Comparative Analysis of Bat Genomes Provides Insight into the Evolution of Flight and Immunity
Guojie Zhang et al.
Science 339, 456 (2013); DOI: 10.1126/science.1230835

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of August 13, 2014):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/339/6118/456.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2012/12/19/science.1230835.DC1.html

This article cites 51 articles, 29 of which can be accessed free:
http://www.sciencemag.org/content/339/6118/456.full.html#ref-list-1

This article has been cited by 13 articles hosted by HighWire Press; see:
http://www.sciencemag.org/content/339/6118/456.full.html#related-urls

This article appears in the following subject collections:
Genetics
http://www.sciencemag.org/cgi/collection/genetics
Comparative Analysis of Bat Genomes Provides Insight into the Evolution of Flight and Immunity

Guojie Zhang,1,2,∗† Christopher Cowled,3 Zhengli Shi,4 Zhiyong Huang,5 Kimberly A. Bishop-Lilly,5 Xiaodong Fang,1 James W. Wynne,1 Zhiqiang Xiong,1 Michelle L. Baker,5 Wei Zhao,3 Mary Tachedjian,1 Yabing Zhu,2 Peng Zhou,5,3 Xuanting Jiang,1 Justin Ng,3 Lan Yang,3 Lijun Wu,1 Jin Xiao,2 Yue Feng,3 Yuanxin Chen,1 Xiaoqing Sun,1 Yong Zhang,3 Glenn A. Marsh,3 Gary Crameri,3 Christopher C. Broder,3 Kenneth G. Frey,5 Lin-Fa Wang,1,3,∗† Jun Wang3,6,8,9†

Bats are the only mammals capable of sustained flight and are notorious reservoir hosts for some of the world’s most highly pathogenic viruses, including Nipah, Hendra, Ebola, and severe acute respiratory syndrome (SARS). To identify genetic changes associated with the development of bat-specific traits, we performed whole-genome sequencing and comparative analyses of two distantly related species, fruit bat Pteropus alecto and insectivorous bat Myotis daudivi. We discovered an unexpected concentration of positively selected genes in the DNA damage checkpoint and nuclear factor-κB pathways that may be related to the origin of flight, as well as expansion and contraction of important gene families. Comparison of bat genomes with other mammalian species has provided new insights into bat biology and evolution.

References and Notes
28. Supplementary materials, including materials and methods, are available on Science Online.

Acknowledgments: We thank M. Rashband for providing the βIV-spectrin and sodium channel antibodies and M. Ericsson, B. Kasthuri, R. Schalek, G. Hao, and C. Speer for help in the preparation of hippocampal tissue slices. This work is supported in part by the NIH and the Collaborative Innovation Award from the HHMI. K.Z. is a HHMI investigator.

Supplementary Materials
www.sciencemag.org/cgi/content/full/science.1232251/DC1
Materials and Methods
Figs. S1 to S10
References (38–42)
1 November 2012; accepted 5 December 2012
Published online 13 December 2012.
10.1126/science.1232251
ity played a key role in its evolution (3), yet the by-products of oxidative metabolism [such as reactive oxygen species (ROS)] can produce harmful side effects including DNA damage (4). We hypothesize that genetic changes during the evolution of flight in bats likely included adaptations to limit collateral damage caused by by-products of elevated metabolic rate. Another phenomenon that has sparked intense interest in recent years is the discovery that bats maintain and disseminate numerous deadly viruses (5). In this context, we further hypothesize that the long-term coexistence of bats and viruses must have imposed strong selective pressures on the bat genome, and the genes most likely to reflect this are those directly related to the first line of antiviral defense—the innate immune system.

We performed high-throughput whole-genome sequencing of individual wild-caught specimens of *P. alecto* and *M. davidii* using the Illumina HiSeq platform (6). More than 100 × coverage high-quality reads were obtained for *P. alecto* and *M. davidii*, which resulted in high-quality assemblies (tables S1 to S3 and fig. S1). The two bat genomes, at ~2 Gb, were smaller in size than other mammals (7) (fig. S2), whereas the number of genes we identified was similar to those of other mammals (21,392 and 21,705 in *P. alecto* and *M. davidii*, respectively) (fig. S3). Both species displayed a high degree of heterozygosity at the whole-genome level (0.45% and 0.28% in *P. alecto* and *M. davidii*, respectively) (tables S4 and S5), whereas repetitive content accounted for slightly less than one-third of each genome (tables S6 and S7). We identified a novel endogenous viral element derived from *Saimiriine herpesvirus 2* that has expanded to 126 copies in *P. alecto* (table S8 and fig. S4). Gene family expansion and contraction analysis (tables S9 to S12) revealed significant expansion (*P < 0.05*) of 71 gene families in *M. davidii* compared with only 13 in *P. alecto*, which may be related to a recent wave of DNA transposon activity (8).

We screened all nuclear-encoded bat genes to identify those for which a single orthologous copy was unambiguously present in both bat species.

<table>
<thead>
<tr>
<th>Trait</th>
<th><em>Myotis davidii</em></th>
<th><em>Pteropus alecto</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Common name</td>
<td>David’s Myotis</td>
<td>Black flying fox</td>
</tr>
<tr>
<td>Suborder</td>
<td>Yangochiroptera</td>
<td>Yinpterochiroptera</td>
</tr>
<tr>
<td>Distribution</td>
<td>China</td>
<td>Australia, PNG, Indonesia</td>
</tr>
<tr>
<td>Habitat</td>
<td>Rock cavities</td>
<td>Trees, mangroves, rainforest</td>
</tr>
<tr>
<td>Diet</td>
<td>Insectivorous</td>
<td>Frugivorous, nectarivorous</td>
</tr>
<tr>
<td>Hibernation</td>
<td>Hibernates Nov-May</td>
<td>No</td>
</tr>
<tr>
<td>Echolocation</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Viral reservoir</td>
<td>Potential</td>
<td>Yes</td>
</tr>
</tbody>
</table>

![Fig. 1. Comparison of bat biological traits. *P. alecto* and *M. davidii* represent two distinct Chiropteran suborders and demonstrate diverse evolutionary adaptations. PNG, Papua New Guinea.](image)

![Fig. 2. Phylogenomic analysis. Maximum-likelihood phylogenomic analysis of 2492 genes from *M. davidii*, *P. alecto*, and eight mammalian species. Divergence time estimates in blue, gene family expansion events in green, and gene family contraction events in red. MRCA, most recent common ancestor.](image)
as well as in human, rhesus macaque, mouse, rat, dog, cat, cattle, and horse. From this, 2492 genes were used to perform maximum-likelihood and Bayesian phylogenetic analysis (Fig. 2 and figs. S5 to S7). All phylogenetically informative signals, including concatenated nucleotides and amino acids, vigorously supported bats as a member of Pegasoidea (Chiroptera + Perissodactyla + Carnivora) (9), with the bat lineage diverging from the Equus (horse) lineage ~88 million years ago, buttressed by findings at the transcript level (Equus the Carnivora) (amino acids, vigorously supported bats as a mem-
signals, including concatenated nucleotides and figs. S5 to S7). All phylogenetically informative
and Bayesian phylogenomic analysis (Fig. 2 and
rat, dog, cat, cattle, and horse. From this, 2492
as well as in human, rhesus macaque, mouse,

total of 2492 genes were used to perform maximum-likelihood and Bayesian phylogenetic analysis (Fig. 2 and figs. S5 to S7). All phylogenetically informative signals, including concatenated nucleotides and amino acids, vigorously supported bats as a member of Pegasoidea (Chiroptera + Perissodactyla + Carnivora) (9), with the bat lineage diverging from the Equus (horse) lineage ~88 million years ago, buttressed by findings at the transcript level (figs. S5 to S7).

The incongruence between nuclear and mitochondrial trees likely reflects rapid evolution of the mitochondrial genome of the bat ancestor during the evolution of flight (3).

To identify mechanisms that facilitated the origin of flight in bats, we surveyed genes involved in detection and repair of genetic damage.

A high proportion of genes in the DNA damage checkpoint–DNA repair pathway were found to be under positive selection in the bat ancestor, including ATM, the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs), RAD50, KU80, and MDM2 (Fig. 3A and Table 1). We propose that these changes may be directly related to minimizing and/or repairing the negative effects of ROS generated as a consequence of flight. Additionally in this pathway, TP53 and BRCA2 were shown to be under positive selection in M. davidii, whereas LIG4 was under positive selection in P. alecto (Table 1). Bat-specific mutations in a nuclear localization signal in p53 and a nuclear export signal in MDM2 (Fig. 3B and fig. S9) may affect subcellular localization and function in both species (11, 12). Other candidate flight-related genes under positive selection in the bat ancestor included COL3A1, involved in skin elasticity, and CACNA2D1, which has a role in muscle contraction (table S13).

We next examined genes of the innate immune system (Table 1). Positively selected genes in the bat ancestor included c-REL, a member of the nuclear factor κB (NF-κB) family of transcription factors, which also contained amino acid changes potentially affecting inhibitor of NF-κB (IκB) binding (fig. S10). In addition to diverse roles in innate and adaptive immunity (13), c-REL plays a role in the DNA damage response by activating ATM (14) and CLSPN (15), whereas ATM is also an upstream regulator of NF-κB (16). The DNA damage response plays an important role in host defense and is a known target for virus interaction (17), which raises the possibility that changes in DNA damage response mechanisms during selection for flight could have influenced the bat immune system.

It is intriguing that both P. alecto and M. davidii have lost the entire locus containing the PYHIN gene family, including AIM2 and IFIT16, both of which are involved in sensing microbial DNA and the formation of inflammasomes (fig. S11). The association between PYHIN genes and cell cycle regulation in other species (18) hints that loss of the PYHIN family in bats may be connected to changes in the DNA damage pathway, because at least one PYHIN gene is present in all other major groups of eutherian mammals (19). NLRP3, triggered by both viral infection and ROS in other mammals (20), plays an analogous role to AIM2 in inflammasome assembly and was also under positive selection in the bat ancestor (Table 1).

Natural killer (NK) cells provide a first line of defense against viruses and tumors and include two families of NK cell receptors; killer-cell immunoglobulin-like receptors (KIRs), encoded by genes in the leukocyte receptor complex (LRC), and killer cell lectin-like receptors (KLRs, also known as Ly49 receptors), encoded within the natural killer gene complex (NKC). KLRs and KIRs were entirely absent in P. alecto and reduced to a single Ly49 pseudogene in M. davidii (table S14). KIR-like receptors identified in other species (21) were also absent from both P. alecto and M. davidii genomes, which was supported by transcript analysis in P. alecto (10). This likely indicates that bat NK cells use a novel class of receptors to recognize classical major histocompatibility complex class I molecules. Furthermore, additional LRC members of the immunoglobulin superfamily [including sialic acid–binding immunoglobulin-like lectins (SIGLECs), leukocyte
immunoglobulin-like receptors (IL1Rs), carci-noembryonic antigen-related cell adhesion mole-cules (CEACAMs), and leukocyte-associated immunoglobulin-like receptors [LILRs]) have under-gone considerable gene duplication in *M. davidii* and other mammals yet have almost completely failed to expand in *P. alecto* (fig. S12). As the genes encoded within the LRC bind a variety of ligands and play multiple roles in immune regulation, these observations have diverse implications for differences in immune function between *P. alecto* and *M. davidii* and between bats and other mammals.

We identified seven complete and two partial copies of the digestive enzyme RNASE4 in *M. davidii* (table S15), whereas *P. alecto* RNASE4 has acquired a frameshift mutation resulting in loss of catalytic residues (fig. S13). We also identified critical amino acid changes in *M. davidii* RNASE4 genes (relative to the mammalian consensus) that suggest diversification of substrate specificity (fig. S13). With a proven role in host defense against RNA viruses (21), RNASE4 expansion in *M. davidii* may have implications for virus resistance but may also reflect the insectivorous diet of *M. davidii*, in contrast with that of *P. alecto*, which consumes predominantly fruit, flowers, and nectar.

*M. davidii* also differs from *P. alecto* in aspects including hibernation and echolocation (Fig. 1). Bile salt–stimulated lipase (BSSL), capable of hydrolyzing triglycerides into monoglycerides and subsequently releasing digestible free fatty acids, has been specifically expanded in *M. davidii* compared with *P. alecto* and other mammals (fig. S14). In addition, we observed six candidate genes related to hibernation, which showed positive selection in *M. davidii* and three other hibernating species relative to nonhibernators (table S16).

Seven echolocation-related genes, including new candidates WNT8A and FOS (a subunit of the AP–1 transcription factor), had significantly higher ratio of nonsynonymous to synonymous substitutions (dN/dS) in the echolocating *M. davidii* branch relative to non-echolocating branches (table S17). Of note, the third exon in *M. davidii* FOXP2 had even greater variation from the mammalian consensus than two previously identified variable sites (fig. S15), which suggests a specific transcript variant is involved in echolocation (23).

In summary, comparative analysis of *P. alecto* and *M. davidii* genomes has provided insight into the phylogenetic placement of bats and has revealed evidence of genetic changes that may have contributed to their evolution. Gene duplication events played a particularly prominent role in the evolution of *Myotis* bats and may have helped contribute to their speciation. Concentration of positively selected genes in the DNA damage checkpoint pathway in bats may indicate an important step in the evolution of flight, whereas evidence of change in components shared by the DNA damage pathway and the innate immune system raises the interesting possibility that flight-induced adaptations have had inadvertent effects on bat immune function and possibly also life expectancy (24). The data generated by this study will help to address major gaps in our understanding of bat biology and to provide new directions for future research.

References and Notes

4. A. Barzilai, G. Rotman, Y. Shiloh, DNA Repair (Amsterdam) 1, 3 (2002).
6. Materials and methods are available as supplementary materials on Science Online.

Acknowledgments: We thank H. Field, C. Smith, and M. Yu for helping source genomic DNA; K. Ikhana and J. J. Boomsma for constructive discussion; and M. Cowled for graphics assistance. We acknowledge financial support from the China National Genbank at Shenzhen, CSIRO (Office of the Chief Executive Science Leaders Award, Julius Award); The Australian

Table 1. DNA damage checkpoint and innate immune genes under positive selection in the bat lineages. The rate ratio $\omega$ of dN/dS was calculated using multiprotein alignments of *P. alecto* and *M. davidii* sequences with orthologous sequences from human, rhesus macaque, mouse, rat, dog, cattle, and horse. $\omega$0 is the average ratio in all branches, $\omega$1 is the average ratio in nonbat branches, and $\omega$2 is the ratio in the bat branch. A low $P$ value indicates that the $\omega$2 model fits the data better than the $\omega$1 model.

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Symbol</th>
<th>Gene</th>
<th>$\omega$0 (average)</th>
<th>$\omega$1 (other)</th>
<th>$\omega$2 (target)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ancestor</td>
<td>TLR7</td>
<td>Toll-like receptor 7</td>
<td>0.2821</td>
<td>0.2670</td>
<td>2.7778</td>
<td>3.54E-07</td>
</tr>
<tr>
<td>ATM</td>
<td>0.20996</td>
<td>0.19595</td>
<td>0.7163</td>
<td>1.34E-05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDM2</td>
<td>0.13358</td>
<td>0.12615</td>
<td>0.81085</td>
<td>4.05E-04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NLRP3</td>
<td>0.1788</td>
<td>0.1714</td>
<td>1.1884</td>
<td>1.93E-04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP3K7</td>
<td>0.0216</td>
<td>0.0194</td>
<td>0.4786</td>
<td>8.93E-03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAD50</td>
<td>0.09657</td>
<td>0.09343</td>
<td>0.28882</td>
<td>7.95E-03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRKDC</td>
<td>0.23036</td>
<td>0.22768</td>
<td>0.45155</td>
<td>6.80E-03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KUB8</td>
<td>0.31145</td>
<td>0.30436</td>
<td>0.91747</td>
<td>3.75E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c-REL</td>
<td>0.2495</td>
<td>0.2403</td>
<td>1.5717</td>
<td>1.11E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. alecto</td>
<td>TBK1</td>
<td>TANK-binding kinase 1</td>
<td>0.0643</td>
<td>0.0522</td>
<td>0.2930</td>
<td>1.29E-09</td>
</tr>
<tr>
<td>LIG4</td>
<td>0.12033</td>
<td>0.11376</td>
<td>0.24797</td>
<td>8.91E-04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL18</td>
<td>0.5298</td>
<td>0.4532</td>
<td>1.7647</td>
<td>2.66E-04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFNG</td>
<td>0.5010</td>
<td>0.4527</td>
<td>1.3282</td>
<td>4.89E-03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISG15</td>
<td>0.2069</td>
<td>0.1909</td>
<td>0.4387</td>
<td>2.63E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDX58</td>
<td>0.3040</td>
<td>0.2923</td>
<td>0.4661</td>
<td>1.23E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. davidii</td>
<td>IFNAR1</td>
<td>Interferon (α, β, and ω) receptor 1</td>
<td>0.4954</td>
<td>0.4723</td>
<td>31.0924</td>
<td>7.00E-03</td>
</tr>
<tr>
<td>TP53</td>
<td>0.25623</td>
<td>0.23933</td>
<td>0.48123</td>
<td>7.00E-03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRCA2</td>
<td>0.49002</td>
<td>0.47732</td>
<td>0.64213</td>
<td>1.31E-03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRAK4</td>
<td>0.1670</td>
<td>0.1583</td>
<td>0.3531</td>
<td>1.96E-02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Tunable Signal Processing Through Modular Control of Transcription Factor Translocation

Nan Hao,1,2 Bogdan A. Budnik,3 Jeremy Gunawardena,3 Erin K. O’Shea3,2*

Signaling pathways can induce different dynamics of transcription factor (TF) activation. We explored how TFs process signaling inputs to generate diverse dynamic responses. The budding yeast general stress–responsive TF Msn2 acted as a tunable signal processor that could track, filter, or integrate signals in an input-dependent manner. This tunable signal processing appears to originate from dual regulation of both nuclear import and export by phosphorylation, as mutants with one form of regulation sustained only one signal-processing function. Versatile signal processing by Msn2 is crucial for generating distinct dynamic responses to different natural stresses. Our findings reveal how complex signal-processing functions are integrated within a single molecule and provide a guide for the design of TFs with “programmable” signal-processing functions.

Many transcription factors (TFs) display diverse activation dynamics in response to various external stimuli (1–4). To investigate how TFs process upstream signals, we studied the Saccharomyces cerevisiae general stress–responsive TF Msn2 (5). In the absence of stress, Msn2 is phosphorylated by protein kinase A (PKA) and localized to the cytoplasm; in response to stress, Msn2 is dephosphorylated and translocates to the nucleus, where it induces gene expression (5).

Natural stresses elicit highly variable dynamics of Msn2 nuclear translocation (Fig. 1A) (6, 7), which are thought to result from oscillatory signaling inputs (presumably PKA activity) (8). To study how Msn2 processes oscillatory PKA inputs, we used an engineered yeast strain (6) carrying mutations in all three PKA isoforms that put stress–responsive TF Msn2 acted as a tunable signal processor that could track, filter, or integrate signals in an input-dependent manner. This tunable signal processing appears to originate from dual regulation sustained only one signal-processing function. Versatile signal processing by Msn2 is crucial for generating distinct dynamic responses to different natural stresses. Our findings reveal how complex signal-processing functions are integrated within a single molecule and provide a guide for the design of TFs with “programmable” signal-processing functions.

1Harvard University Faculty of Arts and Sciences Center for Systems Biology, Cambridge, MA 02138, USA. 2Howard Hughes Medical Institute, Department of Molecular and Cellular Biology, and Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. 3Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: erin_oshea@harvard.edu

Fig. 1. Tunable signal-processing behaviors of Msn2. (A) Illustration of the distinct single-cell dynamic responses of Msn2 to various stresses. (B) Steady-state abundance of Msn2 in the nucleus in response to various concentrations of 1-NM-PP1. In response to each concentration of 1-NM-PP2, Msn2 exhibited uniform and stable nuclear localization in single cells and did not exhibit stochastic fluctuations as observed in natural stress responses. Open circles: responses to different concentrations of 1-NM-PP1; closed circles: responses to 3 μM and 0.2 μM 1-NM-PP1, which are used as high- and low-amplitude inputs, respectively, for the following analyses. AU, arbitrary unit. (C) Averaged single-cell time traces of Msn2 nuclear translocation (bottom: n = 50 cells; error bar: single-cell variances) in response to oscillatory inputs with high and low amplitudes (top). (Left) High-amplitude input produced by 3 μM 1-NM-PP1; (right) low-amplitude input produced by 0.2 μM 1-NM-PP1. Pulse duration of 3 min; pulse interval of 2 min. To emphasize the fact that 3 μM 1-NM-PP1 elicits a steady-state response that is about twice the response elicited by 0.2 μM 1-NM-PP1, the top y axes are not presented on a linear scale.

Figure 1