Next generation sequencing and analysis of a conserved transcriptome of New Zealand’s kiwi

Sankar Subramanian1,2, Leon Huynen1,2, Craig D Millar3, David M Lambert1,2*

Abstract

Background: Kiwi is a highly distinctive, flightless and endangered ratite bird endemic to New Zealand. To understand the patterns of molecular evolution of the nuclear protein-coding genes in brown kiwi (Apteryx australis mantelli) and to determine the timescale of avian history we sequenced a transcriptome obtained from a kiwi embryo using next generation sequencing methods. We then assembled the conserved protein-coding regions using the chicken proteome as a scaffold.

Results: Using 1,543 conserved protein coding genes we estimated the neutral evolutionary divergence between the kiwi and chicken to be ~45%, which is approximately equal to the divergence computed for the human-mouse pair using the same set of genes. A large fraction of genes was found to be under high selective constraint, as most of the expressed genes appeared to be involved in developmental gene regulation. Our study suggests a significant relationship between gene expression levels and protein evolution. Using sequences from over 700 nuclear genes we estimated the divergence between the two basal avian groups, Palaeognathae and Neognathae to be 132 million years, which is consistent with previous studies using mitochondrial genes.

Conclusions: The results of this investigation revealed patterns of mutation and purifying selection in conserved protein coding regions in birds. Furthermore this study suggests a relatively cost-effective way of obtaining a glimpse into the fundamental molecular evolutionary attributes of a genome, particularly when no closely related genomic sequence is available.

Background

DNA sequencing technologies have enabled us to decipher molecular sequences of individual organisms. Conventional DNA sequencing methods relying on fluorescent dideoxy terminators and capillary separation revolutionized sequencing and allowed the first constructions of complete genomes of a number of species from simple prokaryotes to higher vertebrates [1,2]. However the costs involved in eukaryotic genome-sequencing projects using these methods has been very high and thus such projects generally require the collaboration of several well-funded institutes. Furthermore the time required for sequencing and assembling such genomes can span several years. The advent of Next Generation DNA sequencing has reduced the time and expense of complete genome sequencing by orders of magnitude [3,4]. For example using next generation sequencing methods the complete genome of a human individual was completed in eight weeks by spending only a fraction of cost incurred using conventional BAC and shotgun-based cloning and sequencing methods [5].

The major limitation of a next generation sequencing approach is that the length of the sequence reads produced was until recently only 25-200 bases, as opposed to over a kilobase generated by conventional capillary-based sequencing methods. Although short sequence reads do not limit the amount of sequence data collected, this can hamper the assembly of the short sequence reads into large contigs. Therefore the availability of a genome of a closely related organism is generally required (to act as a scaffold) for the successful assembly of a new genome using the next generation sequencing method. For example, the assembly of the genomes of Neanderthal and Woolly Mammoth was possible only because of the availability, respectively, of complete human and African elephant genomes [6,7].

* Correspondence: d.lambert@griffith.edu.au
1Griffith School of Environment and the School of Biomolecular and Physical Sciences, Griffith University, 170 Kessels Road, Nathan, Qld 4111 Australia
Full list of author information is available at the end of the article

© 2010 Subramanian et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Recently new algorithms have been developed to assemble genomes de novo without using a closely related scaffold genome [8]. However for precise genome assembly using this software, substantial sequence coverage is required. For example, using next generation sequencing and de novo assembly the complete genome of a Chinese panda was obtained at 50× (times) coverage [9]. Although the sequence required (150 gigabases) was generated in only one month, the cost amounted to several million US dollars. Therefore without the availability of a closely related organism, assembling complete genomes de novo would still be financially restrictive for most research laboratories. An alternative would be to use low-coverage next generation sequencing to sequence and assemble only the transcribed regions of a genome (transcriptome).

Large-scale comparative genomic studies of avian genomes started soon after the chicken genome became available [10]. These studies revealed mutational rate differences among chromosomes by comparative analyses of chicken and turkey genomes [11,12]. With the availability of the complete genome of the zebra finch [13] a number of studies have examined genome-wide patterns of molecular evolution and gene expression in avian genomes [14-17]. However all these studies were performed using only the Neognathae birds.

In the current study we have sequenced and assembled a number of conserved protein-coding regions of an early stage transcriptome of a Paleognathae bird, the North Island Brown kiwi (Apteryx australis mantelli). Kiwi are flightless birds endemic to New Zealand, and are of high conservation importance. The purpose of this study was to isolate conserved transcribed sequences of kiwi in order to understand firstly the patterns of mutation and selection amongst protein coding sequences and secondly to use these nuclear gene sequences to reanalyze the divergence time between Paleo- and Neognathae birds.

Results and Discussion
Preliminary kiwi transcriptome analysis
Initial cDNA made from kiwi embryo k11-15 was tested for coverage by amplification of the temporally and spatially restricted developmental genes tbx5, cry1, pax6, BMP4, ptx1, hoxB1, hoxB8, and hoxD12. All genes were successfully amplified from the kiwi cDNA suggesting a comprehensive early stage kiwi transcriptome. As expected, comparison of the amplified kiwi gene sequences with those in GenBank consistently gave the chicken homologue as the closest match.

Assembly of the conserved regions of Kiwi protein-coding genes
To assemble over 75,000 FLX sequence reads we used the well-annotated chicken proteome. Kiwi sequences were used as a query to search over 22,000 chicken proteins using BLASTX. Significant hits were garnered and redundant kiwi reads were excluded. Garnered kiwi sequences were assembled into large contiguous segments. We used Blast2seq to align chicken proteins with translated assembled reads of kiwi. Finally we retained only the regions of the assembled reads where at least 90% of the amino acids aligned with their respective chicken orthologs. Although this resulted in a significant reduction in the number of kiwi genes retained, such an approach was required to minimize lineage specific duplicates. Furthermore we are confident that this approach will exclude most of the transcribed pseudogenes, which generally have a high rate of substitution. This stringent approach identified 1,543 kiwi genes. The locations of the chicken-orthologs of kiwi genes on the chicken chromosomes are shown in Figure 1. The average length of coding regions was 168 bp (Figure 2). Our approach to assemble the transcriptome is somewhat similar to a previous method, SCARF [18]. However SCARF is suitable only if the reference scaffold genome is closely related to the target genome with a synonymous divergence of <0.1.

Embryonic expression levels of kiwi genes
Since the sequences generated in this study were from the mRNAs of a kiwi embryo, the copy numbers of the

Figure 1 Chromosomal locations of orthologous kiwi genes recovered in this study on the chicken genome
genes reflect their expression levels. We found two genes with very high (> 2000 copies) expression levels. While we identified (through the functional annotation of its chicken ortholog) the product of one of these genes as a neuropeptide the function of the other could not be determined. We grouped kiwi genes into three categories based on their levels of expression. Roughly 14% of the genes were highly expressed (> 5 copies) while 50% of the genes appeared to be present as single copies in kiwi embryo (Figure 3A). Conversely the small fraction of highly expressed genes constituted approximately 59% of the expressed mRNAs in the embryo, whereas only 13% of the mRNAs were from genes with low expression levels (Figure 3B). For this analysis we did not include the two genes with very high expression levels (mentioned above) to avoid bias due to these outliers. Gene annotations show that most of the highly expressed genes were those involved in protein synthesis such as ribosomal proteins, elongation factors, tRNA synthetases, as well as structural proteins such as collagen.

Evolutionary divergence at neutral and constrained sites
An important parameter used extensively in genome analyses is the rate of molecular evolution. To examine the rate of neutral evolution we compared the concatenated kiwi transcripts with those of chicken. A likelihood-based distance analysis showed that divergence at neutral synonymous sites was 0.465 substitutions per site. Conversely, divergence at nonsynonymous positions was only 0.071 substitutions per site; over an order of magnitude less than that of neutral divergence and suggests that high levels of purifying selection are acting on amino acid replacement sites. This is perhaps not surprising as the genes analyzed in this study are expressed early in embryonic development and therefore most are expected to be highly conserved. The average ratio of nonsynonymous- to synonymous divergences (dN/dS) was 0.15 (± 0.002), which is comparable to that estimated using chicken-zebra finch sequences for the genes expressed in zebra finch embryo [16]. A previous study using zebra finch and emu (another palaeognathae bird) estimated the divergences at synonymous and nonsynonymous positions to be 0.47 and 0.04 respectively [19]. However we
found a much higher divergence between zebra finch and kiwi at synonymous (0.53) and nonsynonymous (0.07) positions, which suggests that the rate of evolution in kiwi might be faster than that of emu.

The distribution of the \(dN/dS\) ratio estimated for individual genes shows that over 60% of the genes were highly constrained (\(dN/dS < 0.1\)) (Figure 4). On the other hand only 5% of the genes had a high (> 0.5) \(dN/dS\) ratio. Among these we found 25 genes that appear to be evolving under positive selection (\(dN/dS > 1.0\)). However the difference between the \(dN\) and \(dS\) estimates was not statistically different (higher) for any gene. This is primarily due to the high variance in the estimates, caused by the short sequence length of the partial genes used in this analysis and this adversely affects \(dS\) estimates over \(dN\) estimates, as synonymous sites are fewer than replacement sites. A previous study observed a higher rate of evolution in the zebra finch lineage compared to chicken using anole lizard as an outgroup [17]. We examined this using kiwi as an outgroup and found an approximately 50% higher rate of evolution in the zebra finch lineage compared to chicken at synonymous (49%) as well as at nonsynonymous (57%) positions.

To compare sequence divergence between the birds and mammals we obtained the orthologous human and mouse genes and used the alignment that contained only the aligned regions from kiwi, chicken, human and mouse. Therefore we obtained the expression levels for individual genes. Therefore we obtained the expression levels for each gene in each species. The expression levels were determined for kiwi genes using an embryonic chicken library (with 22000 ESTs) [26]. Only those genes that had a high expression level (> 5 copies of mRNA) were selected.

Previous studies showed a low nonsynonymous divergence in highly expressed genes, owing to a higher selection constraint on these genes [24,25]. We examined this for the kiwi-chicken species pair and found that this also holds true for these avian species. We found that the highly expressed genes are generally highly constrained and conversely the rate of evolution at nonsynonymous sites of the genes with low expression levels showed a large variation [see also [25]]. Therefore a simple correlation analysis does not capture the actual relationship between expression level and the rate of nonsynonymous site evolution. Hence, we estimated the divergences at synonymous and nonsynonymous sites of genes with high, medium and low expression levels by concatenating the genes in each group. The nonsynonymous divergence of genes with low expression levels was found to be 50% higher than that of the genes with a high expression level (0.082 vs 0.054) and this difference was statistically significant (\(P < 0.0001\)) using a Z-test (Figure 5A, red). However the divergence at synonymous sites of lowly expressed genes was not significantly different to that of the highly expressed genes (0.464 vs 0.458 \(P = 0.58\)) (Figure 5B, red).

The expression levels determined for kiwi genes might be influenced by the level of coverage of individual genes. Therefore we obtained the expression levels for the orthologous chicken genes using an embryonic chicken library (with 22000 ESTs) [26]. Only those genes that had a high expression level (> 5 copies of mRNA) were selected. Molecular data based studies estimates suggest a ~115 My split for primates-rodents divergence [20] and a ~130 My split for paleo-neognathae birds [21,22]. Furthermore fossil-based estimates also suggest a similar divergence time for the primates-rodents split (62 My - 100 My) and for palaeognathae-neognathae split [23].
mRNA) in kiwi as well as in chicken embryos were designated as highly expressed genes. Similarly we determined the genes with medium and low expression levels. Although this reduced the number of genes in our analysis, this stringent approach also produced similar relationships between expression levels and the divergences (Figure 5A and 5B, blue). This analysis showed an almost two fold higher nonsynonymous divergence for lowly expressed genes than that of the genes with high expression level (0.090 Vs 0.047, \( P < 0.0001 \)). The difference in the synonymous divergences between the genes with low and high expression levels was only 14% but was statistically significant (0.483 Vs 0.425, \( P = 0.002 \)). This result suggests very weak selection in synonymous sites of conserved genes of birds.

**Functional categories of proteins expressed in the kiwi embryo**

To determine the functions associated with the kiwi proteins we obtained the functional annotations of their orthologous chicken counterparts from the Biomart resource ENSEMBL http://www.ensembl.org. Our search using the gene ontology classification of chicken genes resulted in the identification of gene function for ~500 genes. We found that the majority of the kiwi genes (41%) code for DNA/RNA- or protein-binding proteins (Figure 6A). This is not surprising as these proteins are largely involved in the control of transcription; an important regulatory activity expected in developing embryos. Furthermore proteins that regulate translation were found to constitute roughly 8%. Other major fractions of proteins detected were those that perform housekeeping functions such as catalysis (enzymes) (22%) and transport (17%).

We also determined the types metabolic pathways with which the kiwi genes are involved. This was done using information from the KEGG pathway database http://www.genome.jp/kegg/pathway.html for the chicken genes, which were then assigned to the corresponding orthologous kiwi genes. Figure 6B shows that roughly a quarter of kiwi genes were involved in genetic information processing pathways, which includes DNA replication and transcription. Approximately 27% and 21% of the genes were found to be associated with cellular and metabolic processes respectively. The remaining 14% of genes were shown to be involved in immunity and the endocrine system and 11% of the genes were associated with environmental processing including transport and signal transduction.

**Divergence times between Paleo- and Neognathae**

We attempted to determine the divergence time between paleo- and neognathae birds as well as that between Galloanseriformes and the remaining neognathae birds. For this purpose we compared the genomes of neognathae birds (chicken and zebra finch), a reptile (anole lizard) and used a mammal (human) as an outgroup. The orthologous sequences of these genomes were obtained using reciprocal BLAST. This resulted in an orthologous dataset of 702 genes and included only those protein-coding sequences that align with the kiwi transcripts. The nucleotide sequences of all genes were concatenated and the branch lengths were estimated.
using BASEML under a general time reversible model of sequence evolution using the tree topology shown in Figure 7. First we conducted a simple direct estimation of the divergence times using pair-wise ML distances and a calibration time of 255 My (million years) between reptiles and birds. The latter was obtained from fossil studies [27]. This yielded a time of 130 My for the paleo- and neognathae split (using a kiwi and chicken/zebra finch pair) and 110 My for Galloanseriformes and other neognathae birds (chicken and zebra finch pair). Similar divergence times were obtained using a fossil calibration time of 310 My for the bird-mammal divergence [28]. We also performed a more rigorous Bayesian approach using the program Mutidivtime [29]. The divergence time estimates were 132 My (80 My - 170 My) for the paleo- and neognathae birds and 120 My (70 My - 150 My) for chicken-zebra finch pair, which are similar to those obtained using the direct method (Figure 7). Furthermore, we estimated the divergence times using BEAST without forcing any phylogenetic relationship among the species. Using the three birds and the anole lizard sequences we obtained the divergence times estimate by calibrating the molecular clock using the avian-reptile fossil based divergence time of 255 My. This analysis produced an estimate of 157 My (143 My -173 My), which is slightly higher than that obtained using the previous method. However, the confidence or HPD (Highest Posterior Density) intervals obtained from BEAST were within those obtained using Multidivtime. The divergence time computed by BEAST for the chicken-zebra finch split was 122 My (110 My -134 My), which is similar to that obtained using Multidivtime. Furthermore, we estimated the divergence times directly using the neutral synonymous evolutionary rates estimated by a previous study. A previous study using over 8,000 protein coding genes from chicken, zebra finch and anole lizard estimated the rate of neutral synonymous site evolution to be $1.23 \times 10^{-9}$ to $2.21 \times 10^{-9}$ [17]. Using an average rate of $1.7 \times 10^{-9}$
and a synonymous divergence of 0.452 we estimated the divergence time between kiwi (Paleognathae) and chicken (Neognathae) to be 133 My. Similarly using this rate and the neutral distance of 0.417, the divergence time for chicken-zebra finch pair was computed to be 123 My. Clearly the divergence times estimated in this study using the three different methods are largely similar. Furthermore these time estimates are comparable to those obtained using complete avian mitochondrial genomes [21,22].

Conclusions
Using next generation sequencing technology our study provides some important insights into the conserved kiwi transcriptome. The neutral divergence at conserved protein coding genes of kiwi and chicken was found to be comparable to the synonymous divergence between human and mouse. However the divergence at amino acid replacement positions of these birds is much higher than the mammals suggesting a greater selective pressure in the latter. Similar to the observations from the studies on mammals, a negative relationship between gene expression levels and rate of protein evolution was found in birds. This study provides divergence time estimates between paleognathae and neognathae birds based on >700 nuclear genes.

The conserved kiwi transcriptome data reported here are useful for further specific studies on kiwi genetics and will assist future complete kiwi genome sequencing efforts, specifically in aiding genome assembly and determining gene structure. Importantly, our study provides a cost effective way to perform preliminary genome-based analyses and allows examination of some fundamental developmental and evolutionary processes of a species in the absence of a closely related genome.

Methods
Kiwi embryo
A young male kiwi embryo (sample k11-15) was kindly provided by Suzanne Bassett from the University of Otago, New Zealand. The embryo was void of any discernable structures and resembled an asymmetric gelatinous mass of approximately 15 mm in diameter. The small size and lack of obvious features suggested the embryo was at a very early stage of tissue building.

RNA extraction and preliminary transcriptome characterization
Several 2 mm² slices were removed from equispaced regions around the kiwi embryo, combined, and total RNA was extracted using Trizol™, then precipitated with ethanol and resuspended in 50 μl of Milli-Q water. Five microlitres of total RNA was reverse transcribed in a 50 μl volume containing 50 mM Tris-Cl pH 8.8, 75 mM KCl, 5 mM MgCl₂, 10 mM DTT, 100 ng oligodT₁₈, 0.5 mM of each dNTP, and 200 U of Moloney Murine Leukemia Virus reverse transcriptase. The mix was incubated at 42°C for 1 hr and extracted with phenol-chloroform. The complementary DNA (cDNA) was precipitated with ammonium acetate and ethanol, washed with 80% ethanol, and the resulting pellet was resuspended in 40 ul of H₂O. Complementary DNA was amplified in 10 μl reactions containing 50 mM Tris-Cl pH 8.8, 20 mM (NH₄)₂SO₄, 2.5 mM MgCl₂, 1 mg/ml BSA, 200 μM of each dNTP, 40 ng of each primer, and ~0.3 U of platinum Taq (Invitrogen). The reaction mix was overlaid with mineral oil and subjected to amplification in a Hybaid OmniGene thermal cycler using the following parameters: 94°C for 2 min (× 1), 94°C for 20 sec, 54°C for 20 sec, 72°C for 20 sec (× 15), and then 94°C for 20 sec, 50°C for 20 sec, and 72°C for 20 sec (× 30). Amplified DNAs were detected by agarose gel electrophoresis in Tris-borate-EDTA buffer (TBE), stained with 50 ng/ml ethidium bromide in TBE, and then visualized over UV light. Positive amplifications were purified by centrifugation through ~40 μl of dry Sephadryl™S300HR and then sequenced at the Allan Wilson Centre Genome Sequencing Service using Applied Biosystems (ABI) BigDye® Terminator v3.1 chemistry and an ABI3730 Genetic Analyzer. The primers used were designed to a selection of developmental and regulatory genes. In all cases the primers spanned an intron of at least 500 bp. The primers pairs used and genes targeted were: tbx5_2Fii-agtccaaagagctgcaggctga and tbx5_4R-catcgcctggtaataaatcct; cryn-F-tctgtgacagcatgtcagga and crynR-tctgtgtagaaaaattcacgcca; px6F-acactcgcgacagtcagag and px6R-acaacttcggctggagctct; BMP4-F-tgctgcagatgtttgggct and BMP4R-acaacttcggctggagctct; px1F-gccacctttcacgcggagg and px1R-gctcatggagttgaagaaggg; hxB1F-cgacaccttgattgaaagaggt and hxB1R-tctgtgattgaaagaggt; hxB8F-caaatccagaggtcattacac and hxB8R-gctgtgtagctggcttgaggt; hxD12F-tcaacattgaactgcaagt and hxD12R-ctgtgattgaaacacatgtttgtgaggt; hxB1F-cgacaccttgattgaaagaggt and hxB1R-tctgtgattgaaagaggt; hxB8F-caaatccagaggtcattacac and hxB8R-gctgtgtagctggcttgaggt; hxD12F-tcaacattgaactgcaagt and hxD12R-ctgtgattgaaacacatgtttgtgaggt; hxB1F-cgacaccttgattgaaagaggt and hxB1R-tctgtgattgaaagaggt; hxB8F-caaatccagaggtcattacac and hxB8R-gctgtgtagctggcttgaggt; hxD12F-tcaacattgaactgcaagt and hxD12R-ctgtgattgaaacacatgtttgtgaggt.

Transcriptome preparation and amplification for FLX sequencing
Approximately 10 μg of total RNA was reverse transcribed using oligodT₁₈ as outlined above, and second strand cDNA was synthesized in the same tube by adding 40 μl of × second strand buffer (100 mM Tris-Cl pH 7.0, 25 mM MgCl₂, 450 mM KCl, 50 mM (NH₄)₂SO₄), 4 μl of a 10 mM solution of each dNTP, 7 ul of 100 mM DTT, 20 U of E. coli DNA polymerase I, and water to 200 ul. The mix was incubated at room temp for 2 hrs, before the addition of 5 U of T4 DNA polymerase I to blunt the dsDNA ends, and the dsDNA was purified by phenol-chloroform extraction and ethanol precipitation. The dsDNA pellet was resuspended in
emPCR I shotgun sequencing kit (Roche).

DNA capture beads for 16 emulsions reactions of an the kiwi transcriptome library was annealed to enough density of 1.5 copies per beads was chosen. Following this, library was not titrated. Instead an emPCR loading density of 1.5 copies per beads was chosen. Following this, the kiwi transcriptome library was annealed to enough DNA capture beads for 16 emulsions reactions of an emPCR I shotgun sequencing kit (Roche).

Assembly of FLX sequence reads

The amplified kiwi cDNA library was prepared and sequenced using 454 FLX sequencing chemistry, which generated 75,632 sequence reads with an average length of 171 bp. These reads were used as queries to search a database of 22,000 chicken proteins downloaded from GenBank. We used BLASTX to translate the coding sequences into all six reading frames. Significant threshold levels were based on query protein length, as described before [30]. Homopolymer tracts and adaptors were removed using perl scripts as well as by manual examination. The number of kiwi reads that had significant hits with the chicken proteome were found to be 23,417 (31%). If there were more than two overlapping fragments the most frequent base was used to determine the consensus. These sequences were assembled using the criteria of a 20 bp identical overlap. Using blast2seq, chicken proteins and the translated segments were aligned and assembled. We used a stringent approach and extracted only the regions of kiwi coding sequences that had at least 90% identity to those of chicken. This conservative approach resulted in identification of 1,543 kiwi protein-coding genes with an average aligned (with chicken) length of 168 bp. This alignment did not show any bias in the genic location of the kiwi reads, as roughly 49% of the reads are from the 3’ terminus of the genes and 51% are from the 5’ terminus. Redundant (identical or subsets of) sequences were excluded from further analysis.

Evolutionary rate estimation

The rate of evolution was estimated using conserved kiwi-chicken sequences. Sequence alignments from all genes were concatenated and the divergences at synonymous and nonsynonymous sites were estimated using the pair-wise option of the CODEML program [31]. This approach was also followed for groups of genes such as highly expressed genes. Furthermore pair-wise likelihood distances at synonymous and nonsynonymous sites were also estimated for individual genes. Since the lengths of the sequences recovered were short (~168 bp) the dS estimates for individual genes are subjected to higher stochastic errors than the dN estimates, which might result in overestimation of the dN/dS ratios due to very small dS values. However we found only 10 genes with a very low dS (< 0.01) and therefore the genic dN/dS ratios obtained for most of the genes were not affected by this overestimation.

Estimation of divergence times

Using a reciprocal BLAST hit approach [32] orthologous genes from three other vertebrates zebra finch, anole lizard, and human were also obtained. Protein sequences of five genomes (including chicken and kiwi) were aligned using CLUSTALW [33] and only those regions that aligned with the partial kiwi proteins were extracted. cDNA sequences of all these genomes were aligned using the protein alignments as a guide. The resultant 702 genes from five genomes were concatenated. Since the phylogenetic relationship between these five species is well known, the tree topology (as in Figure 7) was used to obtain the branch lengths using the program BASEML from PAML [31]. For this analysis a GTR+gamma (five categories) model was used.

To estimate divergence times we followed a Bayesian based approach implemented in the software Multidivtime [29,34]. The molecular clock was calibrated using the well documented fossil-based estimate of 255 My (252 My - 257 My) for the reptile-avian split [27] and the human sequence was used as an outgroup. The lower and upper constraints used in the program were 230 My and 280 My. We used 255 My as the expected time between the (ingroup) root to the tip (rttime). The prior rate was calculated as the ratio of the median of the branch lengths from root-to-tip and the time elapsed as per the suggestion given in the documentation (Thorne and Kishino 2002). The prior standard deviation was kept as 50 My. Other priors used were as outlined in the Multidivtime documentation. Furthermore we used BEAST [35] to estimate the divergence times without constraining any phylogenetic relationship among the species. For this purpose we used three birds and the lizard protein coding sequences. We used the Tamura Nei +Gamma model for sequence evolution and the reptile-bird fossil based divergence time to calibrate the molecular clock.
Nature 11 Paleontological evidence to date the tree of life. Mol Biol Evol 2007, 24:26-53.

19. Kunstner A, Wolf JBW, Backstrom N, Schierup M, Balakrishnan CN, Day L, Nelson J, Chinwalla A, Hou S, Mardis ER, Wilson RK:

20. Kumar S, Hedges SB:

21. Kumar S, Hedges SB:

22. Pereira SL, Baker AJ: A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Mol Biol Evol 2006, 23:1731-1740.

23. Benton MJ, Donoghue PC: Paleontological evidence to date the tree of life. Mol Biol Evol 2007, 24:26-53.

24. Pal C, Papp B, Hurst LD: Highly expressed genes in yeast evolve slowly. Genetics 2001, 158:927-931.

25. Subramanian S, Kumar S: Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. Genetics 2004, 168:373-381.

26. Subramanian S: Neutrality and the evolution of coding usage bias in eukaryotic genomes. Genetics 2008, 178:2429-2432.

27. Reiz RR, Muller J: Molecular timescales and the fossil record: a paleontological perspective. Trends Genet 2004, 20:237-241.

28. Benton MJ: The fossil record 2. London: Chapman and Hall, London, 1993.

29. Thorne JL: MULTIDISTANCE 2003 [http://statgen.ncsu.edu/thorne/multidistance.html].

30. Duret L, Mouchiroud D, Gouy M: Hovergen - a Database of Homologous Vertebrate Genes. Nucleic Acids Res 1994, 22:2360-2365.

31. Yang ZH: PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 2007, 24:1586-1591.

32. Atchulo SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997, 25:3389-3402.

33. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan P, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: Clustal W and clustal X version 2.0. Bioinformatics 2007, 23:2947-2948.

34. Thorne JL, Kishino H: Divergence time and evolutionary rate estimation with multilocus data. Syst Biol 2002, 51:689-702.

35. Drummond AJ, Rambaut A: BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 2007, 7:214.

36. Subramanian S, Kumar S: Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. Genetics 2004, 168:373-381.

37. Reiz RR, Muller J: Molecular timescales and the fossil record: a paleontological perspective. Trends Genet 2004, 20:237-241.

38. Duret L, Mouchiroud D, Gouy M: Hovergen - a Database of Homologous Vertebrate Genes. Nucleic Acids Res 1994, 22:2360-2365.

39. Yang ZH: PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 2007, 24:1586-1591.

40. Atchulo SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997, 25:3389-3402.

41. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan P, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: Clustal W and clustal X version 2.0. Bioinformatics 2007, 23:2947-2948.

42. Thorne JL, Kishino H: Divergence time and evolutionary rate estimation with multilocus data. Syst Biol 2002, 51:689-702.

43. Drummond AJ, Rambaut A: BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 2007, 7:214.

Cite this article as: Subramanian et al.: Next generation sequencing and analysis of a conserved transcriptome of New Zealand’s kiwi. BMC Evolutionary Biology 2010 10:387.

doi:10.1186/1471-2148-10-387

Submit your next manuscript to BioMed Central and take full advantage of:

• Convenient online submission
• Thorough peer review
• No space constraints or color figure charges
• Immediate publication on acceptance
• Inclusion in PubMed, CAS, Scopus and Google Scholar
• Research which is freely available for redistribution

Submit your manuscript at www.biomedcentral.com/submit