

PRIMARY RESEARCH

Open Access

Evidence that DNA repair genes, a family of tumor suppressor genes, are associated with evolution rate and size of genomes



Konstantinos Voskarides^{1*} , Harsh Dweep² and Charalambos Chrysostomou³

Abstract

Adaptive radiation and evolutionary stasis are characterized by very different evolution rates. The main aim of this study was to investigate if any genes have a special role to a high or low evolution rate. The availability of animal genomes permitted comparison of gene content of genomes of 24 vertebrate species that evolved through adaptive radiation (representing high evolutionary rate) and of 20 vertebrate species that are considered as living fossils (representing a slow evolutionary rate or evolutionary stasis). Mammals, birds, reptiles, and bony fishes were included in the analysis. Pathway analysis was performed for genes found to be specific in adaptive radiation or evolutionary stasis respectively. Pathway analysis revealed that DNA repair and cellular response to DNA damage are important (false discovery rate = 8.35×10^{-5} ; 7.15×10^{-6} , respectively) for species evolved through adaptive radiation. This was confirmed by further genetic in silico analysis ($p = 5.30 \times 10^{-3}$). Nucleotide excision repair and base excision repair were the most significant pathways. Additionally, the number of DNA repair genes was found to be linearly related to the genome size and the protein number (proteome) of the 44 animals analyzed ($p < 1.00 \times 10^{-4}$), this being compatible with Drake's rule. This is the first study where radiated and living fossil species have been genetically compared. Evidence has been found that cancer-related genes have a special role in radiated species. Linear association of the number of DNA repair genes with the species genome size has also been revealed. These comparative genetics results can support the idea of punctuated equilibrium evolution.

Keywords: Genomics, Evolutionary genetics, Natural selection, Rapid evolution, Speciation, Mutagenesis rate, Evolutionary medicine, Molecular evolution

Background

Adaptive radiation is a well-known phenomenon in evolutionary biology, where a taxon is split in multiple species which become adapted in a variety of environments in short evolutionary time. Although this phenomenon is mostly known in islands like the great examples of Darwin finches [1] and the Hawaiian drosophilas, other major adaptive radiations have occurred in other animals like cichlids, bats, and cetaceans [2–5]. It is very likely that common evolutionary and molecular processes have been followed in all taxa that have experienced adaptive radiation [6, 7]. No such common molecular pathways have been identified so far.

We could consider living fossil species and adaptive radiation as two very different evolutionary strategies: slow evolutionary rate versus rapid evolutionary rate respectively. Living fossils are characterized by morphological stasis, low taxonomic diversity, and certain rareness. Quantitative criteria have been published recently [8, 9]. The apparent absence of diversification and their morphological stability suggest highly effective adaptations that reduce the need for phenotypic change, regardless of environmental or genetic changes [8, 10]. Living fossils are frequently referred to as an example of evolutionary success and evolutionary stasis [11, 12]. Evolutionary stasis is a common finding in the fossil record [13]. The punctuated equilibrium theory of evolution is based on these fossil observations [14, 15]. Characteristic examples of taxa that are considered by most biologists as living

* Correspondence: kvoskar@ucy.ac.cy

¹Medical School, University of Cyprus, Kallipoleos 75, 1678 Nicosia, Cyprus
Full list of author information is available at the end of the article



fossils are the crocodylians, coelacanths, and ornithorhynchus. Like in the case of adaptive radiation, our knowledge is insufficient for any special genes that are under selection in living fossil species.

This study was mainly aiming at the identification of any common molecular pathways that contributed to a special evolutionary process in animals. We are mostly interested on genes that are related with disease, since evolutionary studies may contribute to a better understanding of the function of those genes. We supposed that living fossil species (LF) and radiated species (R, those that have been evolved through adaptive radiation) represent two animal categories with a very different rate and form of evolution. We took advantage of the plentiful animal genomes that have been sequenced since presently, and we performed an analytical comparative genetics study. Strict inclusion and statistical criteria were applied (see the “Methods” section). In total, 20 LF and 24 R vertebrate genomes (bony fishes, reptiles, birds, mammals) have been analyzed. Interestingly, only one major genetic difference was revealed related to DNA repair genes, one of the most important categories of tumor suppressor genes.

Methods

Species included in this study—genome data

The literature was carefully searched for all animal species that can be characterized as living fossils (LF) (slow evolutionary rate) or radiated (R) (they have experienced adaptive radiation). Additional inclusion criteria are as follows: species with a completed genome project, species with available annotation and gene symbol data (for reliable interspecies comparison). Annotation of genomes has been performed by the submitters under the same NCBI standards. We included animal classes with representative species in both living fossil species and radiated species for a reliable comparison. Genome and gene data used for this work are updated since April of 2019, according to Genome and Gene databases of NCBI (<https://www.ncbi.nlm.nih.gov/>). In total, 44 species were included in this analysis.

Gene analysis

Official gene symbols were used for comparison among species. A custom algorithm was developed for finding all common genes in the LF species group and in the R species group. Next, the two lists of common genes were compared. This was performed through the “unique values” function of Excel 2016. After comparison, two gene lists were created: genes that are common in LF but not found in R and genes that are common in R and not found in LF. We considered that these genes are probably associated with a special type of evolutionary process. Genes were analyzed under the concept of

presence/absence. Copy numbers were not considered. All gene lists can be found in Additional file 1: Table S1.

Pathway analysis and DNA repair gene analysis

Panther 14.1 online software [16, 17] was used for pathway analysis of the two LF and R unique gene lists. The software analyzes the submitted gene lists with reference to the human genome. Two algorithms of the software were used: pathway and reactome profile analysis. Results were compared between LF and R to find any pathways that are unique in any of the two evolutionary processes. False discovery rate (FDR) is the statistical outcome that is a special type of adjusted p value. Significant level alpha was set to 0.0001 for highly reliable results.

To confirm if DNA repair genes represent a major genetic difference between the two vertebrate categories, all 44 species' genomes were analyzed for their content in DNA repair genes. An updated list of all 151 known DNA repair genes was used [18]. Content analysis (presence/absence) was performed using the official gene symbols. An extra search was performed using the gene aliases for any missed misnamed genes. Content analysis was performed through the “duplicate values” function of Excel 2016. Results in detail can be found in Additional file 2: Table S2.

Statistical analysis

All statistical analysis needed for this work was performed through the statistical package STATAv.13 (StataCorp LLC, Texas, USA). The basic statistical analysis included univariate linear regression and independent t test (two-tailed). The heat map was performed through the “color gradient” function of Excel 2016. Significant level alpha was set to 0.01 for identifying the most significant categories of DNA repair genes.

Results and discussion

Species analyzed

Strict inclusion criteria were applied for the 44 species analyzed in this study. Several fossil and molecular studies that are cited below justify the classification “living fossil” or “radiated.” A more detailed description of “living fossil” species can be found in the book *Living Fossils* of [19]. Additionally, the 20 LF species satisfy the very accurate living fossil quantification system of [9]. Genome projects information can be found in Table 1.

The 20 LF species or taxa are as follows (common names, scientific names are found in Table 1): aardvark [20], platypus [21, 22], opossum [23, 24], elephant shrew [25], giant panda [26], koala [23, 27], Philippine tarsier [28], pelican [29], New Zealand wren [30, 31], speckled

Table 1 Living fossil (LF) vertebrate species and radiated (R) vertebrate species analyzed in this study, with genome and proteome information

Species	Genome size (Mb)	Protein number	Genome projects
Mammals—LF			
L1. <i>Orycteropus afer</i>	4566	25,544	https://www.ncbi.nlm.nih.gov/genome/annotation_euk/
L2. <i>Ornithorhynchus anatinus</i>	1924	24,786	[22, 73]
L3. <i>Monodelphis domestica</i>	3598	49,112	[74]
L4. <i>Elephantulus edwardii</i>	4066	25,209	https://www.ncbi.nlm.nih.gov/genome/annotation_euk/
L5. <i>Ailuropoda melanoleuca</i>	2364	36,506	[75, 76]
L6. <i>Phascolarctos cinereus</i>	3398	46,908	[77]
L7. <i>Carlito syrichta</i>	3454	33,081	[78]
Mammals—R			
R1. <i>Myotis brandtii</i>	2107	40,808	[79]
R2. <i>Pteropus alecto</i>	1986	39,227	[80]
R3. <i>Rousettus aegyptiacus</i>	1941	48,803	[81]
R4. <i>Myotis davidii</i>	2060	33,106	[80]
R5. <i>Hipposideros armiger</i>	2237	45,831	[82]
R6. <i>Myotis lucifugus</i>	2035	43,106	[83]
R7. <i>Pteropus vampyrus</i>	2198	43,628	https://www.ncbi.nlm.nih.gov/genome/annotation_euk/
R8. <i>Miniopterus natalensis</i>	1803	29,787	[84]
R9. <i>Eptesicus fuscus</i>	2027	49,822	https://www.ncbi.nlm.nih.gov/genome/annotation_euk/
R10. <i>Microcebus murinus</i>	2487	59,023	https://www.ncbi.nlm.nih.gov/genome/annotation_euk/
R11. <i>Propithecus coquereli</i>	2798	28,194	https://www.ncbi.nlm.nih.gov/genome/annotation_euk/
R12. <i>Tursiops truncatus</i>	2478	38,849	[85, 86]
R13. <i>Balaenoptera acutorostrata</i>	2432	37,625	[87, 88]
R14. <i>Physeter catodon</i>	2512	50,591	https://www.ncbi.nlm.nih.gov/genome/annotation_euk/
R15. <i>Orcinus orca</i>	2373	27,870	[85, 89, 90]
R16. <i>Lipotes vexillifer</i>	2429	26,901	[91]
R17. <i>Delphinapterus leucas</i>	2358	49,714	[92]
Birds and reptiles—LF			
L8. <i>Pelecanus crispus</i>	1161	16,298	[93]
L9. <i>Acanthisitta chloris</i>	1036	16,077	[93]
L10. <i>Colius striatus</i>	1076	15,797	[93]
L11. <i>Cariama cristata</i>	1132	16,125	[93]
L12. <i>Tinamus guttatus</i>	1047	17,873	[93]
L13. <i>Opisthocomus hoazin</i>	1203	14,878	[93]
L14. <i>Crocodylus porosus</i>	2085	28,676	[38, 94]
L15. <i>Alligator mississippiensis</i>	2162	42,388	[38, 94]
L16. <i>Gavialis gangeticus</i>	2415	27,294	[38, 94]
L17. <i>Alligator sinensis</i>	2271	43,105	[95]
Birds and reptiles—R			
R18. <i>Geospiza fortis</i>	1065	16,724	[67, 93, 96]
R19. <i>Parus major</i>	1020	39,666	[97, 98]
R20. <i>Anolis carolinensis</i>	1799	34,827	[69]
Bony fishes—LF			
L18. <i>Scleropages formosus</i>	742	32,859	[39, 99, 100]

Table 1 Living fossil (LF) vertebrate species and radiated (R) vertebrate species analyzed in this study, with genome and proteome information (*Continued*)

Species	Genome size (Mb)	Protein number	Genome projects
L19. <i>Lepisosteus oculatus</i>	946	41,647	[40]
L20. <i>Latimeria chalumnae</i>	2798	34,251	[101, 102]
Bony fishes—R			
R21. <i>Notothenia coriiceps</i>	637	31,979	[103]
R22. <i>Maylandia zebra</i>	957	46,173	[59, 104]
R23. <i>Pundamilia nyererei</i>	830	38,583	[59, 105]
R24. <i>Haplochromis burtoni</i>	831	44,653	[59, 105, 106]

mousebird [32], red-legged seriema [33], tinamou [34], hoatzin [35–37], crocodilians [38], arowana [39], spotted gar [40], and coelacanth [12, 41].

The 24 R species or taxa are as follows (common names, scientific names are found in Table 1): bats [42–44], dolphins and whales [45, 46], lemurs [47–49], medium ground finch [50, 51], great tit [51], Carolina anole [52–55], black rockcod [56–58], and three cichlid species [59–62].

Gene and pathway analysis

Evolutionary stasis and rapid evolutionary speciation can be characterized as opposite evolutionary procedures or at least very different evolutionary phenomena. This is the first study that compares genetically those two very different categories of vertebrate species. Gene or annotation information was inadequate for most invertebrate LF or R species, so they were not included in this study.

The procedure we followed is very simple. We downloaded the annotated genome information for all 44 species. Then, we found the common genes in LF species and the common genes in R species, creating two separate gene lists (Additional file 1: Table S1). The next step was to compare the two lists to find any genes that are common in LF but not found in R species and genes that are common in R but not found in LF species. We consider that these genes may be under selection since they are found only in species with a special evolutionary profile. In total, 1534 genes were found to be specific for LF species and 2263 genes to be specific for R species.

Analysis of the two final gene lists was performed by Panther 14.1 software, under two algorithms: pathways (biological processes) and reactome. We looked for unique biological processes and reactomes in LF- and R-specific genes respectively. Using the strict criterion of $FDR \leq 0.0001$, only one process/pathway was found to be significant in R-specific genes by both algorithms, this being DNA repair (DNA repair and cellular response to

DNA damage; $FDR = 8.35 \times 10^{-5}$ and 7.15×10^{-6} , respectively). Not any common significant pathways came out in the biological processes and reactome analyses for LF-specific genes. Step by step analysis and all analytical output can be found in Additional file 1: Table S1. The flowchart of analysis can be found in Table 2.

DNA repair gene analysis

In order to confirm the pathway analysis results, we analyzed the 44 genomes for their content in DNA repair genes, using a list of all known DNA repair genes since presently (updated list of Wood et al. [18]). Subcategories of DNA repair genes were also considered in the analysis. Results in detail can be found in Additional file 2: Table S2. The results highly confirmed the previously performed pathway analysis (Table 3). R species' genomes are significantly enriched in DNA repair genes ($p = 5.3 \times 10^{-3}$). The most significant subcategories are the nucleotide excision repair ($p = 5.00 \times 10^{-4}$) and base excision repair ($p = 9.80 \times 10^{-3}$). Many other subcategories seem to be significantly enriched in R species under the criterion of $p < 0.05$. Conserved DNA damage response and non-homologous end-joining are not significant at all (Table 3). A heat map diagram shows that indeed the R species' genomes are enriched in DNA repair genes in comparison with the LF species, especially for mammals, reptiles, and birds (Fig. 1).

The top 20 genes with the highest existence rate in R species in relation to LF species can be found in Additional file 2: Table S2. Eleven out of the top 20 (55%) are genes related with nucleotide excision repair and base excision repair. All gene rates are available in Additional file 2: Table S2.

Genome and proteome size analysis

Interestingly, the number of DNA repair genes is linearly related with the genome size and the number of proteins ($p < 1.00 \times 10^{-4}$). We used genome and proteome data

Table 2 Flowchart and main outcomes of each analysis performed in this study

Analysis	Outcome
Identification of all vertebrate species that can be characterized as living fossil or radiated species, with available whole genome sequencing data and complete gene annotation	20 living fossil species 24 radiated species
Genes in common per group	Living fossil species: 2861 genes in common Radiated species: 3590 genes in common
Genes in common per group, not found in the other group	Living fossil species: 1534 unique genes Radiated species: 2263 unique genes
Pathway (biological processes) and reactome analyses, unique ones	Living fossil species: 0 pathways, 2 reactomes Radiated species: 7 pathways, 2 reactomes
Significant process revealed by both algorithms	Living fossil species: None Radiated species: DNA repair and cellular response to DNA damage (FDR = 8.35×10^{-5} ; 7.15×10^{-6} , respectively)
Search for 151 known DNA repair genes in the 45 species' genomes Mean comparison analysis	More DNA repair genes in radiated species than in living fossil species ($p = 5.3 \times 10^{-3}$) Most significant gene subcategory: Nucleotide excision repair ($p = 5.00 \times 10^{-4}$)
Linear regression: DNA repair genes number vs genome size or protein number	Genome size/protein number is linearly related with the number of DNA repair genes ($p < 1.0 \times 10^{-4}$)

FDR false discovery rate

Table 3 Mean comparison (independent *t* test, two-tailed) between living fossil (LF) and radiated species (R), for each category of DNA repair genes and altogether (degrees of freedom, 42)

DNA repair gene category	Species group	Mean number	Std. dev.	[95% conf. interval]	<i>t</i> value	<i>p</i> value
Base excision repair (BER)	20 (LF)	16.75	3.274704	15.21739, 18.28261	-2.7067	9.80×10^{-3}
	24 (R)	18.92	1.976309	18.08214, 19.75119		
Conserved DNA damage response	20 (LF)	15.2	2.876401	13.8538, 16.5462	-2.2001	3.34×10^{-2}
	24 (R)	16.83	2.03591	15.97364, 17.69302		
Direct reversal of damage	20 (LF)	2.9	0.3077935	2.755948, 3.044052	-0.1872	0.8524
	24 (R)	2.92	0.2823299	2.797449, 3.035884		
DNA polymerases	20 (LF)	14.3	1.688974	13.50954, 15.09046	-2.4279	1.96×10^{-2}
	24 (R)	15.38	1.244553	14.84947, 15.90053		
Editing and processing nucleases	20 (LF)	6.5	1.100239	5.985072, 7.014928	-2.4341	1.93×10^{-2}
	24 (R)	7.25	0.9440892	6.851346, 7.648654		
Fanconi anemia	20 (LF)	13.55	0.9445132	13.10795, 13.99205	-2.2591	2.91×10^{-2}
	24 (R)	14.21	0.9770927	13.79574, 14.62092		
Homologous recombination	20 (LF)	21.2	2.261811	20.14144, 22.25856	-1.6880	9.88×10^{-2}
	24 (R)	22.33	2.180281	21.41268, 23.25399		
Mismatch excision repair (MMR)	20 (LF)	8.35	0.8127277	7.969632, 8.730368	-1.7706	8.39×10^{-2}
	24 (R)	8.79	0.8329709	8.439934, 9.143399		
Non-homologous end-joining	20 (LF)	6.55	0.6863327	6.228786, 6.871214	-0.8497	0.4003
	24 (R)	6.71	0.5500329	6.476075, 6.940592		
Nucleotide excision repair (NER)	20 (LF)	25.15	3.528456	23.49863, 26.80137	-3.8043	5.00×10^{-4}
	24 (R)	28.46	2.186503	27.53505, 29.38161		
All DNA repair genes	20 (LF)	130.45	14.56916	123.6314, 137.2686	-2.9417	5.30×10^{-3}
	24 (R)	141.79	10.99003	137.151, 146.4324		

(<https://www.ncbi.nlm.nih.gov/>) of the 44 vertebrate species (Fig. 2). The two linear associations are independently significant since genome size is not linearly related with the number of proteins (Fig. 2). It is well known that genome size is not related with organism complexity [63]; thus, we consider that this association is not due to increased complexity of large genomes. Not any association was found when genome size means of LF and R species were compared (results not shown).

This result may also explain Drake's rule. This is about the density of accumulated mutations per generation (mutagenesis rate) that is roughly inversely proportional to genome size [64–66]. Here, we found that larger genomes have more DNA repair genes (and possibly lower mutagenesis rate, if DNA errors are corrected at a higher rate) that may explain Drake's rule, being unexplained for years.

Why DNA repair genes

There is evidence that LF species are evolving slower than R species. Additionally, some data show that mutagenesis and nucleotide diversity [59, 67] may be higher in R species than in LF species and that some R species with huge bodies (whales) have duplicated DNA repair genes to be protected by cancer [68, 69]. According to these data, we could hypothesize that R species may be at risk due to high mutation load. This could be balanced with more DNA repair genes, repairing as much DNA damages as possible. It seems that DNA repair at the nucleotide level (nucleotide excision repair and base excision repair) is more important than other DNA repair pathways (Table 3, Additional file 2: Table S2). Another explanation is that LF species are probably more protected from spontaneous DNA changes since due to the vast evolutionary time that they exist, stabilizing selection has formed their genome in a way that they are protected from random DNA changes that could change their general morphological features. Certain genes in LF genomes may act in a canalizing way that keeps these species in a narrow state of development and evolution since they are evolutionary successful. R species are not characterized by those features, and probably they need more or certain DNA repair genes to continue to diversify under a non-deleterious mutagenesis rate. We could consider that this is the first evidence for genes related with punctuated equilibrium evolution (long evolutionary stasis followed by short speciation explosions) [14, 15].

The fact that the number of DNA repair genes is related with the genome and proteome size is quite logical since larger genomes need more protection from spontaneous mutagenesis. This is the first time that a class of genes has been associated with genome size and number of proteins in animals.

Conclusions

A big number of genomes have been compared under the prism of evolutionary stasis and adaptive radiation. The analysis concluded that DNA repair genes might play a previously unknown significant role in evolution. It seems that more DNA repair genes are found in vertebrate taxa that have experienced recent adaptive radiation. Additionally, DNA repair genes were found to be statistically associated with the genome size and protein number in vertebrates. DNA repair genes are considered as tumor suppressor genes. There is evidence that tumor suppressor genes are related to environmental adaptation in humans [70, 71] and selective pressures along the evolution of mammals [72]. We can imagine that certain evolutionary procedures may be DNA repair-dependent, this showing the way for future analyses and experiments.

Additional files

Additional file 1: Table S1. Pathway analysis by PANTHER (XLSX 137 kb)

Additional file 2: Table S2. DNA repair gene analysis (XLSX 95 kb)

Acknowledgements

Not applicable

Authors' contributions

KV conceived the study idea, analyzed and interpreted most of the data of this study, and prepared the first draft of the paper. HD contributed to genomic data retrieval and proofread the paper. CC performed a custom-made bioinformatics algorithm for the initial comparative genomics data analysis and proofread the paper. All authors read and approved the final manuscript.

Funding

Not any funding exists for this research.

Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹Medical School, University of Cyprus, Kallipoleos 75, 1678 Nicosia, Cyprus.

²The Wistar Institute, Philadelphia, PA, USA. ³The Cyprus Institute, Nicosia, Cyprus.

Received: 12 March 2019 Accepted: 20 May 2019

Published online: 07 June 2019

References

1. Barlow N. The voyage of the beagle. *Nature*. 1932;129(3255):439.
2. Emerson BC. Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Mol Ecol*. 2002;11(6):951–66.

3. Gavrilets S, Losos JB. Adaptive radiation: contrasting theory with data. *Science*. 2009;323(5915):732–7.
4. Ecology. Book review: the ecology of adaptive radiation. *Ecology*. 2002; 83(2):591–2.
5. Simões M, Breitzkreuz L, Alvarado M, Baca S, Cooper JC, Heins L, et al. The evolving theory of evolutionary radiations. *Trends Ecol Evol*. 2016; 31(1):27–34.
6. Kapralov MV, Votintseva AA, Filatov DA. Molecular adaptation during a rapid adaptive radiation. *Mol Biol Evol*. 2013;30(5):1051–9.
7. Zhou T-C, Irwin DM, Shen Y-Y, Zhang Y-P, Liang L, Pan X-W, et al. Adaptive evolution of the Hox gene family for development in bats and dolphins. *PLoS One*. 2013;8(6):e65944.
8. Combosch DJ, Lemer S, Ward PD, Landman NH, Giribet G. Genomic signatures of evolution in nautilus—an endangered living fossil. *Mol Ecol*. 2017;26(21):5923–38.
9. Bennett DJ, Sutton MD, Turvey ST. Quantifying the living fossil concept. *Palaeontol Electron*. 2018;21(1):1–25.
10. Kin A, Blazejowski B. The horseshoe crab of the genus *Limulus*: living fossil or stabilomorph? *PLoS One*. 2014;9(10):e108036.
11. Mans BJ, de Klerk D, Pienaar R, Latif AA. Nuttalliella namaqua: a living fossil and closest relative to the ancestral tick lineage: implications for the evolution of blood-feeding in ticks. *PLoS One*. 2011;6(8):e23675.
12. Schmutz J, Miyake T, Powers TP, Ruddle FH, Myers RM, Grimwood J, et al. Complete HOX cluster characterization of the coelacanth provides further evidence for slow evolution of its genome. *Proc Natl Acad Sci*. 2010;107(8):3622–7.
13. Hunt G, Hopkins MJ, Lidgard S. Simple versus complex models of trait evolution and stasis as a response to environmental change. *Proc Natl Acad Sci*. 2015; 112(16):4885–90.
14. Gould SJ, Eldredge N. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology*. 1977;2:115–51.
15. Bitterman ME. Phyletic differences in learning. *Am Psychol*. 1965;20:396–410.
16. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. PANTHER version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*. 2017;45(D1):D183–9.
17. Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, Thomas PD. Protocol update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat Protoc*. 2019;14(3):703–21.
18. Wood RD, Mitchell M, Lindahl T. Human DNA repair genes, 2005. *Mutat Res*. 2005;577(1–2):275–83.
19. Simon C. Living fossils. Eldredge N, Stanley SM, editors. *Science news* 2007; 121(17):284.
20. Perelman PL, Yang F, Robinson TJ, Harrison WR, Graphodatsky AS, Pardini AT, et al. Reciprocal chromosome painting among human, aardvark, and elephant (superorder Afrotheria) reveals the likely eutherian ancestral karyotype. *Proc Natl Acad Sci*. 2003;100(3):1062–6.
21. Musser AM. Review of the monotreme fossil record and comparison of palaeontological and molecular data. *Comp Biochem Physiol A Mol Integr Physiol*. 2003;136(4):927–42.
22. Warren WC, Hillier LDW, Marshall Graves JA, Birney E, Ponting CP, Grützner F, et al. Genome analysis of the platypus reveals unique signatures of evolution. *Nature*. 2008;453(7192):175–83.
23. Archer M, Beck RMD, Hand SJ, Weisbecker V, Godthelp H. Australia's oldest marsupial fossils and their biogeographical implications. *PLoS One*. 2008;3(3):e1858.
24. Palma RE, Spotorno AE. Molecular systematics of marsupials based on the rRNA 12S mitochondrial gene: the phylogeny of Didelphimorphia and of the living fossil Microbiotheriid *Dromiciops gliroides* Thomas. *Mol Phylogenet Evol*. 1999;13(3):525–35.
25. Robinson TJ, Fu B, Ferguson-Smith MA, Yang F. Cross-species chromosome painting in the golden mole and elephant-shrew: support for the mammalian clades Afrotheria and Afroinsectiphilia but not Afroinsectivora. *Proc R Soc B Biol Sci*. 2004;271(1547):1477–84.
26. Hunt RM, Jaeger M, Zhu Q, Dong W, Liu J, Ciochon RL, et al. The first skull of the earliest giant panda. *Proc Natl Acad Sci*. 2007;104(26):10932–7.
27. Black KH, Archer M, Hand SJ, Godthelp H. First comprehensive analysis of cranial ontogeny in a fossil marsupial—from a 15-million-year-old cave deposit in northern Australia. *J Vertebr Paleontol*. 2010;30(4):993–1011.
28. Rossie JB, Ni X, Beard KC. Cranial remains of an Eocene tarsier. *Proc Natl Acad Sci USA*. 2006;103(12):4381–5.
29. Louchart A, Tourment N, Carrier J. The earliest known pelican reveals 30 million years of evolutionary stasis in beak morphology. *J Ornithol*. 2011; 152(1):15–20.
30. Taylor Smith BL, McComish BJ, Hartig G, England R, Penny D, McLenachan PA (Trish), et al. New Zealand passerines help clarify the diversification of major songbird lineages during the Oligocene. *Genome Biol Evol* 2015; 7(11):2983–2995.
31. Worthy TH, Hand SJ, Nguyen JMT, Tennyson AJD, Worthy JP, Scofield RP, et al. Biogeographical and phylogenetic implications of an early Miocene wren (Aves: Passeriformes: Acanthisittidae) from New Zealand. *J Vertebr Paleontol*. 2010;30(2):479–98.
32. Ksepka DT, Stidham TA, Williamson TE. Early Paleocene landbird supports rapid phylogenetic and morphological diversification of crown birds after the K–Pg mass extinction. *Proc Natl Acad Sci*. 2017;114(30):8047–52.
33. Mayr G, Noriega J. A well-preserved partial skeleton of the poorly known early Miocene seriema *Noriegavis santacruzensis* (Aves, Cariamidae). *Acta Palaeontol Pol*. 2013;60(3):589–98.
34. Mitchell KJ, Llamas B, Soubrier J, Rawlence NJ, Worthy TH, Wood J, et al. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science*. 2014;344(6186):898–900.
35. Mayr G, Alvarenga H, Mourer-Chauviré C. Out of Africa: fossils shed light on the origin of the hoatzin, an iconic Neotropical bird. *Naturwissenschaften*. 2011;98(11):961–6.
36. Miller AH. A fossil hoatzin from the Miocene of Colombia. *Auk*. 2012; 70(4):484–9.
37. Mayr G, De Pietri VL. Earliest and first Northern Hemispheric hoatzin fossils substantiate Old World origin of a “neotropic endemic”. *Naturwissenschaften*. 2014;101(2):143–8.
38. Jarvis ED, McCormack J, Ray DA, Ramakodi MP, Lyons E, McCarthy FM, et al. Three crocodylian genomes reveal ancestral patterns of evolution among archosaurs. *Science*. 2014;346(6215):1254449.
39. Tan MH, Austin CM, Hammer MP, Gan HM, Croft LJ. Whole genome sequencing of the Asian Arowana (*Scleropages formosus*) provides insights into the evolution of Ray-finned fishes. *Genome Biol* 2015;7(10):2885–95.
40. Sun Y, Volf J-N, Venkatesh B, Ravi V, Holland PWH, Barrell D, et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet*. 2016;48(4):427–37.
41. Qiao T, Lu J, Jia L, Zhu M, Zhao W, Yu X. Earliest known coelacanth skull extends the range of anatomically modern coelacanths to the early Devonian. *Nat Commun*. 2012;3(1):1–8.
42. Monteiro LR, Nogueira MR. Evolutionary patterns and processes in the radiation of phyllostomid bats. *BMC Evol Biol*. 2011;11(1):1–23.
43. Jones KE, Bininda-Emonds ORP, Gittleman JL. Bats, clocks, and rocks: diversification patterns in Chiroptera. *Evolution*. 2006;59(10):2243.
44. Shi JJ, Rabosky DL. Speciation dynamics during the global radiation of extant bats. *Evolution*. 2015;69(6):1528–45.
45. Nielsen R, Ho SYW, Rabosky DL, Fordyce RE, Steeman ME, Willerslev E, et al. Radiation of extant cetaceans driven by restructuring of the oceans. *Syst Biol*. 2009;58(6):573–85.
46. Thewissen JGM, Williams EM. The early radiations of Cetacea (Mammalia): evolutionary pattern and developmental correlations. *Annu Rev Ecol Syst*. 2002;33(1):73–90.
47. Martin RD. Adaptive radiation and behaviour of the Malagasy lemurs. *Philos Trans R Soc Lond Ser B Biol Sci*. 1972;264(862):295–352.
48. Thalmann U. Biodiversity, phylogeography, biogeography and conservation: lemurs as an example. *Folia Primatol*. 2007;78(5–6):420–43.
49. Herrera JP. Testing the adaptive radiation hypothesis for the lemurs of Madagascar. *R Soc Open Sci*. 2017;4(1):1–12.
50. Tebbich S, Sterelnik I, Teschke I. The tale of the finch: adaptive radiation and behavioural flexibility. *Philos Trans R Soc B Biol Sci*. 2010;365(1543):1099–109.
51. Cooney CR, Bright JA, Capp EJ, Chira AM, Hughes EC, Moody CJA, et al. Mega-evolutionary dynamics of the adaptive radiation of birds. *Nature*. 2017;542(7641):344–7.
52. Pinto G, Mahler DL, Harmon LJ, Losos JB. Testing the island effect in adaptive radiation: rates and patterns of morphological diversification in Caribbean and mainland Anolis lizards. *Proc R Soc B Biol Sci*. 2008;275(1652):2749–57.
53. Yoder JB, Clancey E, Des Roches S, Eastman JM, Gentry L, Godsoe W, et al. Ecological opportunity and the origin of adaptive radiations. *J Evol Biol*. 2010;23(8):1581–96.
54. Surget-Groba Y. Lizards in an evolutionary tree: ecology and adaptive radiation of anoles. *Amphibia-Reptilia*. 2010;32(1):141–2.

55. Joseph T. Ecology and evolution of Darwin's finches. *J Evol Biol.* 1986; 1(3):281–3.
56. Hu Y, Ghigliotti L, Vacchi M, Pisano E, Detrich HW, Albertson RC. Evolution in an extreme environment: developmental biases and phenotypic integration in the adaptive radiation of antarctic notothenioids. *BMC Evol Biol.* 2016;16(1):142.
57. Matschiner M, Hanel R, Salzburger W. On the origin and trigger of the notothenioid adaptive radiation. *PLoS One.* 2011;6(4):e18911.
58. Colombo M, Damerou M, Hanel R, Salzburger W, Matschiner M. Diversity and disparity through time in the adaptive radiation of Antarctic notothenioid fishes. *J Evol Biol.* 2015;28(2):376–94.
59. Galibert F, Fan S, Sanchez-Pulido L, Nikaido M, Przybylski D, Simakov O, et al. The genomic substrate for adaptive radiation in African cichlid fish. *Nature.* 2014;513(7518):375–81.
60. Seehausen O. African cichlid fish: a model system in adaptive radiation research. *Proc R Soc B Biol Sci.* 2006;273(1597):1987–98.
61. Takahashi T, Koblmüller S. The adaptive radiation of cichlid fish in Lake Tanganyika: a morphological perspective. *Int J Evol Biol.* 2011;2011:1–14.
62. Kocher TD. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet.* 2004;5(4):288–98.
63. Patrushev LI, Minkevich IG. The problem of the eukaryotic genome size. *Biochem Mosc.* 2009;73(13):1519–52.
64. Drake JW. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci.* 2006;88(16):7160–4.
65. Yamaguchi O, Mukai T. Variation of spontaneous occurrence rates of chromosomal aberrations in the second chromosomes of *Drosophila melanogaster*. *Genetics.* 1974;78(4):1209–21.
66. Sung W, Lynch M, Miller SF, Doak TG, Ackerman MS. Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci.* 2012; 109(45):18488–92.
67. Rubin C-J, Berglund J, Grabherr M, Martinez-Barrio A, Wang C, Webster MT, et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature.* 2015;518(7539):371–5.
68. Tollis M, Robbins J, Webb AE, Kuderna LFK, Caulin AF, Garcia JD, et al. Return to the sea, get huge, beat cancer: an analysis of cetacean genomes including an assembly for the humpback whale (*Megaptera novaeangliae*). *Mol Biol Evol.* 2019. <https://doi.org/10.1093/molbev/msz099>.
69. Lander ES, Haussler D, Castoe TA, Glor RE, Organ CL, Grabherr M, et al. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature.* 2011;477(7366):587–91.
70. Voskarides K. Group selection may explain cancer predisposition and other human traits' evolution. *J Mol Evol.* 2018;86(3–4):184–6.
71. Voskarides K. Combination of 247 genome-wide association studies reveals high cancer risk as a result of evolutionary adaptation. *Mol Biol Evol.* 2018;35(2):473–85.
72. Vicens A, Posada D. Selective pressures on human cancer genes along the evolution of mammals. *Genes.* 2018;9(12):1–13.
73. Martin HC, Batty EM, Hussin J, Westall P, Daish T, Kolomyjec S, et al. Insights into platypus population structure and history from whole-genome sequencing. *Mol Biol Evol.* 2018;35(5):1238–52.
74. Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, et al. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature.* 2007;447(7141):167–77.
75. Li R, Fan W, Tian G, Zhu H, He L, Cai J, et al. The sequence and de novo assembly of the giant panda genome. *Nature.* 2010;463(7279):311–7.
76. Shan L, Nie Y, Hu Y, Wang X, Xiu Y, Ma T, et al. Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. *Proc Natl Acad Sci.* 2017;114(5):1081–6.
77. Johnson RN, O'Meally D, Chen Z, Etherington GJ, Ho SYW, Nash WJ, et al. Adaptation and conservation insights from the koala genome. *Nat Genet.* 2018;50(8):1102–1111. Available from: <https://doi.org/10.1038/s41588-018-0153-5>
78. Voss R, Clawson H, Warren WC, Noll A, Minx P, Churakov G, et al. Genome sequence of the basal haplorhine primate *Tarsius syrichta* reveals unusual insertions. *Nat Commun.* 2016;7(1):1–11.
79. Moskalev AA, Lee S-G, Lyapunov AN, Zhu Y, Sun Y, Chen G, et al. Genome analysis reveals insights into physiology and longevity of the Brandt's bat *Myotis brandtii*. *Nat Commun.* 2013;4(1):1–8.
80. Zhang G, Cowled C, Shi Z, Huang Z, Bishop-Lilly KA, Fang X, et al. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science.* 2013;339(6118):456–60.
81. Pavlovich SS, Lovett SP, Koroleva G, Guito JC, Arnold CE, Nagle ER, et al. The Egyptian rousette genome reveals unexpected features of bat antiviral immunity. *Cell.* 2018;173(5):1098–1110.e18.
82. Lei M, Pang E, Mu S, Hua P, Zheng G, Dong D, et al. The genomes of two bat species with long constant frequency echolocation calls. *Mol Biol Evol.* 2016;34(1):20–34.
83. Jaffe DB, Massingham T, Clawson H, Zuk O, Kheradpour P, Wen J, et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature.* 2011;478(7370):476–82.
84. Eckalbar WL, Schlebusch SA, Mason MK, Gill Z, Parker AV, Booker BM, et al. Transcriptomic and epigenomic characterization of the developing bat wing. *Nat Genet.* 2016;48(5):528–36.
85. Alföldi J, Mancina A, Qin X, Gilbert MTP, Liu Y, Vinař T, et al. Convergent evolution of the genomes of marine mammals. *Nat Genet.* 2015;47(3):272–5.
86. Kim HW, Vijay N, Park C, Jin S, Zhang J, Park J-K, et al. Population genomic analysis reveals contrasting demographic changes of two closely related dolphin species in the last glacial. *Mol Biol Evol.* 2018; 35(8):2026–33.
87. Yim HS, Cho YS, Guang X, Kang SG, Jeong JY, Cha SS, et al. Minke whale genome and aquatic adaptation in cetaceans. *Nat Genet.* 2014;46(1):88–92.
88. Malde K, Seliussen BB, Quintela M, Dahle G, Besnier F, Skaug HJ, et al. Whole genome resequencing reveals diagnostic markers for investigating global migration and hybridization between minke whale species. *BMC Genomics.* 2017;18(1):1–11.
89. Foote AD, Vijay N, Ávila-Arcos MC, Baird RW, Durban JW, Fumagalli M, et al. Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nat Commun.* 2016;7:11693.
90. Moura AE, Van Rensburg CJ, Pilot M, Tehrani A, Best PB, Thornton M, et al. Killer whale nuclear genome and mtDNA reveal widespread population bottleneck during the last glacial maximum. *Mol Biol Evol.* 2014;31(5):1121–31.
91. Zhou X, Sun F, Xu S, Fan G, Zhu K, Liu X, et al. Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. *Nat Commun.* 2013;4:2708.
92. Warren R, Marra M, Li I, Troussard A, Taylor G, Chan A, et al. The genome of the beluga whale (*Delphinapterus leucas*). *Genes.* 2017;8(12):378.
93. Fjeldsa J, Burge SW, Campos PF, Johnson WE, Schubert M, Jarvis ED, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science.* 2014;346(6215):1311–20.
94. Chong AY, Braun EL, Castoe TA, Iguchi T, Khan S, Isberg SR, et al. Sequencing three crocodylian genomes to illuminate the evolution of archosaurs and amniotes. *Genome Biol.* 2013;13(1):415.
95. Wan QH, Pan SK, Hu L, Zhu Y, Xu PW, Xia JQ, et al. Genome analysis and signature discovery for diving and sensory properties of the endangered Chinese alligator. *Cell Res.* 2013;23(9):1091–105.
96. Webster MT, Han F, Grant PR, Grant BR, Andersson L, Lamichhaney S. Rapid hybrid speciation in Darwin's finches. *Science.* 2017;359(6372):224–8.
97. Veronika N, Laine, Toni I, Gossmann, Kyle M, Schachtschneider, Colin J, Garroway, Ole Madsen, Koen J. F. Verhoeven, Victor de Jager, Hendrik-Jan Megens, Wesley C. Warren, Patrick Minx, Richard P. M. A. Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat Commun.* 2016;7(1):1–9.
98. Qu Y, Tian S, Han N, Zhao H, Gao B, Fu J, et al. Genetic responses to seasonal variation in altitudinal stress: whole-genome resequencing of great tit in eastern Himalayas. *Sci Rep.* 2015;5:14256.
99. Yu H, Kabilov M, Zhao X, Peng C, You X, Wang J, et al. The Asian arowana (*Scleropages formosus*) genome provides new insights into the evolution of an early lineage of teleosts. *Sci Rep.* 2016;6(1):1–17.
100. Li J, Bian C, Hu Y, Mu X, Shen X, Ravi V, et al. A chromosome-level genome assembly of the Asian arowana, *Scleropages formosus*. *Sci Data.* 2016;3:160105.
101. Nikaido M, Noguchi H, Nishihara H, Toyoda A, Suzuki Y, Kajitani R, et al. Coelacanth genomes reveal signatures for evolutionary transition from water to land. *Genome Res.* 2013;23(10):1740–8.
102. Christoffels A, Przybylski D, Lander ES, Litman GW, Scharf M, Aken B, et al. The African coelacanth genome provides insights into tetrapod evolution. *Nature.* 2013;496(7445):311–6.
103. Shin SC h, Ahn DH w, Kim SJ i, Pyo CW o, Lee H, Kim MK, et al. The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. *Genome Biol.* 2014;15(9):468.

104. Conte MA, Kocher TD. An improved genome reference for the African cichlid, *Mtreaclima zebra*. *BMC Genomics*. 2015;16(1):1–13.
105. Meier JI, Marques DA, Wagner CE, Excoffier L, Seehausen O. Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol Biol Evol*. 2018;35(6):1489–506.
106. Baldo L, Santos ME, Salzburger W. Comparative transcriptomics of eastern African cichlid fishes shows signs of positive selection and a large contribution of untranslated regions to genetic diversity. *Genome Biol Evol*. 2011;3(1):443–55.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

