell Biol 30, 1041–1048.
Khan, I., Maldonado, E., Vasconcelos, V., O'Brien, S.J., Johnson, W.E., and Antunes, A. (2014). Mammalian keratin associated proteins (KRTAPs) subgenomes: disentangling hair diversity and adaptation to terrestrial and aquatic environments. BMC Genomics 15, 779.
Kim, E.B., Fang, X., Fushan, A.A., Huang, Z., Lobanov, A.V., Han, L., Marino, S.M., Sun, X., Turanov, A.A., Yang, P., et al. (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature 479, 223–227.
Kim, T.G., Kim, M., Lee, J.J., Kim, S.H., Je, J.H., Lee, Y., Song, M.J., Choi, Y., Chung, Y.W., Park, C.G., et al. (2015). CCCTC-binding factor controls the homeostatic maintenance and migration of Langerhans cells. J Allergy Clin Immunol 136, 713–724.
Kirkegaard, S.S., Wulff, T., Gammeltoft, S., and Hoffmann, E.K. (2013). KCNK5 is functionally down-regulated upon long-term hypotonicity in ehrlich ascites tumor cells. Cell Physiol Biochem 32, 1238–1246.
Lai-Cheong, J.E., and McGrath, J.A. (2009). Structure and function of skin, hair and nails. Medicine 37, 223–226.
Leistra, W., Hoyer, M., Kik, M., and Sinke, J. (2003). Recurrent skin problems in a manatee: the diagnostic approach. Tijdschr Diergeneeskd 128, 140–144.
Lewandoski, M., Sun, X., and Martin, G.R. (2000). Fgf8 signalling from the AER is essential for normal limb development. Nat Genet 26, 460–463.
Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., et al. (2010). *De novo* assembly of human genomes with massively parallel short read sequencing. Genome Res 20, 265–272.
Lu, S.C. (2009). Regulation of glutathione synthesis. Mol Aspects Med 30, 42–59.
Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20, 2878–2879.
Marsh, H., O'Shea, T.J., and Reynolds III, J.E. (2011). Ecology and Conservation of the Sirenia: Dugongs and Manatees. Cambridge: Cambridge University Press.
Martinez-Levasseur, L.M., Gendron, D., Knell, R.J., O'Toole, E.A., Singh, M., and Acevedo-Whitehouse, K. (2011). Acute sun damage and photoprotective responses in whales. Proc R Soc B 278, 1581–1586.
Meyer, W., Liumsiricharoen, M., Suprasert, A., Fleischer, L.G., and Hewicker-Trautwein, M. (2013). Immunohistochemical demonstration of keratins in the epidermal layers of the Malayan pangolin (Manis javanica), with remarks on the evolution of the integumental scale armour. Eur J Histochem 57, 27.
Moll, R., Schiller, D.L., and Franke, W.W. (1990). Identification of protein IT of the intestinal cytoskeleton as a novel type I cytokeratin with unusual properties and expression patterns. J Cell Biol 111, 567–580.
Naso, M.F., Liang, B., Chris Huang, C., Song, X.Y., Shahied-Arruda, L., Belkowski, S.M., D'Andrea, M.R., Polkovitch, D.A., Lawrence, D.R., Griswold, D.E., et al. (2007). Dermokine: an extensively differentially spliced gene expressed in epithelial cells. J Invest Dermatol 127, 1622–1631.
Nelson, D.S., Halteren, A., Quispel, W.T., Bos, C., Bovée, J.V.M.G., Patel, B., Badalian-Very, G., Hummelen, P., Ducar, M., Lin, L., et al. (2015). *MAP2K1* and *MAP3K1* mutations in langerhans cell histiocytosis. Genes Chromosomes Cancer 54, 361–368.
Orange, J.S., Stone, K.D., Turvey, S.E., and Krzewski, K. (2004). The Wiskott-Aldrich syndrome. Cell Mol Life Sci 61, 2361–2385.
Panteleyev, A.A., Paus, R., Ahmad, W., Sundberg, J.P., and Christiano, A.M. (1998). Molecular and functional aspects of the hairless (*hr*) gene in laboratory rodents and humans. Exp Dermatol 7, 249–267.
Panteleyev, A.A., Botchkareva, N.V., Sundberg, J.P., Christiano, A.M., and Paus, R. (1999). The role of the hairless (*hr*) gene in the regulation of hair follicle catagen transformation. Am J Pathol 155, 159–171.
Po-Chedley, D.S., and Shadle, A.R. (1955). Pelage of the porcupine, Erethizon dorsatum dorsatum. J Mammal 36, 84–95.
Premi, S., Wallisch, S., Mano, C.M., Weiner, A.B., Bacchiocchi, A., Wakamatsu, K., Bechara, E.J.H., Halaban, R., Douki, T., and Brash, D.E. (2015). Chemiexcitation of melanin derivatives induces DNA photoproducts long after UV exposure. Science 347, 842–847.
Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). *De novo* identification of repeat families in large genomes. Bioinformatics 21, i351–i358.
Prum, R.O., and Torres, R.H. (2004). Structural colouration of mammalian skin: convergent evolution of coherently scattering dermal collagen arrays. J Exp Biol 207, 2157–2172.
Pucci, S., Fisco, T., and Zonetti, M.J. (2012). XRCC5 (X-ray repair complementing defective repair in Chinese hamster cells 5 (double-

strand-break rejoining)). Atlas Genets Cytogenets Oncol Haematol doi: 10.4267/2042/48234.
Ravni, A., Qu, Y., Goffinet, A.M., and Tissir, F. (2009). Planar cell polarity cadherin *Celsr1* regulates skin hair patterning in the mouse. J Invest Dermatol 129, 2507–2509.
Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Rosas, F.C.W. (1994). Biology, conservation and status of the Amazonian manatee *Trichechus inunguis*. Mammal Rev 24, 49–59.
Samuelov, L., Sprecher, E., and Paus, R. (2015). The role of P-cadherin in skin biology and skin pathology: lessons from the hair follicle. Cell Tissue Res 360, 761–771.
Segade, F. (2010). Glucose transporter 10 and arterial tortuosity syndrome: the vitamin C connection. FEBS Lett 584, 2990–2994.
Sen, G.L., Reuter, J.A., Webster, D.E., Zhu, L., and Khavari, P.A. (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. Nature 463, 563–567.
Shapiro, M.D., Kronenberg, Z., Li, C., Domyan, E.T., Pan, H., Campbell, M., Tan, H., Huff, C.D., Hu, H., Vickrey, A.I., et al. (2013). Genomic diversity and evolution of the head crest in the rock pigeon. Science 339, 1063–1067.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.
Simons, M., Gloy, J., Ganner, A., Bullerkotte, A., Bashkurov, M., Krönig, C., Schermer, B., Benzing, T., Cabello, O.A., Jenny, A., et al. (2005). Inversin, the gene product mutated in nephronophthisis type II, functions as a molecular switch between Wnt signaling pathways. Nat Genet 37, 537–543.
Smit, A.F., Hubley, R., and Green, P. (1996). RepeatMasker Open-3.0.
Song, H., Hu, J., Chen, W., Elliott, G., Andre, P., Gao, B., and Yang, Y. (2010). Planar cell polarity breaks bilateral symmetry by controlling ciliary positioning. Nature 466, 378–382.
Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res 34, W435–W439.
Stary, A., and Sarasin, A. (2001). XPC (xeroderma pigmentosum, complementation group C).
Stewart, G.S., Wang, B., Bignell, C.R., Taylor, A.M.R., and Elledge, S.J. (2003). MDC1 is a mediator of the mammalian DNA damage checkpoint. Nature 421, 961–966.
Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34, W609–W612.
Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. System Biol 56, 564–577.
van Bressem, M.F., Reyes, J.C., Félix, F., Echegaray, M., Siciliano, S., Di Beneditto, A.P., Flach, L., Viddi, F., Avila, I.C., Herrera, J.C., et al. (2007). A preliminary overview of skin and skeletal diseases and traumata in small cetaceans from South American waters. Lat Am J Aquat Mamm 6, 7–42.
van Bressem, M.F., Raga, A.J., Di Guardo, G., Jepson, P.D., Duignan, P.J., Siebert, U., Barrett, T., Santos, M.O., Moreno, I.B., Siciliano, S., et al. (2009). Emerging infectious diseases in cetaceans worldwide and the possible role of environmental stressors. Dis Aquat Org 86, 143–157.
Wang, Y., Chang, H., and Nathans, J. (2010). When whorls collide: the development of hair patterns in frizzled 6 mutant mice. Development 137, 4091–4099.
Wilson, B., Arnold, H., Bearzi, G., Fortuna, C.M., Gaspar, R., Ingram, S., Liret, C., Pribanic, S., Read, A.J., Ridoux, V., et al. (1999). Epidermal diseases in bottlenose dolphins: impacts of natural and anthropogenic factors. Proc R Soc Lond B 266, 1077–1083.
Wood, R.D., Mitchell, M., Sgouros, J., and Lindahl, T. (2001). Human DNA repair genes. Science 291, 1284–1289.
Wu, X., Quondamatteo, F., Lefever, T., Czuchra, A., Meyer, H., Chrostek, A., Paus, R., Langbein, L., and Brakebusch, C. (2006). Cdc42 controls progenitor cell differentiation and beta-catenin turnover in skin. Genes Dev 20, 571–585.
Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res 35, W265–W268.
Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24, 1586–1591.
Yang, Y., He, S., Wang, Q., Li, F., Kwak, M.J., Chen, S., O'Connell, D., Zhang, T., Pirooz, S.D., Jeon, Y.H., et al. (2016). Autophagic UVRAG promotes UV-induced photolesion repair by activation of the CRL4 DDB2 E3 ligase. Mol Cell 62, 507–519.
Yim, H.S., Cho, Y.S., Guang, X., Kang, S.G., Jeong, J.Y., Cha, S.S., Oh, H. M., Lee, J.H., Yang, E.C., Kwon, K.K., et al. (2014). Minke whale genome and aquatic adaptation in cetaceans. Nat Genet 46, 88–92.
Yoshida, K., and Miki, Y. (2004). Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. Cancer Sci 95, 866–871.
Yu, M., Finner, A., Shapiro, J., Lo, B., Barekatain, A., and McElwee, K.J. (2006). Hair follicles and their role in skin health. Expert Rev Dermatol 1, 855–871.
Zhou, J., Liu, M., Fleming, A.M., Burrows, C.J., and Wallace, S.S. (2013). Neil3 and NEIL1 DNA glycosylases remove oxidative damages from quadruplex DNA and exhibit preferences for lesions in the telomeric sequence context. J Biol Chem 288, 27263–27272.
Zhu, H., Enaw, J.O.E., Ma, C., Shaw, G.M., Lammer, E.J., and Finnell, R. H. (2007). Association between *CFL1* gene polymorphisms and spina bifida risk in a California population. BMC Med Genet 8, 12.

SUPPORTING INFORMATION

The supporting information is available online at https://doi.org/10.1007/s11427-020-1841-5. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.