SENSORY REWIRING IN AN ECHOLOCATOR: GENOME-WIDE MODIFICATION OF RETINOGENIC AND AUDITORY GENES IN THE BAT *MYOTIS DAVIDII*

Nicholas J. Hudson1*, Baker ML2, Hart NS3, Wynne JW2, Gu Q4,5, Huang Z6, Zhang G6, Ingham AB1, Wang L-F2,7 and Reverter A1

1 Computational and Systems Biology, CSIRO Agriculture Flagship, Queensland Bioscience Precinct, 306 Carmody Road, Brisbane, Queensland, Australia
2 Australian Animal Health Laboratory, CSIRO Biosecurity Flagship, Geelong, Victoria, Australia
3 School of Animal Biology and the Oceans Institute, University of Western Australia, Crawley, WA 6009, Western Australia, Australia
4 College of Information Sciences and Technology, Donghua University, Shanghai 201620, China
5 Faculty of Medicine, Imperial College London, South Kensington, London SW7 2AZ, England
6 BGI-Shenzhen, Shenzhen, 518083, China
7 Program in Emerging Infectious Diseases, Duke-NUS Graduate Medical School, Singapore

*Author for correspondence

Dr. Nicholas James Hudson
nick.hudson@csiro.au

© The Author(s) 2013. Published by the Genetics Society of America.
Abstract

Bats comprise 20% of all mammalian species and display a number of characteristics, including true flight, echolocation and a heightened ability to resist viral load that uniquely position this group for comparative genomic studies. Here we searched for evidence of genomic variation consistent with sensory rewiring through bat evolution. We focussed on two species with divergent sensory preferences. *M. davidii* is a bat species which echolocates and possesses dim- but not day-light adapted vision while the black flying fox (*Pteropus alecto*) has highly developed day vision but does not echolocate. Using the naked mole rat as a reference, we found 5 functional genes (*CYP1A2, RBP3, GUCY2F, CRYBB1 and GRK7*) encoding visual proteins that have degenerated into pseudogenes in *M. davidii* but not *P. alecto*. In a second approach genome-wide Codon Usage Bias (CUB) was compared between the two bat species. This CUB ranking systematically enriched for vision-related (*CLN8, RD3, IKZF1, LAMC3, CRX, SOX8, VAX2, HPS1, RHO, PRPH2, SOX9*) and hearing-related (*TPRN, TMIE, SLC52A3, OTOF, WFS1, SOD1, TBX18, MAP1A, OTOS, GPX1, USH1G*) machinery in *M. davidii* but not *P. alecto*. All vision and hearing genes selectively enriched in *M. davidii* for which orthologs could be identified were also more biased in the echolocating *M. lucifugus* than the non-echolocating *P. vampyrus*. We suggest that the existence of codon bias in vision and hearing-related genes in a species that has evolved echolocation implies CUB is part of evolution’s toolkit to rewire sensory systems. We propose that the two genetic changes (pseudogene formation and CUB) collectively paint a picture of that incorporates a combination of destruction and gain-of-function. Together, they help explain how natural selection has reduced physiological costs associated with the development of a smaller eye poorly adapted to day vision, but that also contribute to enhanced dim light vision and the hearing adaptations consonant with echolocation.

Key words: evolution, vision, codon usage bias, pseudogene, bat
Introduction

As pointed out previously (JONES et al. 2013) bats perceive the world using a range of sensory mechanisms, some of which have undergone specialisation (ALTRINGHAM and FENTON 2003). *M. davidii* and *P. alecto* are Chiropterans that utilise different habitats, prefer different diets and occupy different sensory niches (ZHANG et al. 2013). Previous work has compared and contrasted their genomes to shed light on the genetic basis on a host of their phenotypic similarities and differences. Recently, a new connection was drawn between bat immunity and bioenergetic demands based on gene family expansions and contractions and other lines of genetic evidence (ZHANG et al. 2013). In taking a sensory perspective of the two species, *P. alecto* forages through vision and smell but cannot echolocate. In contrast, *M. davidii* is an echolocator thought to have relatively poor day vision. It exhibits the loss of function often observed when biological systems avoid paying the cost of unnecessary capacity (WEIBEL et al. 1991).

Basic morphological data in echolocating bats - such as differences in the lens, cornea, vitreous body shape and proportion of rods and cones – are consistent with reduced visual acuity (BHATNAGAR 1975; PETTIGREW 1986; FUZESSERY et al. 1993; KALKO et al. 1996). A complication to the sensory ecology of myotid bats is that while they possess a smaller eye, this does not exclusively imply degradation. Recent work in *M. brandtii* suggests this echolocator may be adapted to dim light conditions such as found at dawn and dusk and also UV (SEIM et al. 2013). These visual adaptations are related to photopigment opsins, rhodopsin and downstream signalling components (SEIM et al. 2013). The molecular basis of visual adaptation in *M. davidii* per se is unknown.

Taking a broader sensory perspective, a very recently published analysis based on the ratio of non-synonymous to synonymous substitutions common to both echolocating dolphins and echolocating bats identified genes encoding sensory proteins, particularly vision and hearing, as striking signatures of convergent evolution (PARKER et al. 2013). The hearing adaptation relates to the echolocation facility, a multimodal capability comprising both specialised vocalisation on the one hand and signal reception on the other. A recent review of the genetic basis of the sensory ecology of echolocating bats can be found here (JONES et al. 2013). Genes thought to be of particular interest for echolocation per se include those encoding proteins related to vocalisation (*FOXP2*) and enhanced hearing (*SLC26A5*).

There is some debate over whether bat echolocation is 1) a derived characteristic not present in the common ancestor that has subsequently evolved in multiple echolocating lineages by convergent evolution, or 2) was present in the ancestral species then selectively lost (TEELING 2009). Either way, our aim is to contribute to the elucidation of the genetic basis of sensory rewiring, with a focus on these two species. We compare the output to two other bat species for whom genome wide data already exists, the echolocator *M. lucifugus* and the non-echolocator *P. vampyrus*. Pteropodids
diverged from other bats 55 mya, while the myotid bats are considered by some authorities to be a more recent lineage that split 15 mya. Assembled genomes for these four bats provide great scope for comparative genomics. Indeed, it has been pointed out that the future for understanding the molecular basis of sensory biology is promising, with comparative genomics having great potential (JONES et al. 2013). Here, we find a combination of hypothesis-driven and hypothesis-free approaches to be mutually reinforcing, shedding molecular insight into the apparent rewiring of M. davidii’s visual modality, and also complementary changes in the hearing modality which we suggest likely relate to echolocation.

Our motivation for exploring genome-wide patterns of codon usage bias (CUB) in the context of comparative mammalian evolution was our recent finding that pathway-level alterations in translation efficiency appear to have an under-appreciated role in driving evolutionary change through gain of function - even in complex eukaryotes such as mammals and birds (HUDSON et al. 2011). We previously observed systematic CUB in genes fundamental to important characteristics of each lineage, such as cell wall and chloroplast physiology in plants, mitochondrial function in birds and hair formation in mammals (HUDSON et al. 2011). Other recent work has also reinforced a likely role for CUB in human evolution. For example, a large scale tissue-specific analysis in humans previously found that the tRNA Adapation Index significantly correlated with expression levels both globally and in a range of tissues taken in isolation (WALDMAN et al. 2010). These lines of evidence support a role for CUB in mammalian evolution, but genome-wide CUB analysis has never previously been applied to any bat species.

Here, we apply a very sensitive CUB detection tool that, when applied genome-wide, finds a coordinated pattern of enrichment for genes encoding vision and hearing-related machinery in the echolocating M. davidii but not the non-echolocating fruitbat P. alecto. Orthologs of those vision and hearing genes were also more biased in M. lucifugus than P. vampyrus, although the effect was not so strong. The outlier visual proteins tend to be highly expressed in human eye tissue but this is not always the case, implying no simple relationship to abundance. Also based on comparative human data, some of these genes are expected to have highly restricted tissue-specific expression patterns, whereas others are widely expressed. Given observations we have made applying the same method in a number of other genomes, we submit that the most likely explanation to be selection on translation efficiency that manifests at the pathway level. We propose that the coordinate impact of these changes (pseudogenes and CUB) is sensory rewiring, presumably helping to manipulate the Myotis retinogenic pathway towards the development of a smaller, low light-adapted eye with additional changes relating to the enhanced hearing that contributes to the echolocation facility.

Materials and Methods

(a) Hypothesis-driven Pseudogene analysis
A list of vision related genes obtained from the gene ontology database (GO:0007601: visual perception) and visual system from the QIAGEN pathway library (http://www.qiagen.com/Products/Genes%20and%20Pathways/) defined presence, copy number and nature of corresponding genes in the M. davidii and P. alecto genomes. Proteins from human vision related genes and their neighboring genes were downloaded from Ensembl release 64. The longest transcript was chosen to represent each gene in cases when alternative splicing variants have been shown to exist. We then subjected all proteins to tBlastn analysis against the bat genomes with the similarity cutoff threshold of e-value=1e^-5. With the human protein set as a reference, we found the best hit for each human protein in the two bat genomes, with the criteria that more than 30% of the aligned sequence showed an identity above 30%. We used GeneWise algorithm (BIRNEY et al. 2004) (with parameters -genesf -tfor -quiet) to define the detailed exon-intron structure of each bat gene and to identify potential pseudogenes among the bat visual genes.

Initially, we targeted the examination of the bat orthologs for vision-related pseudogenes identified in the naked mole rat Heterocephalus glaber (KIM et al. 2011), a species with poor vision adapted to life in the dark. Genes containing frame shifts or premature stop codons were considered candidates. We filtered as follows: (1) To avoid reported frame shifts or premature stop codons that were due to a flaw in the GeneWise algorithm, we aligned all human proteins to their corresponding loci in the human genome, and genes with frame shifts or premature stop codons in human-to-human alignments were filtered; (2) Using the results of the human-to-human alignment from GeneWise, candidate pseudogenes with obvious splicing errors near their frame shifts or premature stop codons were filtered; (3) Candidate pseudogenes with a low number of reads covering their frame shift or premature stop codon sites were considered assembly errors. Genes with a considerable number of reads due to genotype variation at these sites were treated as heterozygous and filtered.

(b) Hypothesis-free Genome-wide CUB analysis

A number of CUB statistics have been proposed, and there is little consensus on the optimal approach. A very recent review outlining the pros and cons of some commonly applied CUB quantification approaches can be found here (BEHURA and SEVERSON 2013). Our CUB statistic uses the concept of entropy which estimates data regularities based on a combination of order and proportion. Its strength lies in its broad applicability and sensitivity. Because it is not systematically influenced by gene length, and accounts for amino acid composition, it can be fairly applied within and between genomes. However, it does not account for background GC content, nor does it determine whether bias is towards preferred or non-preferred codons. For this reason, background GC content and a more detailed analysis of codon usage must be retrospectively assessed on a case by case basis (i.e. in those genes whose entropic properties have been determined as extreme and therefore worthy of deeper investigation).
In brief, we determined the extent of CUB for every coding sequence using an information theory-based statistic as previously described (HUDSON et al. 2011). For each observed coding sequence we computed the entropy of the nucleic acid and compared it against the average entropy of 20 random sequences. The random sequences were generated in a way that they would encode the same amino acid sequence, but where the codons were selected at random. Because entropy provides a measure of data regularity, the differential entropy between the observed sequence and the random corresponds to the extent of regularity attributable to CUB.

We first ranked all coding sequences on extent of CUB on a within-genome basis. To ascertain the extent of functional enrichment analysis the ranked lists for *Myotis* and *Pteropus* were subjected to the GOrilla webtool (EDEN et al. 2009). GOrilla uses hypergeometric statistics to determine if particular biological processes are enriched within any ranked input set of genes. We used the single ranked list option based on human functional annotation, and imported SWISSPROT IDs which is one of the designated preferred input identifiers for GOrilla analysis. We also compared relative codon bias in the orthologs in common between the two species (of which we identified 6748). This secondary analysis helps identify genes that may not be extreme on a within genome basis, but still possess markedly different properties between the two species.

Because the GOrilla enrichment relies on incomplete annotation of functional processes, and can overlook functional connections even in well annotated genes, we used the top enrichment as a focus for further manual curation, based on mining the PubMed literature database. A subset of outlier CUB genes were retrospectively assessed for regional GC content. This was computed from the gene itself, plus 10 kb upstream from the start codon and 10 kb downstream from the stop codon.

*Independent validation of divergent CUB*

The differential entropy statistic is one of many ways of measuring CUB. In an effort to provide an independent line of evidence for differential CUB in the two bat species, and subsequent differential functional enrichment, we also ran the previously published Effective Numbers of Codons (ENC) (WRIGHT 1990), through the CodonW website: http://codonw.sourceforge.net/. According to (BEHURA and SEVERSON 2013) ENC is used to measure how far the codon usage of a gene departs from equal usage of synonymous codons. EC values range from 20 (extreme codon bias when one codon is used exclusively for each amino acid) to 61 (in cases of no codon bias where synonymous codons are equally likely to code the amino acid).

*CUB outlier genes quality checking*

As it is possible that gaps in the genome assembly data can result in artifacts with the cds sequences, the cds sequence data for all *M. davidii* CUB outlier genes were confirmed by comparison with cds
sequences from other species. We used translation BLAST to identify potential assembly problems, including gaps in our cds sequences. The bats *M. brandtii* and *M. lucifugus* plus other mammalian species available in Genbank were used for sequence alignment. In the event of potential missing exons we cross-referenced back to the *M. davidii* genomic assembly to confirm the quality of genomic sequence data.

*CUB analysis in two additional bat species: M. lucifugus and P. vampyrus.*

To detect comparative evidence for CUB in the sensory apparatus in echolocating bats, we computed CUB in an additional echocator (*M. lucifugus*) and non-echocator (*P. vampyrus*). The current genome assemblies from the Ensemble database for *P. vampyrus* (pteVam1) and *M. lucifugus* (Myoluc2.0) were used for this analysis. For these two comparison species we obtained coding sequence data from Biomart, yielding 14,141 annotated cds for *M. lucifugus* and 8,845 for *P. vampyrus*. CUB was estimated using the entropy statistic as previously described.

Results

(a) Hypothesis-driven Pseudogene analysis

Of the 19 visual perception genes identified as inactivated or missing in the naked mole rat, we identified 5 that have also become pseudogenes in *M. davidii* but not *P. alecto* (*CYP1A2, RBP3, GUCY2F, CRYBB1 and GRK7*) (**Figure 1**). Taxonomically the naked mole rat *H. glaber* is placed within rodents. In spite of some exceptional traits the overall properties of the *H. glaber* genome are similar to other mammals (Kim et al. 2011) making the comparison to the two bat species reasonable.

For *CYP1A2, CRYBB1* and *RBP3* single or multiple nucleotide deletions were responsible for frameshift mutations leading to premature stop codons in *M. davidii*. For *CRYBB1* the deletion was 31 nucleotides in the first exon compared to *P. alecto*. In contrast the *GRK7* and *GUCY2F* pseudogenes were caused by single point mutations that created single stop codons. In the case of *GUCY2F* two point mutations had created two stop codons in close vicinity to each other. The similarity between the intact *P. alecto* genes compared to their corresponding regions in the *M.davidii* genome, ranged from 85 to 90%.

(b) Hypothesis-free Genome-wide CUB analysis

We calculated codon usage preferences for the two species genome-wide (**Tables 2 and 3**). These genome-wide figures can be used to help establish preferred and non-preferred codons for that particular species in concert with the available tRNA pool. In turn this helps shed light on the translational implications of CUB. Namely, is CUB likely to increase or decrease the energetic cost and efficiency of translation for the gene in question?
We next used a hypothesis-free screen, based on CUB, to test for pathway functional enrichment. Prior to functional enrichment GOrilla screens the imported gene list, removes duplicates (keeping the only the most highly ranked candidate) then assigns Gene Ontology characteristics where those exist. For *M. davidii* we imported 13,654 genes that could be assigned SWISS PROT identifiers of which 37% were recognised. 6951 were recognised by Gene Symbol, 1731 duplicates were removed, and of the remaining 5220, 5056 had a GO term assigned. For *P. alecto* we imported 12,188 genes that could be assigned to SWISS PROT identifiers of which 40% were recognised. 6183 were recognised by Gene Symbol, 1056 duplicates were removed, and of the remaining 5127, 4956 had a GO term assigned.

*NEFH* (alias *YHU2*) (encoding the neurofilament heavy polypeptide) possessed extreme CUB in both species. Consistent with previous research in other organisms, the Translation Initiation Factor *IF2* (*PRP2*) was among the most biased genes in both species.

Of the biased genes within *M. davidii* we found the top ‘Process’ enrichment to be “*retina development of camera type eye*” (hypergeometric test *P*-value = 0.0000775; FDR q-value = 0.766) based on the prominent bias of the following genes: *CLN8, RD3, IKZF1, LAMC3, CRX, SOX8, VAX2, HPS1, RHO, PRPH2, SOX9* (**Figure 2A; Supplementary file 1**). The remaining functional enrichments reported in descending order by GOrilla for *M. davidii* were “*O-glycan processing*” (*P*-value = 0.000127), “*protein O-linked glycosylation*” (*P*-value = 0.000393), “*adenylate cyclase-inhibiting G-protein coupled receptor signalling pathway*” (*P*-value = 0.000508), “*renal absorption*” (*P*-value = 0.00051), “*renal water transport*” (*P*-value = 0.00051), “*peripheral nervous system neuron axonogenesis*” (*P*-value = 0.000593), “*positive regulation of transcription from RNA polymerase II promoter in response to stress*” (*P*-value = 0.000791), “*sensory perception of light stimulus*” (*P*-value = 0.000853), “*sensory perception of sound*” (*P*-value = 0.000932) and “*negative regulation of neutrophil differentiation*” (*P*-value = 0.000989).

The ‘*sensory perception of light stimulus*’ contained the following additional genes not recognised in the ‘*retina development*’ function, but which play a related physiological role: *OPN4, NR2E3, CNGB1, RDH8, ROM1, CLN6, WFS1, KIFC3, RGR, RP1L1, RDH5, CABP4, AIPL1, RAX2, NRL and USH1G*. The sensory perception of sound comprised the following genes (*TPRN, TMIE, SLC52A3, OTOF, WFS1, SOD1, TBX18, MAP1A, OTOS, GPX1 and USH1G*). Overall, we considered the collective presence of 4 functions in *M. davidii* (retina development, peripheral axonogenesis, light perception and sound perception) clearly associated with sensory rewiring to be noteworthy.

There was no formal vision related enrichment detected for *P. alecto* (**Figure 2B**) or indeed of any other sensory rewiring such as sound perception. The top ‘Biological Process’ enrichments in descending order were “*cellular developmental process*” (*P*-value = 0.0000178), “*regulation of intrinsic apoptotic signalling pathway*” (*P*-value = 0.000559), “*G-protein coupled receptor signalling pathway*” (*P*-value = 0.000508), “*renal absorption*” (*P*-value = 0.00051), “*renal water transport*” (*P*-value = 0.00051), “*peripheral nervous system neuron axonogenesis*” (*P*-value = 0.000593), “*positive regulation of transcription from RNA polymerase II promoter in response to stress*” (*P*-value = 0.000791), “*sensory perception of light stimulus*” (*P*-value = 0.000853), “*sensory perception of sound*” (*P*-value = 0.000932) and “*negative regulation of neutrophil differentiation*” (*P*-value = 0.000989).
pathway” (P-value = 0.000634), “developmental process” (P-value = 0.000192), “regulation of interferon-gamma biosynthetic process” (P-value = 0.000267), “positive regulation of MAPK cascade” (P-value = 0.000354), “cell differentiation” (P-value = 0.000381), “cell surface receptor signalling pathway” (P-value = 0.00056), “regulation of MAPK cascade” (P-value = 0.000652), “regulation of interferon-gamma production” (P-value = 0.000703) and “blood vessel development” (P-value = 0.000813). In both species, there was some indication of enrichment relating to immune function.

When we compared these orthologs of the two genomes directly on a between-genome basis (Figure 3), IKZF1 (also identified in the “retina development of the camera eye”) displayed a prominent outlier value in M. davidii. LAMC3, RHO and SOX9 were genes annotated as vision related that were absent in the P. alecto annotation and so orthologs could not be directly compared. A number of other orthologs were found to possess more bias in Myotis than in Pteropus, some of which were manually explored for their possible connection to the visual system. While not identified as relating to hearing by GOrilla enrichment, the outlier position of MUC19 is of interest, given 1) the general role of mucin family members in signal transduction during hearing (KERSCHNER 2007) and 2) that MUC19 specifically encodes a major gel-forming mucin expressed in the middle ear (KERSCHNER et al. 2009).

However, the mucins are known to bind pathogens, so an immune function or some other role cannot be discounted.

The overall distribution of these data was not symmetrical. There were more relatively biased orthologs in M. davidii than in P. alecto. This may reflect the fact that the Myotis bats are a much more recent lineage that, according to some sources, diverged approximately 15 mya from other bat species. Alternatively, the skewed data may reflect the fact that M. davidii possesses more derived phenotypic characteristics than P. alecto, resulting from the evolution of enhanced echolocation in this species. The genome-wide GC content is similar in the two bat species, with P. alecto (39.7%) slightly more biased than M. davidii (42.7%). This implies genome-wide GC content does not account for the slight skew in the between species distribution.

It is clear from scrutinising the plots that the CUB statistic has the appealing feature of being independent of gene length as the mass of the data in Figure 1 follows a horizontal line in both species. It is also immediately apparent that the vast majority of genes in both bat species do not exhibit extreme CUB, sitting close to the horizontal line. Nevertheless, in both cases there are a small subset of genes, many of which are short and some of which are long, which possess properties of relatively extreme CUB. Moreover, this subset enriches for a particular biological function, even though the physical location, gene length and other molecular details vary. This observation can be generalised. We detected the exact same broad patterns previously identified in 6 other eukaryotic genomes: Arabidopsis, S. cerevisiae, C. elegans, chicken, chimpanzee and human (HUDSON et al. 2007).
i.e. while most genes are unbiased, a small subset of variable length are biased and these enrich for particular functions that appear relevant to that lineage.

We calculated the GC content up and downstream of the differentially biased \textit{IKZF1}, \textit{GPSM1} and \textit{SPRR1} (cornifin) finding them to be either almost identical, or more out of equilibrium in the Pteropodid genome. Table 1 summarises these findings. Therefore regional background GC content is unable to account for the greater CUB in these Myotid coding sequences. Gene length changes between species is also unable to account for the observation of differential CUB, as while it varies, the variation is not systematic. \textit{M. davidii IKZF1} is longer than in \textit{P. alecto}, \textit{GPSM1} is shorter, and \textit{SPRR1} is almost identical – while CUB is substantially more extreme in \textit{M. davidii} in all 3 cases.

\textit{Independent validation of divergent CUB}

The ENC measure of genome-wide CUB reinforced the change in the visual system of \textit{M. davidii}, and partially reinforced the change in its auditory system when compared to \textit{P. alecto}. However, the sensory enrichments observed for \textit{M. davidii} were no longer ranked among the top ten on a within-genome basis.

The specific functional enrichments “retina development in camera-type eye” (based on CLN8, BAX, \textit{PRPH2}, LAMC3, \textit{IKZF1}, SOX9, NPHP4, PFDN5, PVRL1, SOX8, RD3) and “sensory perception of sound” (based on \textit{ASIC2}, \textit{TPRN}, \textit{MYO15A}, WFS1, SOBP, ESPN, GPX1, TIMM13, GRM7, TMIE, SLC52A3, OTOF, SOD1, MAP1A, OTOS, SIX1, USH1G, MYO7A) were both significant in \textit{M. davidii} (\(P = 0.000294\) and \(P = 0.00066\) respectively) but not \textit{P. alecto}, broadly reinforcing the patterns identified by the original analysis performed using the entropy-based statistic. Interestingly, the Myotid bat also had the functional enrichment (\(P = 0.000232\)) “sensory perception of mechanical stimulus” (based on \textit{ASIC2}, \textit{TPRN}, \textit{MYO15A}, \textit{TSHZ3}, WFS1, SOBP, GPX1, TIMM13, GRM7, TMIE, SLC52A3, OTOF, SOD1, MAP1A, OTOS, SIX1, USH1G, MYO7A). The overlap in content with “sensory perception of sound” given genes such as TMIE (trans-membrane of inner ear), and components of the otolith (OTOF and OTOS) suggests this function is also essentially related to inner ear development.

Vision enrichments for \textit{P. alecto} are present but to a lesser extent and do not convincingly highlight sensory rewiring. While \textit{P. alecto} does have vision-related enrichments given the presence of “closure of optic fissure” (\(P = 0.00026\)) “positive regulation of lens epithelial cell proliferation” (\(P = 0.00026\)) and “regulation of lens epithelial cell proliferation” (\(P = 0.00026\)), these were all based on the ranked position of a single gene, \textit{SOX11}. Finally, “Inner ear development” (based on \textit{PDGFA}, \textit{CCM2}, \textit{CYTL1}, \textit{HPCA}, \textit{LHX3}, \textit{GATA3}, \textit{SHH}, \textit{CEBPA}, \textit{SOX2} and \textit{SIX1}) was enriched in Pteropus (\(P = 0.000816\)). The genes comprising the hearing enrichment in \textit{P. alecto} are different to those highlighted in \textit{M. davidii}. 

10
The genome-wide relationship between the two different CUB statistics (entropy versus ENC) was plotted (Figure 4). We discovered a positive relationship (correlations of 0.619 and 0.672 in M. davidii and P. alecto respectively). The entropy CUB statistic appears to discriminate the highly biased genes more sensitively, as evidenced by the skew at the bottom left of both species. This enhanced discrimination of entropy as one assesses CUB in the more biased genes appears a possible driver of any observed discrepancies in functional enrichment.

**CUB outlier genes quality checking**

* *M. davidii* LAMC3, CLN8 and SOX9 contained apparent gaps in the cds following translation BLAST and comparison to other bats and mammals. *IKZF1, WFS1* and *SOX8* were in the expected reading frame but not contiguous. Collectively, these genes possess low quality genomic sequence (string of N’s) in predicted exonic regions that might explain any missing cds sequence compared to other mammals. In contrast, alignment of *M. lucifugus* sequence for *IKZF1, LAMC3, SOX9* and *SOX8* to other mammals is more conserved. This implies these genes may contain genomic assembly errors peculiar to *M. davidii* rather than being indicative of true evolutionary change. *SOD1* appears to be partial sequence only.

**CUB enrichment across two additional bat species: M. lucifugus and P. vampyrus.**

In an effort to find independent comparative evidence for CUB in genes encoding proteins related to the sensory apparatus of echolocating bats, we computed the entropy based CUB in an additional echolocator (*M. lucifugus*) and a non-echolocator (*P. vampyrus*) species (Supplementary File 2; Figure 5). Sensory perception was not enriched in either species according to GOrilla analysis. The top 3 functional enrichments for *M. lucifugus* were G-protein coupled receptor signalling pathway (*P* = 0.000000137), G-protein coupled receptor signalling pathway coupled to cyclic nucleotide messenger (*P* = 0.00000667) and negative regulation of blood pressure (*P* = 0.0000295). For *P. vampyrus* they were G-protein coupled receptor signalling pathway (*P* = 0.000000196), adenylate cyclise-activating G-protein coupled receptor signalling pathway (*P* = 0.0000296) and feeding behaviour (*P* = 0.0000315).

Next, we identified the orthologs of those vision and hearing sensory genes previously enriched in *M. davidii*. In the three vision (*SOX8, HPS1, IKZF1*) and six hearing (*TMIE, SLC52A3, TBX18, OTOS, GPX1* and *USH1G*) instances where both *M. lucifugus* and *P. vampyrus* orthologs could be identified, they were always more biased in *M. lucifugus* than *P. vampyrus* (Figure 6).

Finally, we manually explored the *M. lucifugus* and *P. vampyrus* CUB outliers using the gene ontology information in the Gene function of the NCBI database and literature searching PubMed. In *M. lucifugus* on a within-genome basis the 2nd (*TSPAN10, alias OCSP*) (WISTOW et al. 2002), 15th
(HCRT) (LIU et al. 2011), 20th (FIZI) (MALI et al. 2008), 25th (RTN4R) (WANG et al. 2008) and 34th (TRAPPC6A) (GWYNN et al. 2006) were all vision-related (Figure 5A). TSPAN10 has a very strong tissue-restricted expression in retina based on the BioGPS normalised tissue expression library (WU et al. 2009). In fact, retina is the only tissue out of the 84 assayed where TSPAN10 has an expression level > 3-fold of the mean (Supplementary File 3). FIZI is a transcriptional co-factor that interacts with NRL and CRX to drive expression of rod photoreceptor-specific genes (MALI et al. 2008).

Intriguingly, one of these partners (CRX) was actually prioritised by CUB in M. davidii. TRAPPC6A contributes to melanosome formation, and when mutated in mouse yields a loss of pigmentation phenotype in the retinal epithelial layer (GWYNN et al. 2006).

While there was no formal enrichment for sensory perception in M. lucifugus, the CUB data obtained for the two myotid bats is broadly comparable at a pathway level (vision and hearing). Although some details differ, this highly suggestive of a role for CUB in echolocating bat sensory evolution.

Discussion

In this study we performed a comparative genomic assessment of Myotid and Pteropodid bats using multiple approaches, and the results provide information on the genetic origins of sensory rewiring in an echolocating bat. We began by identifying pseudogenes using the naked mole rat as a comparative species with poor vision. Pseudogenes are relatives of genes that have subsequently lost their protein coding function, usually by disabling mutations such as premature stop codons and frameshifts. The observed correspondence between the naked mole rat and the echolocating bat (CYP1A2, RBP3, GUCY2F, CRYBB1 and GRK7) points to the general importance of these particular genes in controlling mammalian vision, and also to pseudogenes in particular as a genetic signature following loss of physiological function (NEWMAN and ROBINSON 2005; MEREDITH et al. 2013).

The next analysis was a hypothesis-free screen based on the genome-wide extent of CUB. In simple organisms CUB is known to alter translation efficiency and subsequent protein abundance, but the potential role in governing translation efficiency during mammalian evolution is less clear (DURET 2002). Supporting the concept it is known that protein expressed from a foreign gene in human cells can be increased by two orders of magnitude if the sequence is engineered to contain the most commonly used human codons (GUSTAFSSON et al. 2004). However, there is an apparent lack of correlation between codon bias and protein abundance and it is possible for codon usage studies to be systematically influenced by amino acid content, gene length and other potential artefacts (CHAMARY et al. 2006).

Despite these variable findings all these studies have been impacted by an inability to exhaustively interrogate an entire and well annotated genome (BEHURA and SEVERSON 2013). For this reason, recent work that focuses on complete genome screens as well as other factors such as accounting for
tissue and developmental stage, along with this study, may provide the first complete evidence for
correlation between CUB and protein abundance in humans and other mammals (WALDMAN et al.
2010). A recent study concluded that although variation in mutational bias is the dominant force
influencing CUB, translational selection acts as a weak additional factor and that this is widely true
across the vertebrates (DOHERTY and McINERNEY 2013).

On a within-genome basis we were struck with the finding that the top functional enrichment for the
echolocating M. davidii was retina development, a process not enriched at all on a within-genome
basis in P. alecto. The additional M. davidii enrichments for ‘axonogenesis’, ‘light perception’ and
‘sound perception’ that were absent in P. alecto reinforced evidence for sensory rewiring in the
echolocating species. Equivalent analyses run on the additional echolocator M. lucifugus and the
additional non-echolocator P. vampyrus did not receive formal enrichment scores for vision and
hearing. However, manual exploration of the outliers found intriguing support for vision, particularly
through TSPAN10 which has a very prominent tissue specific expression in retina, and the
transcriptional co-activator FIZ1 which interacts with CRX to drive the expression of rod
photoreceptor-specific genes.

To complement our CUB within-genome analysis (Figure 2), we also compared the orthologs of M.
davidii and P. alecto directly (Figure 3). Orthologs identified as substantially more biased in M.
davidii (but not previously identified by the within-genome GOrilla analysis) were manually explored
for possible connections to the visual and hearing systems through targeted PubMed literature
searches. Although these genes are bound together by their shared role in the biological processes of
vision and hearing, they are somewhat diverse in terms of other more detailed molecular and
functional properties. As regards the visual proteins, some are transcription factors (e.g. CRX), while
others are visual pigment proteins (e.g. RHO). Based on the TIGER database (LIU et al. 2008) which
provides normalised tissue mRNA comparisons in humans, many of the vision genes are abundantly
expressed in eye, some are not. Furthermore, some are expressed highly across a wide range of tissues
while others have an extremely restricted expression profile.

Clearly, this range of tissue-level expression profiles renders any simple correlation of CUB to protein
abundance unlikely. However, this does not mean one should discount CUB as a driver of eukaryotic
evolution, as previously argued (WALDMAN et al. 2010). They found significantly higher correlations
observed in adult tissues than the fetus, pointing to the importance of developmental stage in these
kind of assessments (WALDMAN et al. 2010). In that work, related sets of functional gene groups were
found to be translated efficiently in each tissue, pointing to the importance of viewing groups of genes
as components in a broader biological process or pathway (WALDMAN et al. 2010). In the future,
comparison of biased genes to actual expression level in components of the bats sensory system
would be of interest, but unfortunately these data are not presently available.
In order to allow the CUB outliers to be placed in context, we calculated the genome-wide preferred codons for *M. davidii* and *P. alecto* based on all the available coding sequences available from the genome sequencing efforts (Tables 2 and 3). These data can be exploited to make predictions of bias towards or against preferred codons. In the context of explaining the low light adapted vision and small eyes of *M. davidii*, we would predict enhanced translation efficiency (bias towards preferred codons) in negative regulators of eye size and visual pigments adapted for low light, and the converse (non-preferred codons) for positive regulators of eye size and pigments adapted for day vision. However, the developmental pathways controlling retinogenesis are incomplete, and the genetic changes in the highlighted genes need to be assessed on a case by case basis, and await future functional validation. Of the vision related genes highly biased in *M. davidii* IKZF1 encodes a transcription factor called Ikaros family zinc finger 1, an ortholog of which confers competence to mouse retinal progenitor cells (ELLIOTT et al. 2008). Manual exploration identified additional highly differentially biased genes encoding proteins expressed in eye components - GPSM1 (KEROV et al. 2005) and SPRRI (TONG et al. 2006). Enrichments for light and sound perception include numerous additional sensory genes not included in the specific retinogenic enrichment. The observation of extreme CUB in CRX in *M. davidii* and its binding co-factor FIZ1 in *M. lucifugus* supports this molecular pathway as a likely hotspot in vision-related adaptations that have evolved in nocturnal echolocating bats.

Overall, our findings across two echolocating (*M. davidii* and *M. lucifugus*) and two non-echolocating (*P. alecto* and *P. vampyrus*) bats strongly support a recent observation made across echolocating mammals (bats and bottlenose dolphins) using a different comparative genomics approach. In this independent study, it was found that patterns of convergent protein evolution (based on non-synonymous versus synonymous amino acid substitutions) identify pathways contributing to hearing and vision as among the strongest signals (PARKER et al. 2013). The exact overlap between our differential CUB analysis versus the protein evolution study of (PARKER et al. 2013) included the genes RHO and OTOF which are involved in light and sound perception, respectively. The fact that not all genes exist in both analyses suggests that multiple regulatory mechanisms are likely to contribute to change of function.

*NEFH* (alias *YHU2*) (encoding the neurofilament heavy polypeptide) possessed extreme CUB in both *M. davidii* and *P. alecto*, implying it was either present in the ancestor, or has subsequently converged due to a fundamental role in both bat lineages. A broader comparative analysis is needed to shed light on this gene, but unfortunately an annotated ortholog has not been identified in either the *M. lucifugus* or *P. vampyrus* sequences. One could sequence *NEFH* in a range of bats and one or more Laurasiatherian outgroups. *NEFH* is most highly expressed in the peripheral nervous system (LIU et al. 2008) – so a generic sensory explanation seems plausible i.e. one not specific to echolocation.
Finally, a number of unannotated highly biased genes in both *M. davidii* and *P. alecto* were detected. These may represent targets for future functional studies.

**Conclusion**

Overall, our findings provide insight into two genetic mechanisms contributing to evolution’s toolkit for adapting phenotypes. The deterioration of a number of core visual genes in *M. davidii* is consistent with genomic evolution in the naked mole rat (Kim *et al.* 2011) and the poor day vision and small eyes of this species. Pseudogene formation is clearly a loss-of-function decay process. Disabling mutations are allowed to persist because of relaxation of selection pressure on that particular protein. Because mutation is random, if unchecked it will cause DNA sequence to tend towards (random) equilibrium. This contrasts with extreme CUB which enforces structure and regularity on DNA sequence. This enforced regularity is strongly indicative of an applied evolutionary force, and tends to be associated with gain-of-function.

The CUB evidence we have gathered for manipulation of translation efficiency at the biological pathway level has not previously been reported in the context of bat sensory ecology and warrants further investigation. The CUB data collectively imply a suite of adaptations to both the visual system (perhaps positively impacting on adaptation to dim light) and the enhancement of hearing (which presumably relates to echolocation). Overall, the general prioritisation of sensory rewiring in the echolocator is highly resonant with the recent findings of (Parker *et al.* 2013), despite a different line of genetic evidence and indeed different choice of echolocating species.

**Acknowledgments**

We wish to acknowledge fruitful discussions with Dr. James Kijas and Dr. Chris Elvin. This research was facilitated by the CSIRO Animal, Food and Health Sciences post-genomics workshop, held at Coffs Harbour during April 2013 and was supported by CSIRO. QG was supported by in part by the Shanghai Postdoctoral Sustention fund (no. 12R21410200). MLB was supported by an Australian Research Council Future Fellowship (FT110100234).
References


Pettigrew, J. D., 1986 Flying primates? Megabats have the advanced pathway from eye to midbrain. Science 231: 1304-1306.


Wright, F., 1990 The 'effective number of codons' used in a gene. Gene 87: 23-29.


**Figure 1** Gene structure, partial amino acid and nucleotide alignments for five *M. davidii* pseudogenes. Red triangles illustrate the location of stop codons compared to *P. alecto*. Amino acid alignment surrounding the stop codon region in *P. alecto* and *M. davidii*. Asterix represent stop codons. Nucleotide alignment between *P. alecto* and *M. davidii* corresponding to the same region in amino acid alignment. Grey nucleotides represent the region following a frame shift in *M. davidii* that leads to a premature stop codon (*red*).

**Figure 2** Genes showing extreme codon bias (negative differential entropy) on a within-genome basis in (A) *M. davidii* and (B) *P. alecto*, with those encoding vision and hearing related proteins highlighted.

**Figure 3** Orthologs showing differential codon bias (distance from diagonal) between *M. davidii* and *P. alecto* on a between-genome basis. Genes encoding vision and hearing related proteins are highlighted. Of the sensory genes *IKZF1* is particularly prominent. The highly codon biased *LAMC3*, *RHO* and *SOX9* identified by GOrilla as vision-related in *M. davidii* appear to be absent in *P. alecto*. *MUC19*, highly biased in *M. davidii*, is a Mucin family member expressed in the inner ear. NEFH is equally highly biased in both bat species.

**Figure 4** The relationship between ENC and entropy based CUB estimation for (A) *M. davidii* and (B) *P. alecto*.

**Figure 5** Genes showing extreme codon bias (negative differential entropy) on a within-genome basis in (A) *M. lucifugus* and (B) *P. vampyrus*, with those encoding vision and hearing related proteins highlighted.

**Figure 6** Orthologs showing differential codon bias (distance from diagonal) between *M. lucifugus* and *P. vampyrus* on a between-genome basis. Genes encoding vision and hearing related proteins are highlighted. Of the sensory genes, *TSPAN10* which has a highly restricted expression pattern in retina is particularly prominent. All the sensory genes prioritised by CUB in *M. davidii* are more biased in *M. lucifugus* than *P. alecto*. 
Table 1 The genomic properties of 3 vision-related genes showing more CUB in *M. davidii* than *P. alecto*. Background GC% and gene length cannot account for the observation.

<table>
<thead>
<tr>
<th>Gene</th>
<th>GC% <em>M. davidii</em></th>
<th>GC% <em>P. alecto</em></th>
<th>Gene length <em>M. davidii</em></th>
<th>Gene length <em>P. alecto</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>IKZF1</td>
<td>49.6</td>
<td>53.2 / 51.5</td>
<td>1908</td>
<td>228 / 816</td>
</tr>
<tr>
<td>GPSM1</td>
<td>49.1</td>
<td>65.4</td>
<td>705</td>
<td>2118</td>
</tr>
<tr>
<td>SPRRI</td>
<td>44.0</td>
<td>47.1</td>
<td>339</td>
<td>342</td>
</tr>
</tbody>
</table>
Table 2 Patterns of genome-wide codon usage in *Myotis davidii*. These data help set the expectation for whether a given *M. davidii* codon is preferred or non-preferred. Observed bias in any particular gene can therefore be interpreted as likely to increase or decrease the cost and efficiency of translation.

<table>
<thead>
<tr>
<th>Codon</th>
<th>Percentage</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU</td>
<td>15.7</td>
<td>(141141)</td>
</tr>
<tr>
<td>UUC</td>
<td>20.5</td>
<td>(183882)</td>
</tr>
<tr>
<td>UUA</td>
<td>7.2</td>
<td>(64716)</td>
</tr>
<tr>
<td>UUG</td>
<td>12.5</td>
<td>(112218)</td>
</tr>
<tr>
<td>CUA</td>
<td>6.4</td>
<td>(57262)</td>
</tr>
<tr>
<td>CUG</td>
<td>41.3</td>
<td>(371130)</td>
</tr>
<tr>
<td>AUU</td>
<td>14.6</td>
<td>(131554)</td>
</tr>
<tr>
<td>AUC</td>
<td>21.5</td>
<td>(193277)</td>
</tr>
<tr>
<td>AUA</td>
<td>6.9</td>
<td>(62193)</td>
</tr>
<tr>
<td>AUG</td>
<td>22.1</td>
<td>(198262)</td>
</tr>
<tr>
<td>GUU</td>
<td>10.3</td>
<td>(92630)</td>
</tr>
<tr>
<td>GUC</td>
<td>15.4</td>
<td>(137894)</td>
</tr>
<tr>
<td>GUA</td>
<td>6.4</td>
<td>(57396)</td>
</tr>
<tr>
<td>GUG</td>
<td>29.2</td>
<td>(261930)</td>
</tr>
<tr>
<td>CAA</td>
<td>11.8</td>
<td>(106136)</td>
</tr>
<tr>
<td>CAG</td>
<td>36.1</td>
<td>(323855)</td>
</tr>
<tr>
<td>AAU</td>
<td>12.0</td>
<td>(108140)</td>
</tr>
<tr>
<td>AAC</td>
<td>20.0</td>
<td>(179409)</td>
</tr>
<tr>
<td>ACA</td>
<td>13.9</td>
<td>(125180)</td>
</tr>
<tr>
<td>ACG</td>
<td>7.2</td>
<td>(65088)</td>
</tr>
<tr>
<td>GAA</td>
<td>28.7</td>
<td>(25792)</td>
</tr>
<tr>
<td>GAG</td>
<td>41.6</td>
<td>(374083)</td>
</tr>
<tr>
<td>GUA</td>
<td>6.4</td>
<td>(136268)</td>
</tr>
<tr>
<td>GAC</td>
<td>27.4</td>
<td>(246001)</td>
</tr>
<tr>
<td>GCG</td>
<td>7.6</td>
<td>(68234)</td>
</tr>
<tr>
<td>AAG</td>
<td>33.3</td>
<td>(298847)</td>
</tr>
<tr>
<td>AGA</td>
<td>11.5</td>
<td>(103418)</td>
</tr>
<tr>
<td>AGG</td>
<td>12.9</td>
<td>(115789)</td>
</tr>
<tr>
<td>AGU</td>
<td>12.1</td>
<td>(108595)</td>
</tr>
<tr>
<td>AGC</td>
<td>20.5</td>
<td>(183722)</td>
</tr>
<tr>
<td>ACA</td>
<td>13.9</td>
<td>(125180)</td>
</tr>
<tr>
<td>ACG</td>
<td>7.2</td>
<td>(65088)</td>
</tr>
<tr>
<td>GAA</td>
<td>28.7</td>
<td>(25792)</td>
</tr>
<tr>
<td>GAG</td>
<td>41.6</td>
<td>(374083)</td>
</tr>
<tr>
<td>GGA</td>
<td>15.5</td>
<td>(138999)</td>
</tr>
<tr>
<td>GGG</td>
<td>17.3</td>
<td>(155758)</td>
</tr>
</tbody>
</table>
Table 3 Patterns of genome-wide codon usage in *Pteropus alecto*. These data help set the expectation for whether a given *P. alecto* codon is preferred or non-preferred. Observed bias in any particular gene can therefore be interpreted as likely to increase or decrease the cost and efficiency of translation.

<table>
<thead>
<tr>
<th>Codon</th>
<th>Frequency</th>
<th>Count</th>
<th>Codon</th>
<th>Frequency</th>
<th>Count</th>
<th>Codon</th>
<th>Frequency</th>
<th>Count</th>
<th>Codon</th>
<th>Frequency</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU</td>
<td>16.3</td>
<td>156476</td>
<td>UCU</td>
<td>15.2</td>
<td>146151</td>
<td>UAU</td>
<td>11.3</td>
<td>108424</td>
<td>UGU</td>
<td>10.0</td>
<td>96432</td>
</tr>
<tr>
<td>UUC</td>
<td>20.0</td>
<td>192106</td>
<td>UCC</td>
<td>17.7</td>
<td>170229</td>
<td>UAC</td>
<td>14.8</td>
<td>141890</td>
<td>UGC</td>
<td>12.2</td>
<td>116902</td>
</tr>
<tr>
<td>UUA</td>
<td>7.7</td>
<td>74291</td>
<td>UCA</td>
<td>12.0</td>
<td>114942</td>
<td>UAA</td>
<td>0.5</td>
<td>5244</td>
<td>UGA</td>
<td>1.1</td>
<td>10325</td>
</tr>
<tr>
<td>UUG</td>
<td>13.1</td>
<td>125571</td>
<td>UCG</td>
<td>5.2</td>
<td>49969</td>
<td>UAG</td>
<td>0.5</td>
<td>4808</td>
<td>UGG</td>
<td>12.3</td>
<td>117792</td>
</tr>
<tr>
<td>CUU</td>
<td>13.1</td>
<td>125218</td>
<td>CCU</td>
<td>17.6</td>
<td>168660</td>
<td>CAU</td>
<td>10.4</td>
<td>100155</td>
<td>CGU</td>
<td>4.9</td>
<td>46662</td>
</tr>
<tr>
<td>CUC</td>
<td>19.8</td>
<td>190319</td>
<td>CCC</td>
<td>20.7</td>
<td>198601</td>
<td>CAC</td>
<td>15.7</td>
<td>150619</td>
<td>CGC</td>
<td>10.6</td>
<td>101489</td>
</tr>
<tr>
<td>CUA</td>
<td>7.3</td>
<td>69639</td>
<td>CCA</td>
<td>16.7</td>
<td>160345</td>
<td>CAA</td>
<td>12.2</td>
<td>116897</td>
<td>CGA</td>
<td>6.4</td>
<td>61551</td>
</tr>
<tr>
<td>CUG</td>
<td>39.5</td>
<td>379104</td>
<td>CCG</td>
<td>7.7</td>
<td>74204</td>
<td>CAG</td>
<td>35.0</td>
<td>336173</td>
<td>CGG</td>
<td>12.0</td>
<td>114783</td>
</tr>
<tr>
<td>AUU</td>
<td>15.1</td>
<td>144545</td>
<td>ACU</td>
<td>12.9</td>
<td>124104</td>
<td>AAU</td>
<td>16.3</td>
<td>156140</td>
<td>AGU</td>
<td>12.5</td>
<td>119863</td>
</tr>
<tr>
<td>AUC</td>
<td>20.4</td>
<td>196052</td>
<td>ACC</td>
<td>18.5</td>
<td>177795</td>
<td>AAC</td>
<td>18.9</td>
<td>181664</td>
<td>AGC</td>
<td>20.1</td>
<td>192594</td>
</tr>
<tr>
<td>AUA</td>
<td>7.3</td>
<td>70411</td>
<td>ACA</td>
<td>14.6</td>
<td>139725</td>
<td>AAA</td>
<td>24.3</td>
<td>233289</td>
<td>AGA</td>
<td>11.8</td>
<td>112920</td>
</tr>
<tr>
<td>AUG</td>
<td>21.5</td>
<td>206522</td>
<td>ACG</td>
<td>6.9</td>
<td>65967</td>
<td>AAG</td>
<td>31.9</td>
<td>305590</td>
<td>AGG</td>
<td>12.1</td>
<td>116001</td>
</tr>
<tr>
<td>GUU</td>
<td>10.6</td>
<td>101643</td>
<td>GCU</td>
<td>18.5</td>
<td>177510</td>
<td>GAU</td>
<td>21.4</td>
<td>205693</td>
<td>GGU</td>
<td>10.5</td>
<td>100839</td>
</tr>
<tr>
<td>GUC</td>
<td>15.2</td>
<td>146033</td>
<td>GCC</td>
<td>28.8</td>
<td>275932</td>
<td>GAC</td>
<td>26.3</td>
<td>253066</td>
<td>GGC</td>
<td>22.5</td>
<td>215905</td>
</tr>
<tr>
<td>GUA</td>
<td>7.2</td>
<td>69462</td>
<td>GCA</td>
<td>15.7</td>
<td>150363</td>
<td>GAA</td>
<td>29.6</td>
<td>284341</td>
<td>GGA</td>
<td>15.8</td>
<td>151181</td>
</tr>
<tr>
<td>GUG</td>
<td>27.9</td>
<td>267333</td>
<td>GCG</td>
<td>8.0</td>
<td>76818</td>
<td>GAG</td>
<td>40.3</td>
<td>387415</td>
<td>GGG</td>
<td>16.6</td>
<td>159580</td>
</tr>
</tbody>
</table>


Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
Figure 6