The First Chameleon Transcriptome: Comparative Genomic Analysis of the OXPHOS System Reveals Loss of *COX8* in Iguanian Lizards

Dan Bar-Yaacov, Amos Bouskila, and Dan Mishmar*

Department of Life Sciences, Ben Gurion University of the Negev, Beer Sheva, Israel

*Corresponding author: E-mail: dmishmar@bgu.ac.il.

Accepted: August 29, 2013

Data deposition: The transcriptome is available at http://lifeserv.bgu.ac.il/wb/dmishmar/pages/supplementary-files.php). The transcriptome sequences that were generated were deposited in Short Reads Archive (SRA, http://www.ncbi.nlm.nih.gov/sra) with the accession #SRP018939.

Abstract

Recently, we found dramatic mitochondrial DNA divergence of Israeli *Chamaeleo chamaeleon* populations into two geographically distinct groups. We aimed to examine whether the same pattern of divergence could be found in nuclear genes. However, no genomic resource is available for any chameleon species. Here we present the first chameleon transcriptome, obtained using deep sequencing (SOLiD). Our analysis identified 164,000 sequence contigs of which 19,000 yielded unique BlastX hits. To test the efficacy of our sequencing effort, we examined whether the chameleon and other available reptilian transcriptomes harbored complete sets of genes comprising known biochemical pathways, focusing on the nDNA-encoded oxidative phosphorylation (OXPHOS) genes as a model. As a reference for the screen, we used the human 86 (including isoforms) known structural nDNA-encoded OXPHOS subunits. Analysis of 34 publicly available vertebrate transcriptomes revealed orthologs for most human OXPHOS genes. However, OXPHOS subunit *COX8* (Cytochrome C oxidase subunit 8), including all its known isoforms, was consistently absent in transcriptomes of iguanian lizards, implying loss of this subunit during the radiation of this suborder. The lack of *COX8* in the suborder Iguania is intriguing, since it is important for cellular respiration and ATP production. Our sequencing effort added a new resource for comparative genomic studies, and shed new light on the evolutionary dynamics of the OXPHOS system.

Key words: chameleon, oxidative phosphorylation, transcriptome.

Introduction

Massive parallel sequencing (MPS) enables identifying the entire set of transcribed genes (transcriptome) of understudied organisms, thus providing novel genomic resources. However, because there is no genomic reference to those organisms, the short reads generated by MPS must be de novo assembled in order to form sequence contigs, which in turn could be annotated (Kusumi et al. 2011), thus creating reference sequences for further analyses.

Recently, we found sharp mitochondrial DNA (mtDNA) divergence of *Chamaeleo chamaeleon* populations into two geographically distinct groups in Israel: one ranging from the Jezreel Valley to the north and the other ranging from the Jezreel Valley to the south (Bar-Yaacov et al. 2012). The division of mtDNA clusters was absolute, not even a single specimen carrying a northern mtDNA was identified south of the Jezreel Valley and vice versa. Bayesian coalescence analyses (BEAST) (Drummond and Rambaut 2007) supported a long separation (more than 1 million years), which correlated well with the existence of an ancient marine barrier at the Jezreel Valley, exactly where the mtDNA clusters met. We aimed at examining whether the same pattern of mitochondrial divergence could be found in nuclear genes, especially nuclear DNA-encoded mitochondrial genes. However, the lack of genomic resource for any chameleon species posed a major obstacle. Moreover, reptiles in general are understudied with little available genomic resources, mainly harboring mtDNA sequences and few nDNA-encoded genes (Macey et al. 2008; Alfoldi et al. 2011; Kusumi et al. 2011; Tezuka et al. 2012). Recent advances in MPS technologies enabled sequencing the first reptilian genome, the genome of *Anolis carolinensis* (Alfoldi et al. 2011), and more recently, several other reptilian transcriptomes (Schwartz et al. 2010; Castoe

© The Author(s) 2013. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

et al. 2011; Tzika et al. 2011). Here we present the first chameleon transcriptome, its annotation, and its usage to perform comparative genomic analysis that revealed novel insights into the evolution of the entire mitochondrial oxidative phosphorylation (OXPHOS) system in reptiles and other vertebrates. The chameleon transcriptome will constitute a new genomic resource for further genetic studies.

Materials and Methods

RNA Extraction and Sequencing

We received a chameleon specimen that was collected by Israel Nature and Parks Authority personnel after it was hit by a car in the north of Israel. The chameleon was euthanized using isofloran and was dissected several minutes postmortem. Isolated brain, lungs, skeletal muscle, and heart were then snap-frozen in liquid nitrogen. Total RNA was extracted from the above-mentioned tissues using Perfect pure RNA kit (5 Prime). RNA concentration was estimated using nano-drop (NanoDrop Technologies). Clear rRNA bands were visualized on a 1% agarose gel to further assure RNA sample quality. RNA from the four tissues was mixed into a single tube in the following amounts: brain 12.1 μg, lungs 5.3 μg, heart 2.7 μg, and skeletal muscle 5.2 μg. Notably, the RNA from heart constitutes the entire preparation of this tissue; excess of brain RNA was introduced instead, to reach the amount required for sequencing library preparation. The RNA was subjected to library preparation using the SOLiD total RNA-Seq kit and the complete transcriptome was sequenced using the SOLiD 4 platform (Applied Biosystems) at the Hebrew University genomics center. The specimen was recorded in the Hebrew University of Jerusalem, Reptiles collection, Voucher #HUJR-24101, and was stored in −80°C at the Life Sciences Department, Ben Gurion University of the Negev, Beer Sheva, Israel.

Identifying the High-Quality Sequence Data and De Novo Assembly of Sequence Contigs

SOLiD sequencing resulted in 110 million paired reads of 50 and 35 bp (SRA accession number #SRP018939). GALAXY (Giardine et al. 2005) was used to filter out reads that had less than 70% bases with Phred scale greater than 23. This left us with ~55 million paired reads that were subjected to de novo assembly using CLC-Bio assembly cell 4. The best results were received using the default parameters; however, we focused on transcript contigs longer than 100 bp in length. More than 76% of the reads mapped back to the assembled transcriptome, thus confirming that most of the reads were used during the assembly process.

Annotation of the Chameleon Transcriptome

The assembled contigs were annotated using Blast2GO (Conesa et al. 2005) (fig. 1). Specifically, our transcriptome sequences were screened using BlastX against the NCBI NR database. A "HIT" for a contig was listed only if it had a value greater than 1.0E−6. The best hits were ranked according to Blast2GO default parameters. Mapping and annotation steps were performed using the Blast2GO default parameters. Supplementary table S1, Supplementary Material online, summarizes all the assembled contigs and Blast hits which are available online (http://lifeserv.bgu.ac.il/wb/dmishmar/pages/supplementary-files.php, last accessed September 18, 2013). Blast2GO was used to construct a biological process graph using 36,740 human annotated transcripts.

Comparative Analysis of nDNA-Encoded Orthologs of OXPHOS Human Genes in 34 Vertebrates

We downloaded from NCBI all the available RefSeq transcripts of *Pan troglodytes*, *Pongo abelii*, *Nomascus leucogenys*, *Macaca mulatta*, *Callithrix jacchus*, *Sus scrofa*, *Bos taurus*, *Equus caballus*, *Loxodonta africana*, *Ailuropoda melanoleuca*, *Canis lupus familiaris*, *Mus musculus*, *Rattus norvegicus*, *Cricetulus griseus*, *Cavia porcellus*, *Oryctolagus cuniculus*, *Monodelphis domestica*, *Ornithorhynchus anatinus*, *A. carolinensis*, *Taeniopygia guttata*, *Gallus gallus*, *Meleagris gallopavo*, *Xenopus (silurana) tropicalis*, *Danio rerio*, and *Oreochromis niloticus*. We also downloaded available assembled transcripts from recently sequenced vertebrates, including *Thamnophis elegans*, *Python molurus bivittatus*, *Pogona vitticeps*, *Elaphe guttata*, *Trachemys scripta*, *Crocodylus niloticus*, *G. gallus*, *Tetraodon nigroviridis*, *Fugu rubripes* (Jaillon et al. 2004; Schwartz et al. 2010; Castoe et al. 2011; Kai et al. 2011; Tzika et al. 2011). Notably, the recently sequenced *G. gallus* transcriptome gave better results than the available RefSeq transcripts; therefore we used those transcripts in further analysis. We then downloaded 86 known human nDNA-encoded OXPHOS proteins sequences and constructed a local Blast database (Blast 2.2.25+ [Altschul et al. 1997]). Blast screen was performed for each transcriptome against the OXPHOS human genes to identify orthologs. A contig was considered a hit if its similarity value was above 1.0E−5, following recently used threshold (Schwartz et al. 2010; Castoe et al. 2011). Additionally, for each OXPHOS subunit, only contigs having the lowest e-value were further analyzed. Then, in order to exhaust all publicly available data, an additional Blast (TBlastN, BlastP, and BlastN) search was performed for each of the species in which we analyzed RefSeq transcripts, using the entire NCBI database (nr) and all available genomes in NCBI. Figure 2 specifies the identification of each subunit in the transcriptomes and (when available) genomes of each species.

The schematic tree representing all tested species was designed following the taxonomy published in NCBI, which is also consistent with a recently published phylogenetic study (Vidal and Hedges 2009) (http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/taxtree.cgi?db=Mito&taxid=2759&

result=frame&complete=All&init_rankid=1, last accessed September 18, 2013).

Results

The Chameleon Transcriptome

We aimed at sequencing as many unique *C. chamaeleon* transcripts as possible. For this purpose, and to avoid tissue specificity, we subjected mixture of RNA samples from four tissues (brain, lungs, muscle, and heart) extracted from a single *C. chamaeleon* specimen to MPS using the SOLiD ABI platform. MPS yielded a total of 9.35 Gbp in 110 million forward (50 bp) and reverse (35 bp) reads of which 5 Gbp were high-quality reads.

We de novo assembled all the high-quality reads using CLC-BIO assembly cell (default parameters) yielding 164,525 contigs (>100 bp). The average contig length was 169 bp (range: 102–3,085 bp, with 95% of the contigs being no longer than 350 bp) with a mean coverage of 107× per nucleotide position (range: 3.3× to 286,258×, with 95% of the contigs having no more than 160× coverage). This analysis resulted in more sequence contigs than generated by previous studies which used various sequencing technologies, though the average length of our contigs was shorter.

Table 1

Summary Statistics of the Assembly, Blast Hits, and Annotation of the Chameleon Transcriptome

Assembled contigs	**164,525**
Contigs with BlastX hits	42,741
Nonredundant contigs with BlastX hits	19,086
Annotated contigs	36,740
Contigs with no BlastX hits	121,784

Annotation of the Chameleon Transcriptome

As the first step toward understanding the gene content of the *C. chamaeleon* transcriptome contigs, we used Blast2GO

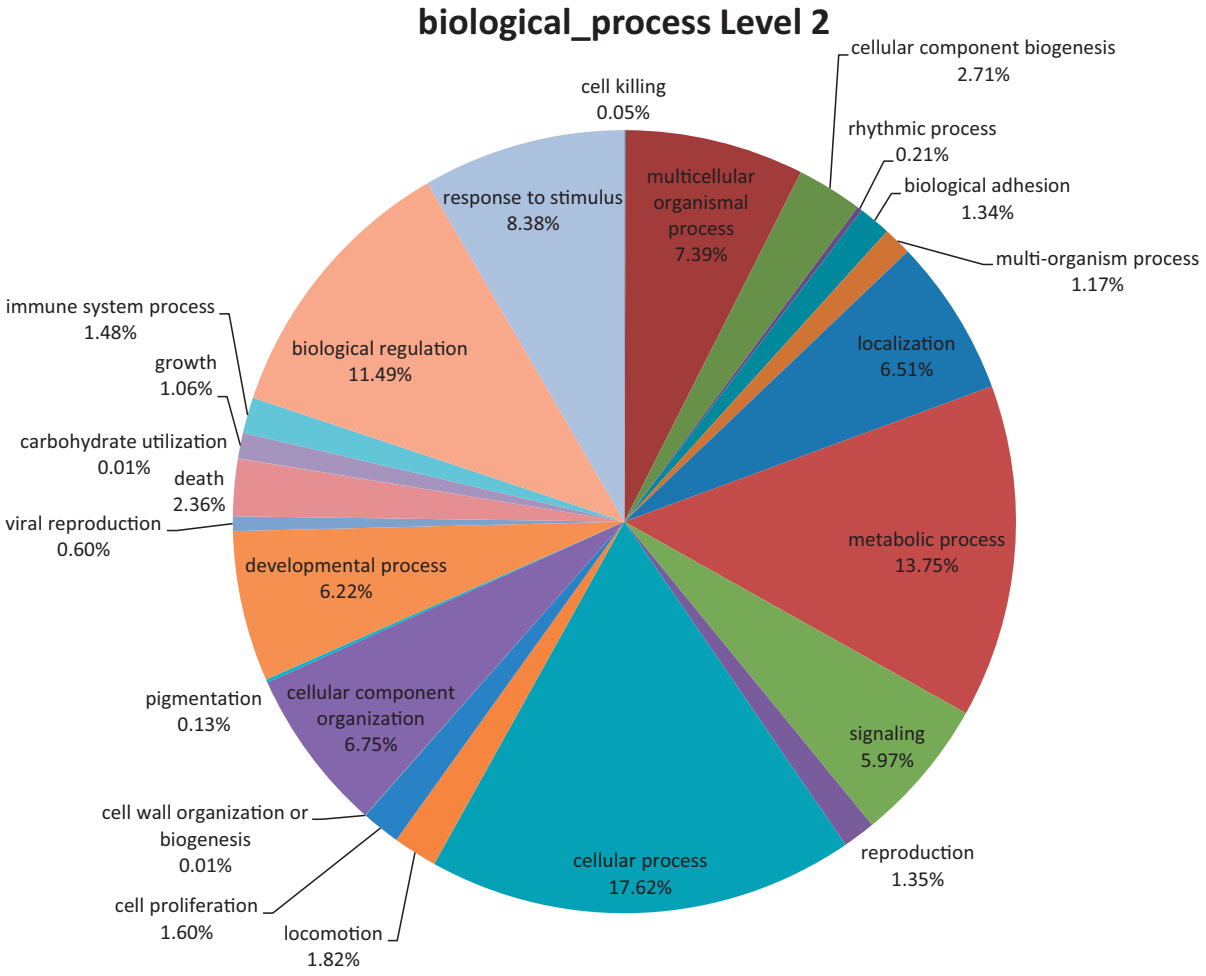

Fig. 1.—Pie chart summary of biological processes in the *Chamaeleo chamaeleone* transcriptome. The chart was assembled using Blast2GO. Notably, certain contigs (genes) could be listed in multiple processes.

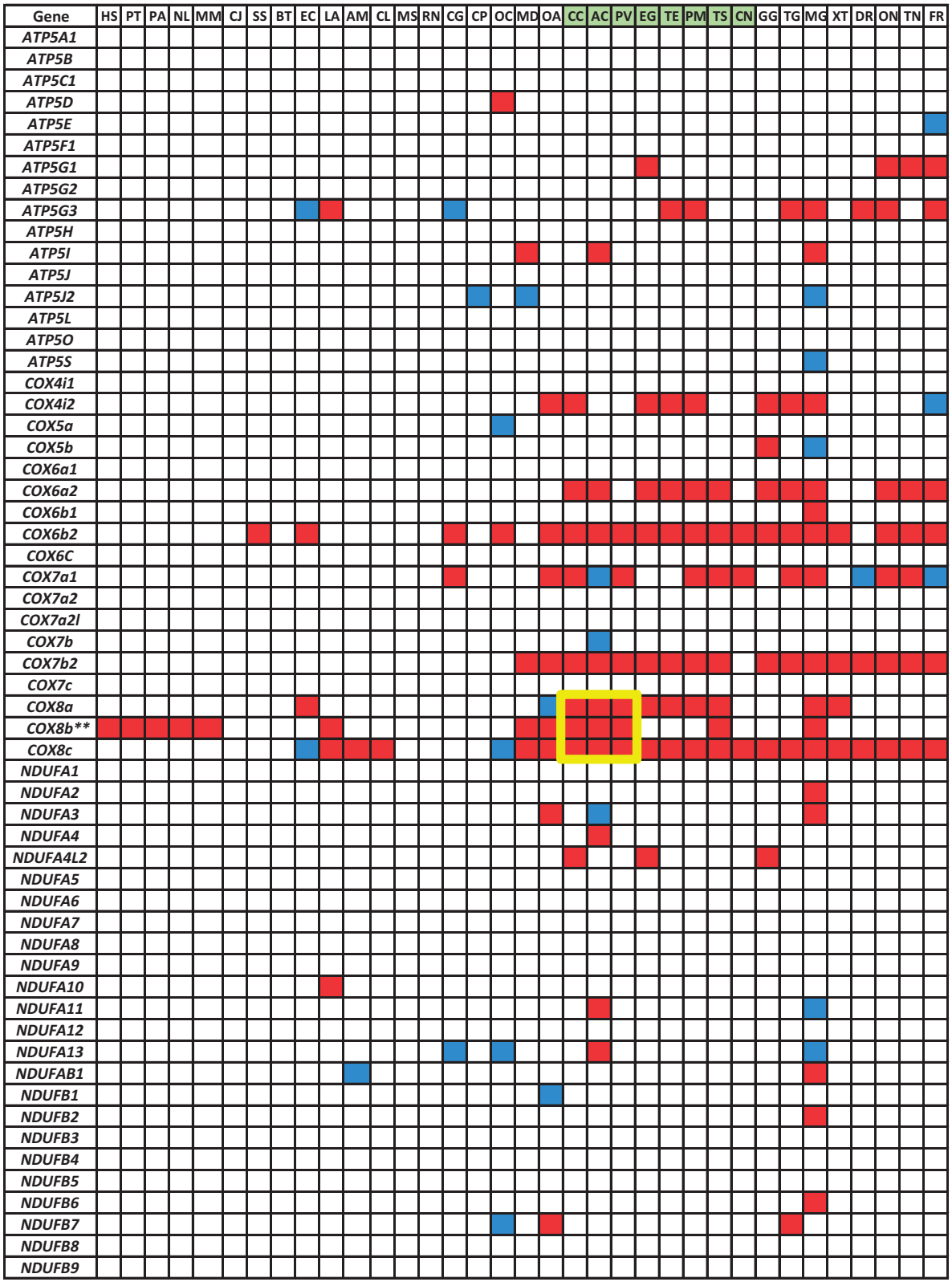

Fig. 2.—Orthologs of nDNA-encoded OXPHOS human genes in 34 vertebrate transcriptomes. Red box: missing ortholog. Blue box: ortholog identified only in the whole genome sequence of the relevant species*. Framed in yellow: missing *COX8* in iguanian lizards. Green background: reptilian species. Species name abbreviations: HS, *Homo sapiens*; PT, *Pan troglodytes*; PA, *Pongo abelii*; NL, *Nomascus leucogenys*; MM, *Macaca mulatta*; CJ, *Callithrix jacchus*;

(continued)

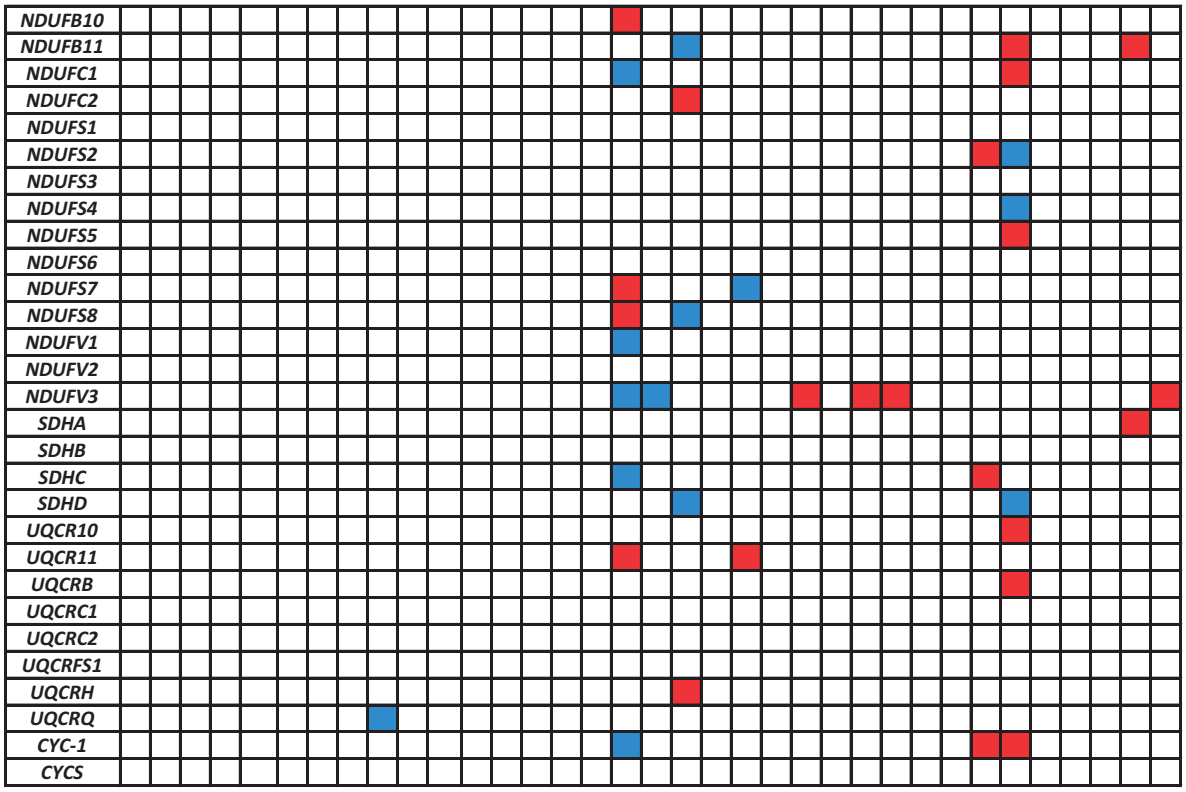

FIG. 2.—Continued

SS, *Sus scrofa*; BT, *Bos taurus*; EC, *Equus caballus*; LA, *Loxodonta africana*; AM, *Ailuropoda melanoleuca*; CL, *Canis lupus familiaris*; MS, *Mus musculus*; RN, *Rattus norvegicus*; CG, *Cricetulus griseus*; CP, *Cavia porcellus*; OC, *Oryctolagus cuniculus*; MD, *Monodelphis domestica*; OA, *Ornithorhynchus anatinus*; CC, *Chamaeleo chamaeleon*; AC, *Anolis carolinensis*; PV, *Pogona vitticeps*; TE, *Thamnophis elegans*; EG, *Elaphe guttata*; PM, *Python molurus bivittatus*; TS, *Trachemys scripta*; CN, *Crocodylus niloticus*; GG, *Gallus gallus*; TG, *Taeniopygia guttata*; MG, *Meleagris gallopavo*; XT, *Xenopus (silurana) tropicalis*; DR, *Danio rerio*; ON, *Oreochromis niloticus*; TN, *Tetraodon nigroviridis*; FR, *Fugu rubripes*. *NCBI genome sequences were not available for the following species: CC, PV, TE, EG, PM, TS, CN, and TN. ***Cox8b* sequence was extracted from a mouse reference as it was absent in humans.

(Conesa et al. 2005). Our identified contigs were screened using the BlastX algorithm against the NR protein database (NCBI). This screen yielded 42,741 BlastX hits with values higher than 1.0E−6 (table 1 and supplementary table S1-sheet 1, Supplementary Material online). As anticipated, the majority of top BlastX hits (the best hit for each contig, namely having the top Blast score according to Blast2GO default parameters, constituting more than 19,000 hits) aligned with *A. carolinensis*. After examining the results, we identified a total of 19,086 nonredundant transcripts (of which 10,095 [52.82%] had orthologs in the available *A. carolinensis* NCBI protein database) (table 1 and supplementary table S1-sheet 2, Supplementary Material online). Biological process analysis (in the lowest level) revealed that the identified transcripts harbored orthologs of genes from all major biological functions, suggesting high transcript representation (fig. 1). Notably, the various biological processes in our analysis are more evenly distributed than those found recently in a python (*P. molurus bivittatus*) transcriptome analysis (Castoe et al. 2011). We interpret this result as possible differences in the sampled tissues in our and the python studies.

Comparative Analysis of nDNA-Encoded OXPHOS Genes in 34 Vertebrates

To assess the quality of our *C. chamaeleon* transcriptome, we aimed at identifying complete sets of genes encompassing well-studied biochemical pathways. To this end, and because of our initial motivation, we focused on the OXPHOS nDNA-encoded genes and analyzed 34 additional publicly available vertebrate transcriptome sequences including 18 mammals, 3 birds, 8 reptiles, 1 amphibian, and 4 bony fish species.

Because the most complete set of nDNA-encoded OXPHOS genes was mainly studied and recorded in humans, we created a local Blast database from 86 human nDNA-encoded OXPHOS structural subunits (including subunit isoforms) and compared all available transcriptome

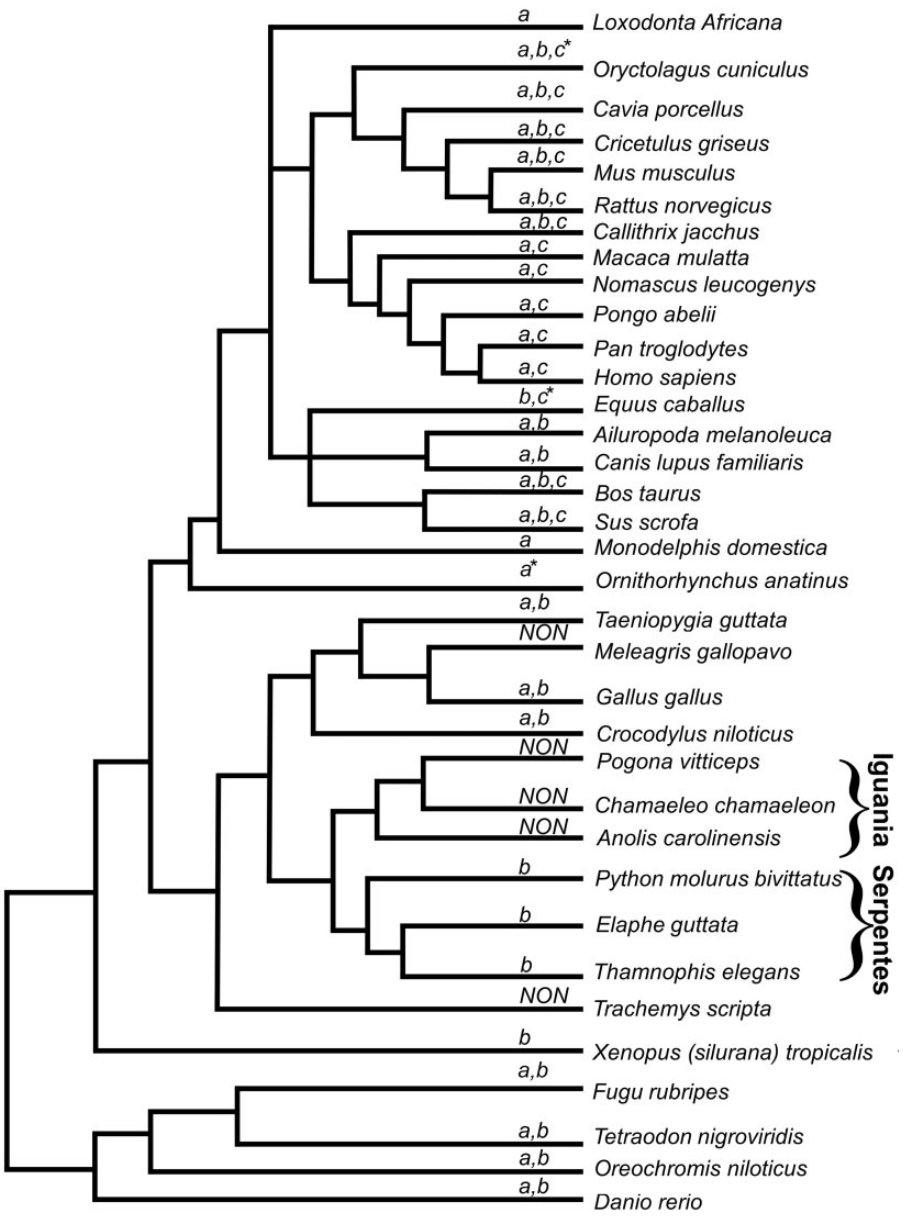

Fig. 3.—Schematic phylogenetic tree demonstrating the presence or absence of *COX8* across the vertebrate phylogeny. Letters above each branch indicate the presence of the relevant *COX8* isoforms. Non: total absence of all *COX8* isoforms. Notably, the suborders Iguania and Serpentes are labeled. The topology of the tree is adopted from NCBI Taxonomy (see URL in Material and Methods), which is also consistent with a recently published phylogenetic study (Vidal and Hedges 2009). *represents detection of *COX8* in the genome sequence of the relevant organism.

sequences with this database. The majority of the human OXPHOS genes had orthologs in the transcriptomes of most studied organisms, whereas the *M. gallopavo* (Turkey) transcriptome yielded the lowest amount of orthologs (56) (fig. 2), possibly reflecting missing data in that organism. This explanation could apply to other species with lower numbers of identified OXPHOS orthologs. In the *C. chamaeleon* transcriptome we identified 78 human orthologs (including isoforms), an amount similar to the other recently sequenced reptilian transcriptomes. The most prominent finding was the lack of *COX8* (including its human isoforms) in all reptile transcriptomes, excluding the crocodile, which is phylogenetically closer to birds than to other reptiles (Gauthier et al. 1989). When we extended our database search to find additional *COX8* isoforms using the mouse *COX8B* sequence as a reference, we identified *COX8B* orthologs in all tested Serpentes (snakes), *T. elegans*, *P. molurus bivittatus*, and *E. guttata*, but not in the examined iguanian lizards, *C. chamaeleon*, *A. Carolinensis*, and *Pog. vitticeps*, as well as the terrapin *Tra. scripta* (fig. 3).

Discussion

Sequencing the chameleon transcriptome added a novel genomic resource of a nonmodel organism from a completely understudied reptilian family (Chamaeleonidae). This genomic resource could be utilized for comparative genomics, ecological research, and species-specific genetic studies. Our approach using RNA that was extracted and mixed from four different tissues generated a high coverage transcriptome while controlling, at least in part, for tissue specificity. The amount of nonredundant transcripts (19,086) that we identified was similar to recently sequenced reptilian transcriptomes, but our contigs (and therefore our transcript sequences) were shorter, likely due to the used sequencing platform (SOLiD ABI). Most of the identified transcripts best aligned to the genome sequence of *A. carolinensis*, which is the only reptile whose genome was completely sequenced and published, thus fortifying the validity of our sequencing effort.

Our comparative genomic analysis of nDNA-encoded OXPHOS genes identified most of the human gene orthologs in the majority of the studied species (transcriptomes and when available whole genome sequence data), thus indicating that both the *C. chamaeleon* and recently sequenced transcriptomes produced quality genomic resources enabling the identification of complete sets of genes in previously understudied organisms. Notably, *COX8* was absent in all examined lizards, which belonged to the suborder Iguania (Macey et al. 1997), a sister taxon of the Serpentes suborder (Vidal and Hedges 2009) (fig. 3). The presence of *COX8* in transcriptomes belonging to representatives of all other examined taxa (mammals, birds, crocodiles, Serpentes, amphibians, and bony fish) suggests that it was lost (at least transcription wise) during the radiation of iguanian lizards. Notably, some individual species lack *COX8* and its isoforms, despite the existence of this gene in closely related sister taxa, such as in the case of the avian *M. gallopavo*. Additionally, *COX8* was also absent in certain species that were the only representatives of their taxa in our analysis (such as the turtle *Tra. scripta*). In such cases, we currently cannot discriminate between possible true absence of the gene and technical partial representation of genes in such transcriptomes. Eventually, sequencing transcriptomes of additional species will likely shed light on the dynamics of the OXPHOS system, in general, and of *COX8* in particular.

Close inspection of figure 3 indicates the presence of orthologs to all identified *COX8* isoforms in some species, and only to some isoforms in others. The identification of orthologs to only a subset of *COX8* isoforms could either be due to the tissue-specific expression of some of these isoforms or due to actual absence of these paralogs from the genomes of some species. However, until high-quality genome sequences of more organisms are available, this caveat cannot be easily resolved.

COX8 was previously shown to be important for cellular respiration and ATP production, by specifically increasing the functional efficiency of OXPHOS complex IV (cytochrome c oxidase) (Patterson and Poyton 1986). It was previously argued that *COX8B* became transcriptionally silenced in humans and other primates, but could be identified in the transcriptomes of other mammals and vertebrates (Goldberg et al. 2003). In our analysis, *COX8B* was found in Serpentes but not in iguanian lizards, implying the complete loss off all *COX8* isoforms in iguanian lizards (figs. 2 and 3). This finding raises the question of functional compensation to maintain the activity of OXPHOS complex IV in iguanian lizards. In conclusion, our sequencing effort added a new resource for chameleon genetics which is useful for comparative genomic studies, and sheds new light on the evolutionary dynamics of the OXPHOS system.

Supplementary Material

Supplementary tables S1 is available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank the Kreitman foundation for a partial scholarship of excellence awarded to D.B.Y. They also thank the Israel Nature and Parks Authority for collecting the chameleon and for issuing the permits for our work on wild chameleons. This work was supported by an Israel Science Foundation grant (grant number 610/12) awarded to D.M.

Literature Cited

Alfoldi J, et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature 477:587–591.

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

Bar-Yaacov D, et al. 2012. Mitochondrial DNA variation, but not nuclear DNA, sharply divides morphologically identical chameleons along an ancient geographic barrier. PLoS One 7:e31372.

Castoe TA, et al. 2011. A multi-organ transcriptome resource for the Burmese Python (*Python molurus bivittatus*). BMC Res Notes. 4: 310.

Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676.

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

Gauthier J, Cannatella D, de Queiroz K, Kluge A, Rowe T. 1989. Tetrapod phylogeny. In: Fernholm B, Bremer K, Jornvall H, editors. The hierarchy of life: proceedings of the 70th Nobel Symposium. Amsterdam (The Netherlands): Elsevier Science Publishers B.V. p. 337–353.

Giardine B, et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 15:1451–1455.

Goldberg A, et al. 2003. Adaptive evolution of cytochrome c oxidase subunit VIII in anthropoid primates. Proc Natl Acad Sci U S A. 100: 5873–5878.

Jaillon O, et al. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431: 946–957.

Kai W, et al. 2011. Integration of the genetic map and genome assembly of fugu facilitates insights into distinct features of genome evolution in teleosts and mammals. Genome Biol Evol. 3:424–442.

Kusumi K, et al. 2011. Developing a community-based genetic nomenclature for anole lizards. BMC Genomics 12:554.

Macey JR, et al. 2008. Socotra Island the forgotten fragment of Gondwana: unmasking chameleon lizard history with complete mitochondrial genomic data. Mol Phylogenet Evol. 49:1015–1018.

Macey JR, Larson A, Ananjeva NB, Papenfuss TJ. 1997. Evolutionary shifts in three major structural features of the mitochondrial genome among iguanian lizards. J Mol Evol. 44:660–674.

Patterson TE, Poyton RO. 1986. *COX8*, the structural gene for yeast cytochrome c oxidase subunit VIII. DNA sequence and gene disruption indicate that subunit VIII is required for maximal levels of cellular respiration and is derived from a precursor which is extended at both its NH2 and COOH termini. J Biol Chem. 261:17192–17197.

Schwartz TS, et al. 2010. A garter snake transcriptome: pyrosequencing, de novo assembly, and sex-specific differences. BMC Genomics 11: 694.

Tezuka A, et al. 2012. Comprehensive primer design for analysis of population genetics in non-sequenced organisms. PLoS One 7:e32314.

Tzika AC, Helaers R, Schramm G, Milinkovitch MC. 2011. Reptilian-transcriptome v1.0, a glimpse in the brain transcriptome of five divergent Sauropsida lineages and the phylogenetic position of turtles. Evodevo 2:19.

Vidal N, Hedges SB. 2009. The molecular evolutionary tree of lizards, snakes, and amphisbaenians. C R Biol. 332:129–139.

Associate editor: Dan Graur