GENETIC BASES OF APOSEMATIC TRAITS: INSIGHTS FROM THE SKIN

TRANSCRIPTIONAL PROFILES OF Oophaga POISON FROGS

Andrés Posso-Terranova$^{1,2\ast}$, José Andres$^{1,3}$.

1. Department of Biology. University of Saskatchewan. 112 Science Pl. Saskatoon. SK. Canada.
14853.

*Corresponding author
Email: andres.posso@usask.ca

All authors contributed equally to this work.
ABSTRACT

Aposematic organisms advertise their defensive toxins to predators using a variety of warning signals, including bright coloration. While most Neotropical poison frogs (Dendrobatidae) rely on crypsis to avoid predators, Oophaga poison frogs from South America advertise their chemical defenses, a complex mix of diet-derived alkaloids, by using conspicuous hues. The present study aimed to characterize the skin transcriptomic profiles of the South American clade of Oophaga poison frogs (O. anchicayensis, O. solanensis, O. lehmanni and O. sylvatica). Our analyses showed very similar transcriptomic profiles for these closely related species in terms of functional annotation and relative abundance of gene ontology terms expressed. Analyses of expression profiles of Oophaga and available skin transcriptomes of cryptic anurans allowed us to propose possible mechanisms for the active sequestration of alkaloid-based chemical defenses and to highlight some genes that may be potentially involved in resistance mechanisms to avoid self-intoxication and skin coloration. In doing so, we provide an important molecular resource for the study of warning signals that will facilitate the assembly and annotation of future poison frog genomes.

KEYWORDS

Dendrobatids, RNA sequencing, transcriptomes, gene ontology, candidate genes.
INTRODUCTION

Aposematic organisms advertise their defensive toxins to predators using a variety of warning signals [1-3]. Among aposematic species, the poison frogs (Dendrobatidae) from the tropical rain forests of Central and South America represent one of the most spectacular examples of warning coloration. While the majority of dendrobatids rely on crypsis to avoid predators, some members of this family are both brightly colored and chemically defended [4]. From a genetic point of view, these aposematic defenses can be defined as a complex phenotype resulting from the integration (i.e. covariation) of different genetic elements related to conspicuousness, bold behavior, unpalatability, diet specialization, etc. While a long history of research has been devoted to understand the genetics of warning coloration in arthropods, particularly in Heliconius butterflies [5, 6], the molecular underpinnings of aposematism in vertebrates, particularly the mechanisms whereby individuals become toxic (or distasteful) remain mostly unknown. In this study, our primary aim was to characterize the skin transcriptomic profiles of Oophaga species as a first attempt to shed light on the molecular genetic bases of aposematic components in poison frogs: the ability to sequester alkaloid-based chemical defenses, warning coloration, and the resistance mechanisms to avoid self-intoxication.

The frogs of this complex inhabit the lowland Pacific rainforests of the Colombian and Ecuadorian Chocó. Preliminary molecular data from nuclear and mitochondrial markers showed a similar genetic background among lineages in contrast to an extraordinary diversity of morphotypes [7]. Individuals from different allopatric lineages can be relatively homogeneous, striped, or spotted, and their colors range from bright red, to orange and yellow (Fig 1) [8]. These polymorphic coloration patterns serve
as a warning signal of their chemical defenses [9], a complex mix of diet-derived alkaloids secreted by the dermal glands [10, 11]. Although these chemicals may involve metabolism and transport through other tissues, the final accumulation of toxins as well as the production of pigmentation cells is performed in the skin tissue of these organisms even during the adult stage [12-14]. Thus, we expected that the genes, pathways, and/or gene networks potentially associated with coloration, alkaloid metabolism, transport and storage, to be expressed in the skin of these lineages.

Although the lineages of the *Oophaga* studied here display a wide variety of warning signals, most of them share a black background coloration (Fig 1). In adult frogs, this melanistic coloration is controlled by the lack of iridophores and xanthophores in the dermal chromatophore unit, as well as the size and dispersion pattern of the melanin-containing organelles (melanosomes) within the dermal melanophores [15]. Thus, we also hypothesized that in adult individuals, genes involved in the amount, size and distribution of the melanosomes should be differentially expressed in the skin. Similarly, because the secreted alkaloids of chemically-defended dendrobatids are known to disrupt the normal ion-channel activity and to alter the neurotransmitter-receptor binding capacity of nerve and muscle cells [16-18], we hypothesized that the fraction of genes differentially expressed in the skin of poison frogs (when compared to non-aposematic species) is also enriched in genes related to alkaloid sequestration, transport and autotoxicity avoidance.
Fig 1. Hierarchical clustering (bootstrap=999) and heatmap contrasting the differentially expressed unigenes from skin tissue between *Oophaga* and cryptic species. Color within the heatmap represent the log2 fold change values. Dark blue indicates upregulation while a lighter coloration indicates downregulation. Rows of the heatmap correspond to the differentially expressed unigenes (n=1,931) detected in this study.
To test these hypotheses, we first assembled the skin transcriptomes of five different Harlequin poison frogs species: *O. histrionica* [19], *O. lehmanni* [20], *O. sylvatica*, [21] and the recently described *O. anchicayensis* and *O. solanensis* [22]. Then, we compared the transcription profiles of these aposematic species with that of a diverse group of cryptic anurans. Functional annotation of unigenes showing a phylogenetic signal (*i.e.* highly represented and differentially expressed in *Oophaga* poison frogs) allowed us to propose plausible metabolic pathways (and candidate genes) related with toxicity and color. Overall, this study provides an important molecular resource for the study of aposematism within poison frogs and will facilitate the future assembling and annotation of the complex dendrobatid genomes [23].

**METHODS**

**Library preparation and sequencing**

Due the conservation status of these frogs (CR, EN, VU; IUCN) and the restrictions imposed by the Ethics Committee, we were allowed to euthanize a very limited number of samples per species. Seven individuals (*O. lehmanni* and *O. anchicayensis*, n=2 per each species; *O. histrionica*, *O. sylvatica*, and *O. solanensis*, n=1; each species) were euthanized in the field with benzocaine gel at 5% by direct application into the mouth cavity. Individual skin samples were taken and stored at -80 °C in RNAlater® (Life Technologies, Carlbad, CA). RNA extractions were performed using Tri-Reagent (Molecular Research Center, Cincinnati, OH). After quality check (Agilent Bioanalyzer Agilent Technologies, Wilmington, DE) samples were used to generate RNA-seq libraries using the Illumina Truseq RNA Sample Prep protocol (Illumina, San Diego, CA).
Diego, CA). Libraries were cleaned using AMPure XP and sequenced on a single Illumina HiSeq2000 lane (TruSeq SBS v. 3) as follows: four libraries (O. anchicayensis, O. histrionica, O. lehmanni and O. sylvatica) were sequenced using single-end (150-bp) read trimming, filtering and removal of adapters was performed using FLEXBAR v.2.5 [24]. All retained reads were ≥ 50bp with an average quality of ≥ 30 and less than two uncalled (ambiguous) bases (S1 Table). For each library, a de novo transcriptome was assembled using TRINITY v.2.1.0 [25] with default settings (kmer=25, minimum contig length=48) and keeping only the longest transcript per cluster for subsequent analyses. Varying these parameters (i.e. default settings) did not result in assemblies with longer N50 values. To account for differences in the quality and sequencing strategy among individual transcriptomes, we constructed a composite de novo reference transcriptome. Briefly, we followed [26] and combined 20% randomly selected reads (>33 million in total) from each of the four single-end libraries (O.
anchicayensis, O. histrionica, O. lehmanni and O. sylvatica). Then, a TRINITY assembly was performed using a minimum contig length of 350 bp. To filter out highly similar contigs that may potentially represent alternatively spliced transcripts, we implemented the error correction module of iAssembler v1.3.2 [27] with default parameters (maximum length of end clips=30 bp, minimum overlap length=40 bp, minimum percent sequence identity=95%).

Gene annotation was conducted using a sequential BLASTX search to both, the available Xenopus transcriptomes and the NCBI non-redundant (nr) database. The composite transcripts were first compared with that of the Xenopus databases retaining annotations with E-values \( \leq 10^{-5} \). Unannotated contigs were then submitted for BLASTX to the nr protein database for possible identification. Gene ontology (GO) annotation and term mapping was performed using Blast2Go v 5.2.0 with default significance cutoffs [28].

To estimate the completeness of each transcriptome, we implemented the expected gene content of Benchmarking Universal Single-Copy Orthologs (BUSCO v3) [29] as implemented in gVolante v1.1.0 [30]. This widely used metrics for transcriptome assembly [31] assesses the quality of any given transcriptome by estimating the completeness of core vertebrate genes predicted to be ubiquitous in eukaryotes. The preprocessing of the reads and all BLASTX analyses were run in the Bugaboo Dell Xeon cluster of the western Canada's WestGrid computer facilities (www.westgrid.ca).

**Transcriptome profiles**

Analyses on the distribution of reads and the relative abundance of different transcripts within each individual sample (i.e. lineage) were carried out by mapping
RNA-seq reads back to the composite *de novo* assembly. In order to identify highly represented unigenes, we first implemented eXpress (probabilistic assignment of ambiguously mapping sequenced fragments) [32] to estimate the effective number of reads (ER) that mapped to the contigs in the reference transcriptome after adjusting for read number and length biases. Then, to compare the proportion of reads that mapped to a transcript in each *Oophaga* RNA-library, we estimated the number of transcripts per million (TPM), a measure of RNA abundance that allows the comparison between samples (sum of all TPMs in each sample are the same) [33]. We visually explored for highly represented unigenes in each RNA-library by using dispersion plots of the ER values and percentile plots of the TPM distribution. Additionally, we performed a Principal Component Analysis (PCA) of the reference contigs dataset and using each library TPMs values as independent variables and selected the overall top expressed unigenes (2%).

To further characterize the skin transcriptome profile of *Oophaga* poison frogs, we compared the transcription levels of annotated unigenes in this clade with those observed in several cryptic (*i.e.* non-aposematic) anurans, including a toad (*Bufo gargarizans*), and five frog species (*Pelophylax nigromaculatus*, *Polypedates megacephalus*, *Rana catesbeiana*, *Rana sylvatica*, and *Xenophrys sangzhiensis*, S2 Table) [34, 35]. RNA-seq data from these species was obtained from the Sequence Read Archive platform (SRA) and represented the only skin-based RNA-seq available to us for comparison purposes. Although transcriptomic datasets are available for phylogenetically closer and cryptic species (*i.e* *Colostethus*) [36], these datasets were not included in our analysis because they are not derived from skin tissue and may bias our
analyses. Read quality, contamination screening, duplicate removal, and quality trimming steps were performed in FLEXBAR v.2.5 as previously described. The filtered raw reads from individual libraries (Oophaga, n=7; cryptic species n=6) were mapped to our composite de novo reference transcriptome using Bowtie2 v2.3.4.1 [37]. To accommodate for sequence divergence among the taxa included in our study (≈75–98 similarity), we followed [38] and ran the alignment algorithm allowing for soft clipping (--local) and with a maximum penalty value of three (--mp 3). Then, we implemented eXpress [32] to estimate the effective number of reads (ER) that mapped to the contigs in the reference transcriptome after adjusting for read number and length biases. Only unigenes with an average of >10 ER/library were included in the differential expression analyses [39].

To identify genes showing a phylogenetic signal in their expression levels and to estimate fold changes in expression levels between aposematic (i.e. Oophaga) and cryptic species, we followed [40] and implemented the R package DEseq [41] to select unigenes with an adjusted p-value of <0.05. To visualize patterns of expression, we constructed a multidimensional scaling plot (Euclidean distances) of the ER data using Past v. 3.18 [42]. A heatmap of the expression differences between Oophaga and cryptic anuran lineages (n<sub>genes</sub>=1,931) was generated using the ‘heatmap.2’ function of the ‘gplots’ package, with Euclidean distances and complete linkage for clustering (bootstrap=999) [43]. BLAST analysis and GO term mapping for the 1,931 unigenes included in this analysis was performed as described above.

**RESULTS**

**Transcriptome assemblies and functional annotation**
Illumina sequencing produced an average number of reads per sample of 118.3 million for paired-end and 37.7 million for single-end libraries. After a stringent read trimming involving the removal of low-quality sequences, duplicated reads and reads containing adapter/primer sequences, we retained over 81% of the initial sequencing data. Paired-end libraries produced a fairly consistent number of reads (S1 Table). After the removal/merging of highly similar contigs (paired-end: 7.5% - 9.6%; single-end: 6.0% - 13.4%), the 150-bp based transcriptomes recovered a large number of contigs ranging from 35,287 (O. anchicayensis) to 107,381 (O. sylvatica). Paired-reads libraries (100-bp; 2X) generated transcriptomes with a relatively similar number of contigs (40,398 for O. solanensis and 60,494 for O. lehmannii). N50 values and average transcript length (AL) were lower for paired-end libraries (N50= 498-561bp; AL= 441.78±379 - 475.93±436bp) than for assemblies produced with single-end libraries (N50=667-1579bp; AL=538.9±1001bp - 668.6±819 bp) (Table 1). Assembled and annotated transcriptomes are available from the authors upon request.
Table 1. Summary statistics and quality assessment estimators of the eight *de novo* assembled transcriptomes generated in this study (*Oophaga* species n=7; composite reference n=1).

<table>
<thead>
<tr>
<th></th>
<th>Paired-end libraries</th>
<th>Single-end libraries</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>O. solanensis</em></td>
<td><em>O. anchicayensis</em></td>
<td><em>O. lehmanni</em></td>
</tr>
<tr>
<td>File size (Mbytes)</td>
<td>22.23</td>
<td>32.06</td>
<td>35.7</td>
</tr>
<tr>
<td># contigs</td>
<td>43,690</td>
<td>63,520</td>
<td>66,937</td>
</tr>
<tr>
<td># contigs after filtration (removal/merging of highly similar contigs)</td>
<td>40,398</td>
<td>57,516</td>
<td>60,494</td>
</tr>
<tr>
<td># bases</td>
<td>19,853,219</td>
<td>28,061,593</td>
<td>31,856,993</td>
</tr>
<tr>
<td># bases after filtration</td>
<td>17,164,733</td>
<td>23,756,585</td>
<td>26,287,312</td>
</tr>
<tr>
<td>Merged contigs</td>
<td>7.5%</td>
<td>9.5%</td>
<td>9.6%</td>
</tr>
<tr>
<td>Average transcript length (AL)</td>
<td>454.41± 405</td>
<td>441.78± 379</td>
<td>475.93 ± 436</td>
</tr>
<tr>
<td>Maximum length</td>
<td>6,115</td>
<td>7,915</td>
<td>8,400</td>
</tr>
<tr>
<td>N50</td>
<td>528</td>
<td>498</td>
<td>561</td>
</tr>
<tr>
<td>Contigs with BLAST hits</td>
<td>46.5%</td>
<td>47.7%</td>
<td>50.8%</td>
</tr>
</tbody>
</table>
Comparative BLAST analysis indicated that our transcriptomes recovered a significant proportion of the *X. laevis* and *X. tropicalis* transcriptomes (50,592 and 41,042 unigenes respectively). Paired-end transcriptomes showed a lower number of significant BLAST hits (18,263 – 29,670) than single-end libraries (18,526 – 40,651) (S1 Fig). In general, individual *de-novo* transcriptomes were of good quality and recovered 51-79% of the core eukaryotic universal single-copy orthologs [29] (S3 Table). The final *de novo* composite skin transcriptome reconstructed from all the RNA reads of the *Oophaga* species yielded 31,498 contigs greater than 350 bp. The overall assembly incorporated 94% of all initial reads and the level of fragmentation was low with half of all base pairs clustered into contigs of 1,316 bp in length or greater. The maximum contig length was 16,041 bp and the AL was 1014.91 ± 948.43 (S2A Fig). Nucleotide-based BLAST analyses (*BLASTX*) revealed that ~63% of the contigs (n=19,732) show significant similarities with either annotated gene products and/or known protein domains (E-value ≤ 10−5) (S2B Fig) and only a small fraction of unigenes (3.4%) showed significant homology to the same annotated transcript. The highest BLAST scores were found with *X. tropicalis* (74%), *X. laevis* (2%) and the green sea turtle *Chelonia mydas* (0.5%). The composite skin transcriptome was of higher quality when compared with individual assemblies, having 88.54% of the ortholog count of BUSCO [29] (S3 Table).

Gene Ontology (GO) assignments were used to classify the functions of the predicted unigenes based on contigs with significant *BLASTX* (E-value ≤ 10−5). Based on GO level II, unigenes were assigned to 27 biological processes (BP), 22 cell components (CC) and 15 molecular functions (MF) (S2C Fig). Some unigenes were associated with multiple GO annotations because a single sequence may be annotated in any or all
categories, giving more GO annotations than sequences annotated [44]. Within BP, ~47% of the annotations were assigned to basic cellular, metabolic processes and biological regulation. The remaining unigenes were involved in a broad range of BP such as response to stimulus (7%), response to stress (7%), localization (6%), developmental process (5%), signal transduction (4%), biogenesis (4%), immune response (2%), reproductive process (1.7%) and cellular adhesion (1%). Within the CC category, other than the essential cell constituents (44%), the membrane components were highly represented in the transcriptome (19%). Within MF, most of the unigenes were assigned to binding and catalytic activities (72%) followed by transporter activity (9%), molecular transducers (7%) and molecular function regulators (4%) (S2C Fig).

Transcriptome profiles

Dispersion plots of ER values in Oophaga RNA libraries showed a common pattern of over-represented unigenes across RNA-libraries. That is, the same group of unigenes showed similar patterns of over-representation no matter RNA library’s origin (i.e. different Oophaga species) or composition (paired vs. single-read libraries) (S3 Fig). The TPM represents a measure of RNA abundance and hence, it provides a general overview of gene expression levels in a particular sample. Interestingly, TPM distribution plots (percentiles) showed a distinct spike in expression levels towards percentile 98% for all our RNA-libraries (S4A Fig). After the conservative filtration of the 2% top expressed unigenes, we selected a subset of 1,437 unigenes with particular higher levels of expression (higher TPM values and PCA outliers, S4B Fig) in Oophaga skin tissue. Homology BLAST analysis of these unigenes revealed that 76% (n=1,092) show significant similarities with annotated gene products and/or known protein domains.
distributed mainly across amphibian (frogs) species ($X. tropicalis$=30%, $X. laevis$=14% and $Rana catesbeiana$=4%).

After adjusting for library size and length bias, the maximum effective number of reads (ER) in $Oophaga$ species ranged from 2,461 ($O. anchicayensis$ single-end) to 5,813 ($O. sylvatica$ – single-end) while in the cryptic species varied from 82,509 ($Rana catesbeiana$) to 452,018 ($Bufo gargarizans$) (higher coverage in cryptic species, S2 Table). This indicated that after the optimization of our alignment parameters, a considerable amount of reads from each RNA-seq library actually mapped to our composite de novo transcriptome. After the removal of contigs with low expression profiles (<10 read counts), our RNA-seq dataset was composed of reads mapping to 20,771 genes. Multidimensional scaling plots (stress<0.05) revealed a clear separation by groups ($Oophaga$ vs. cryptic species, p<0.001 Bonferroni corrected) with higher variation among the cryptic species, as expected given the broad taxonomic range of the members of this group (Fig 2).

About 9% of the highly expressed genes (n= 1,931) also showed a phylogenetic signal (i.e. significant fold changes in expression levels between $Oophaga$ and all other anurans). Most of these genes (n=1,656; 85.8%) show higher levels of expression in $Oophaga$. The clustering analysis (Fig 1) revealed that the toad ($B. gargarizans$) has the closest transcription profile to $Oophaga$ species. One possible explanation is the presence of alkaloid-based toxins in its skin glands [45].
Fig 2. Multidimensional scaling (MDS) plot of 13 RNA-sequencing samples from skin tissue of *Oophaga* (n=7, red dots) and cryptic species (n=6, blue dots). RNA-seq data shown here represent the effective read counts (ER) for 31,498 skin-specific contigs. Color ellipses represent a 95% confidence interval.
Homology BLAST analyses revealed that 65% (n=1,252) show significant similarities with annotated gene products and/or known protein domains (S4 Table; S5A Fig) distributed mainly across amphibian (frogs) species (X. tropicalis=32%, X. laevis=19% and Rana catesbeiana=4%) (S5B Fig). Based on GO mapping level II, these unigenes were assigned to 24 BP, 20 CC and 13 MF (S5C Fig). Multilevel GO term classification assigned the highly expressed unigenes to 16 BP, 5 CC and 9 MF (S5C Fig). Within BP, 13% of unigenes were associated with oxidation-reduction processes, followed by response to stimulus (10%) and regulation of macromolecule metabolic processes (7.3%). Within CC, the integral components of the cell membrane were highly represented (39%), followed by protein complex (27%) and nuclear components (20%). Within MF, most of the differentially expressed unigenes were assigned to metal ion binding activity (21%), followed by protein binding activity and transmembrane transporters (17%), oxidoreductase activity (12%) and structural molecule activity (10%) (S5C Fig). The full functional annotation of these genes allowed us to propose plausible mechanisms that may be involved in the ability to sequester alkaloid-based chemical defenses, warning coloration and the auto-resistance mechanisms to avoid self-intoxication (S5 Table).

**DISCUSSION**

Studies focusing on highly represented and differential gene expression in related lineages have provided valuable insights into the molecular underpinnings of phenotypic variation [46, 47]. These analyses rely on the premises that lineages show contrasting phenotypes of interest, and that most differentially expressed genes relate to the trait of
interest rather than the overall of genomic differentiation between lineages. Here, we examined shared patterns of gene expression in a small clade of closely related aposematic frogs (*Oophaga* species). Because all members of this monophyletic group show contrasting hues and accumulate alkaloid-derived toxins in their dermal glands [48, 49], we assumed that the genes related with these traits are likely to 1) be highly expressed in the skin of *Oophaga* poison frogs and 2) showed contrasting expression patterns between aposematic and cryptic and less (or not) poisonous species. Using expression levels in a multivariate framework, we first identified ortholog genes that are highly represented in *Oophaga* skin tissue and then, we compared with those differentially expressed in the skin of contrasting groups and that show similar patterns of expression among *Oophaga* species. We finally used a functional annotation analyses to highlight a series of genes that are possibly associated with a series of aposematic traits.

The characterization of our skin transcriptomic profiles allowed us to identify genes that are potentially associated with alkaloid transportation and resistance to autotoxicity in poison frogs (Dendrobatidae). Because the alkaloids secreted by these frogs are structurally very similar to those present in plants [18, 50-52], it is reasonable to assume that the transportation and accumulation of alkaloids in these frogs may be carried out by similar systems to those described in plants in which alkaloids are transferred from source to sink organs [53-55]. If so, at least, two alternative membrane mechanisms may help to explain the transport and storage of unmodified diet-derived alkaloid by the specialized cells of skin secretory glands. First, alkaloids in a lipophilic state may freely pass through the cell membranes by simple diffusion and accumulate in acidic secretory lysosomes if they become protonated to form hydrophilic cations. This
ion-trap mechanism is not energy-dependent and does not necessarily require the expression of any transporters. Alternatively, the transportation of alkaloids may be managed by proton-antiport carrier systems in an energy-requiring manner [56-58]. Under this alternative model, diet derived alkaloids may be taken up by ABC transporters [59, 60] and may accumulate in the secretory lysosomes of the skin gland cells by a cation exchanger antiporter (CAX) system, dependent on the pH gradient generated by “vacuolar” type ATPases (V-H⁺-ATPases) and/or pyrophosphatases (V-H⁺-PPases). To find some support for these two potential models that might explain the accumulation and posterior secretion of dietary alkaloids, we investigated the common transcriptional profile across Oophaga species. For each of these species, our differential expression and transcript abundance analysis revealed that at least 15 homologs of different type II CAX are highly expressed in skin tissue (S5 Table). A result suggesting the existence of an active membrane transportation-accumulation mechanism of alkaloids in harlequin poison frogs. Consistent with this hypothesis, we also found putative ABC transporters and at least one V-H⁺-ATPase on the skin transcriptome. All together, these results suggest that active lysosomal exocytosis may play a key role in the secretion of alkaloids in these frogs.

The accumulation of toxic compounds implies that organisms must avoid self-intoxication (auto-resistance). While the membrane transportation mechanisms described above reduce auto-toxicity by compartmentalizing the sequestered alkaloids, other non-alternative excluding mechanisms are likely to contribute to auto-toxicity resistance. In harlequin poison frogs, the major toxic components present in skin tissue are histrionicoxins (HTX) [51]. These alkaloids are known to cause temporary paralysis or
even death by inactivating or blocking voltage-gate ion channels [61]. Resistance to this type of cytotoxic compounds usually arises through either, an increased expression level of P450 enzymes (CYPs) that metabolize the toxin, or through target insensitivity via mutations that reduce the toxin's ability to bind the ion channel itself [62]. Our analyses revealed that five CYP homologs are amongst the most highly expressed genes in the skin of harlequin poison frogs (S5 Table). This result suggests that the oxidative biotransformation of lipophilic alkaloids to hydrophilic compounds [63] is an important auto resistance mechanism in these frogs. Although the overproduction of the CYP-enzymes has been interpreted as evidence for metabolic detoxification [64], there is an alternative explanation for the high expression levels of CYP proteins in skin of harlequin poison frogs. While many of the alkaloids sequestered in the wild are accumulated unchanged in the dermal glands, at least two species of dendrobatids, stereo-selectively hydroxylate the toxic pumiliotoxin PTX-(+)-251D to convert it into a much more potent toxin, the allopumiliotoxin aPTX-(+)-267A [17]. Both PTX-(+)-251D and aPTX-(+)-267A have been isolated from the skin of O. histrionica, and CYPs are known to be involved in the stereo-selective hydroxylation of alkaloids [65]. Thus, it is also possible that the CYPs expressed in the skin of harlequin poison frogs may enhance the antipredator potency of ingested PTXs.

Despite the major role that detoxification enzymes may have in the resistance to diet acquired toxins, mutations conferring constitutive resistance to alkaloids are also likely to be involved in auto-resistance in chemically-defended dendrobatids. In insects, single amino-acid substitution in the voltage-gated sodium channels (Na⁺K⁺-ATPases) are responsible for resistance to host-plant phytochemicals [66-68]. In amphibians, the
only documented example of this this type of target insensitivity is from distantly-related lineages of alkaloid-defended frogs, 24 species of Neotropical dendrobatids and the aposematic Madagascar frog *Mantella aurantiaca* [69]. In all of these cases, they carried different amino acid replacements in the inner pore of the voltage-gated sodium channel Nav1.4 that are not found in other frog [70]. Our analyses revealed eight genes encoding voltage-gated ion channel proteins (VGIC), one of which (8966), encodes the gamma-1 subunit of a sodium channel.

Aposematism in harlequin frogs is a complex phenotype that results from the integration of different elements including unpalatability and conspicuousness. Thus, another of our goals was to identify genes potentially related with warning coloration. Many aposematic organisms, including the *Oophaga* poison frogs studied here exhibit color patterns that show strong color and/or luminance contrast such between black and red, orange or yellow [15]. It has been proposed that such patterning increases aversion learning as well as conspicuousness and distinctiveness from palatable prey [71-76]. In anurans, coloration relates to both the structure and pigment composition of the dermal chromatophore units in which three types of pigment cells (xanthophores, iridophores and melanophores) are laid one on another [77]. In dark/black skin areas of harlequin poison frogs, no xanthophores or iridophores are present and the distal fingers of the melanophores are filled with melanin granules (melanosomes) obscuring the dermis [15]. Thus, genes involved in the amount, size and distribution of the melanosomes are likely to play a significant role in the coloration pattern of these frogs.

In one of the few detailed studies of coloration in anurans [78], authors suggest a common origin for of all the pigment granules found in the cells of the chromatophore: a
primordial organelle (vesicle) derived from the rough endoplasmic reticulum (RER).

According to this model, in the formation of melanosomes, the pre-melanosomes are derived from cisternae of the RER, which then fuse with vesicles derived from the Golgi complex containing tyrosinase enzymes [12, 78-81]. A highly expressed gene identified here (tyrosinase regulator, 25062) may contribute to regulate this mechanism, which in turn might translate into hue differences. Interestingly, among Oophaga species, we have found populations that are characterized by light-brown background coloration as oppose to black [15, 22]. One tantalizing possibility is that this phenotypic difference is indeed associated with the expression differences of these tyrosinase enzymes.

Obviously, there are other molecular and cellular mechanisms that might be associated to the difference between light and dark background coloration. In fishes and amphibians, dark hues are known to be produced by the interaction between high levels of melanocyte-stimulating hormone (α-MSH) and several variants of its transmembrane receptor (MC1R) through the dispersion of melanosomes within the melanophore (by increasing cAMP intracellular levels) [82, 83]. Thus, it is possible that structural or expression differences in the MC1R might contribute to dark phenotypes in Oophaga frogs. While there is a strong evidence that different mutations at MC1R cause either light or dark phenotypes in many mammals, birds and reptiles [84-88], the only two studies conducted in frogs are inconclusive [89, 90]. A detailed inspection of the coding sequences recovered for this gene revealed the presence of different length isoforms, making MC1R a promising candidate gene candidate to explain the differences in background coloration in poison frogs [15]. Finally, our study revealed another two highly and differentially expressed genes (14447-Melanoregulin-; 2038-Melanocortin...
phosphoprotein), which may also contribute to dark hues. In this case, the predicted products of these genes are key proteins that mediate the melanosome transport and distribution in epidermal cells through the formation of a tripartite protein complex [91]. The disruption of the transport protein complex results in pigmentedary dilution and lighter phenotypes by the clustering of melanosomes around the nucleus [92].

Overall, this study demonstrates the utility of using RNA-sequencing with non-model organisms to identify loci that may be of adaptive importance. Altogether, these data enabled us to provide a first global study of *Oophaga* poison frogs transcriptomes and propose potential mechanisms for alkaloid sequestration, auto-resistance to toxic compounds, and variation in coloration and patterns. It is important to recognized that due the sampling restrictions imposed to us (see methods), we are most likely missing some of the genes responsible for these traits. It is also probable that the genes associated with coloration, alkaloid metabolism, transport and storage are differentially expressed not only in skin, but other tissues in the organism. Despite the potential caveats, other molecular studies have shown that some of the genes identified in this study are indeed potentially related with aposematic traits in dendrobatids [15, 70, 93].

Despite the critical impact of genetic basis of coloration in the evolution and diversification of aposematic phenotypes, the genetic architecture of coloration in poison frogs (Dendrobatidae) remains virtually unexplored. Here, we report probably the first RNA-seq study in *Oophaga* poison frogs, a model system for understanding the relationship between toxicity, diet and coloration. The skin-expressed genes that we have identified here provide initial working hypotheses to further unravel the molecular genetics mechanisms to sequester alkaloid-based chemical defenses, warning coloration
and the auto-resistance mechanisms to avoid self-intoxication. Hence, further analysis aiming to compare the amino acid substitutions in the expressed cDNAs will provide insights about the structure and function of each of these genes. Finally, our comparative transcriptome data provide an important new resource to better understand the evolution of warning signals in nature.

Acknowledgements

We would like to thank the former members of Rick Harrison’s lab (*requiescat in pace*) at Cornell University for their valuable discussions and comments on early versions of the manuscript, which was greatly improved by the thoughtful comments of Dr. Andrew Crawford at Universidad de los Andes (Colombia). This study was funded by COLCIENCIAS (Departamento Administrativo de Ciencia, Tecnología e Innovación, Colombia, 529-2011) grant to A.P.-T. and an NSERC-Discovery grant to J.A.A.
REFERENCES


17. Daly JW, Garraffo HM, Spande TF, Clark VC, Ma J, Ziffer H, et al. Evidence for an enantioselective pumiliotoxin 7-hydroxylase in dendrobatid poison frogs of the genus


http://www.nature.com/nmeth/journal/v10/n1/abs/nmeth.2251.html#supplementary-information.


https://www.nature.com/articles/srep24069#supplementary-information.


https://www.nature.com/articles/nmeth.1923#supplementary-information.


61. Daly JW, McNeal ET, Overman LE, Ellison DH. A new class of cardiotonic agents: structure-activity correlations for natural and synthetic analogs of the alkaloid


**SUPPORTING INFORMATION**

**SUPPORTING TABLES**

S1 Table. Number of reads and library type (i.e paired vs. single-end) for each individual RNA-seq experiment.

<table>
<thead>
<tr>
<th>Individual library</th>
<th>Raw number of reads</th>
<th>Reads after filtering</th>
<th>Remaining % of reads</th>
<th>Library type</th>
</tr>
</thead>
<tbody>
<tr>
<td>O. solanensis</td>
<td>128,599,724</td>
<td>105,179,065</td>
<td>81%</td>
<td>Paired reads*</td>
</tr>
<tr>
<td>O. anchicayensis</td>
<td>124,232,804</td>
<td>110,217,128</td>
<td>88%</td>
<td></td>
</tr>
<tr>
<td>O. lehmanni</td>
<td>102,225,680</td>
<td>91,493,848</td>
<td>89%</td>
<td></td>
</tr>
<tr>
<td>O. sylvatica</td>
<td>48,929,329</td>
<td>45,147,654</td>
<td>92%</td>
<td></td>
</tr>
<tr>
<td>O. lehmanni</td>
<td>43,633,080</td>
<td>41,086,379</td>
<td>94%</td>
<td>Single reads</td>
</tr>
<tr>
<td>O. anchicayensis</td>
<td>7,998,636</td>
<td>7,173,285</td>
<td>89%</td>
<td></td>
</tr>
<tr>
<td>O. histrionica</td>
<td>50,204,617</td>
<td>47,376,912</td>
<td>94%</td>
<td></td>
</tr>
<tr>
<td>Reference transcriptome</td>
<td>N/A</td>
<td>33,895,474</td>
<td>100%</td>
<td>Single reads</td>
</tr>
</tbody>
</table>

*. Number of reads correspond to the combined value of both ends libraries.
**S2 Table.** RNA-sequence datasets from cryptic species used for the differential expression analysis of this work.

<table>
<thead>
<tr>
<th>Species</th>
<th>Experiment (SRA*)</th>
<th>RNA-seq library</th>
<th>Total bases</th>
<th>Date available</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Bufo gargarizans</em></td>
<td>SRX2640691</td>
<td>Paired-end (150 b)</td>
<td>19.2 Gbp</td>
<td>2017-03-15</td>
</tr>
<tr>
<td><em>Pelophylax nigromaculatus</em></td>
<td>SRX2640690</td>
<td>Paired-end (150 b)</td>
<td>18.0 Gbp</td>
<td>2017-03-15</td>
</tr>
<tr>
<td><em>Polypedates megacephalus</em></td>
<td>SRX1720193</td>
<td>Paired-end (100 b)</td>
<td>6.3 Gbp</td>
<td>2016-04-25</td>
</tr>
<tr>
<td><em>Rana catesbeiana</em></td>
<td>SRX2988982</td>
<td>Paired-end (100 b)</td>
<td>9.2 Gbp</td>
<td>2017-07-07</td>
</tr>
<tr>
<td><em>Rana sylvatica</em></td>
<td>SRX2989023</td>
<td>Paired-end (100 b)</td>
<td>8.9 Gbp</td>
<td>2017-07-07</td>
</tr>
<tr>
<td><em>Xenophrys sangzhiensis</em></td>
<td>SRX1720191</td>
<td>Paired-end (100 b)</td>
<td>6.8 Gbp</td>
<td>2016-04-25</td>
</tr>
</tbody>
</table>

* Sequence Read Archive accession number from Genbank.

**S3 Table.** Quality assessment report for each individual and the composite reference transcriptome based on the protein count of the core eukaryotic genes (BUSCO, see methods in the main article)

<table>
<thead>
<tr>
<th>Transcriptome</th>
<th>Library</th>
<th>Busco v.2 mapping</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Complete</td>
<td>Partial</td>
<td>Complete + Partial</td>
</tr>
<tr>
<td><em>O. solanensis</em></td>
<td>Paired</td>
<td>20.60%</td>
<td>30.47%</td>
<td>51.07%</td>
</tr>
<tr>
<td><em>O. anchicayensis</em></td>
<td>Paired</td>
<td>24.46%</td>
<td>37.77%</td>
<td>62.23%</td>
</tr>
<tr>
<td><em>O. lehmanni</em></td>
<td>Paired</td>
<td>29.61%</td>
<td>40.77%</td>
<td>70.38%</td>
</tr>
<tr>
<td><em>O. sylvatica</em></td>
<td>Single</td>
<td>45.92%</td>
<td>33.48%</td>
<td>79.40%</td>
</tr>
<tr>
<td><em>O. lehmanni</em></td>
<td>Single</td>
<td>33.91%</td>
<td>39.91%</td>
<td>73.82%</td>
</tr>
<tr>
<td><em>O. anchicayensis</em></td>
<td>Single</td>
<td>9.87%</td>
<td>30.47%</td>
<td>40.34%</td>
</tr>
<tr>
<td><em>O. histrionica</em></td>
<td>Single</td>
<td>39.06%</td>
<td>37.77%</td>
<td>76.83%</td>
</tr>
</tbody>
</table>

**S4 Table.** Summary statistics of the differentially expressed unigenes (see methods) and their BLAST annotations.

<table>
<thead>
<tr>
<th></th>
<th>Differentially expressed unigenes</th>
<th>With annotation</th>
<th>Without annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up-regulated</td>
<td>1656 (85.6%)</td>
<td>996 (51.6%)</td>
<td>660 (34.2%)</td>
</tr>
<tr>
<td>Down-regulated</td>
<td>275 (14.2%)</td>
<td>256 (13.3%)</td>
<td>19 (1%)</td>
</tr>
<tr>
<td>Total</td>
<td>1931 (100%)</td>
<td>1252 (64.9%)</td>
<td>679 (35.2%)</td>
</tr>
</tbody>
</table>
**S5 Table.** Transcripts with potential important functions in the alkaloid sequestration system, auto-resistance to toxic compounds and variation in coloration in harlequin poison frogs. The corresponding composite *Oophaga* transcriptome is available from the authors upon request.

<table>
<thead>
<tr>
<th>Transcript ID</th>
<th>Fold change</th>
<th>padj</th>
<th>BLAST description</th>
<th>Length (bp)</th>
<th>GO Names list</th>
</tr>
</thead>
<tbody>
<tr>
<td>14256</td>
<td>779.11</td>
<td>0.00</td>
<td>Transmembrane protein 41B</td>
<td>1253</td>
<td>C:cell; P:gonad development; P:thyroid gland development; P:luteinizing hormone secretion; P:thyroid hormone generation; P:cellular response to hormone stimulus; P:negative regulation of organ growth; C:extracellular region; F:hormone activity</td>
</tr>
<tr>
<td>26699</td>
<td>517.76</td>
<td>0.00</td>
<td>Transmembrane protein 181</td>
<td>1079</td>
<td>C:extracellular space</td>
</tr>
<tr>
<td>14644</td>
<td>297.83</td>
<td>0.00</td>
<td>Transmembrane protein 116-like</td>
<td>1743</td>
<td></td>
</tr>
<tr>
<td>14399</td>
<td>192.77</td>
<td>0.01</td>
<td>Transmembrane protein 26</td>
<td>1039</td>
<td></td>
</tr>
<tr>
<td>6476</td>
<td>174.07</td>
<td>0.00</td>
<td>Transmembrane protein</td>
<td>1690</td>
<td>F:structural constituent of eye lens</td>
</tr>
<tr>
<td>15428</td>
<td>105.09</td>
<td>0.02</td>
<td>Transmembrane protein 220</td>
<td>697</td>
<td>F:chondroitin sulfate binding; C:nucleus; F:identical protein binding; F:calcium-dependent protein binding; C:cytosol; C:chromaffin granule membrane; P:regulation of transcription, DNA-templated; F:heparin binding; P:negative regulation of cellular process; F:carbohydrate binding; C:apical plasma membrane</td>
</tr>
<tr>
<td>8646</td>
<td>100.29</td>
<td>0.01</td>
<td>Transmembrane protein 69</td>
<td>936</td>
<td></td>
</tr>
<tr>
<td>20904</td>
<td>66.89</td>
<td>0.04</td>
<td>Transmembrane protein 186-like</td>
<td>615</td>
<td></td>
</tr>
<tr>
<td>Rank</td>
<td>Z-score</td>
<td>FDR</td>
<td>Description</td>
<td>Biological Process</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>19857</td>
<td>53.30</td>
<td>0.05</td>
<td>Transmembrane protein 70, mitochondrial</td>
<td>P: 'de novo' pyrimidine nucleobase biosynthetic process; F: orotidine-5'-phosphate decarboxylase activity; P: 'de novo' UMP biosynthetic process</td>
<td></td>
</tr>
<tr>
<td>24171</td>
<td>45.71</td>
<td>0.01</td>
<td>Transmembrane protein 125-like</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19808</td>
<td>37.86</td>
<td>0.04</td>
<td>Keratinocyte-associated Transmembrane protein 2-like</td>
<td>P: negative regulation of potassium ion transmembrane transport; P: positive regulation of myofibroblast contraction; P: response to salt; P: negative regulation of platelet aggregation; P: response to progesterone; P: positive regulation of vasoconstriction; P: negative regulation of gastrin-induced gastric acid secretion; C: neuronal cell body; P: negative regulation of insulin secretion; F: prostaglandin ....</td>
<td></td>
</tr>
<tr>
<td>16724</td>
<td>27.39</td>
<td>0.02</td>
<td>Transmembrane protein 206</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24522</td>
<td>25.75</td>
<td>0.00</td>
<td>Transmembrane protease serine 13-like</td>
<td>F: oxidoreductase activity</td>
<td></td>
</tr>
<tr>
<td>25621</td>
<td>7.42</td>
<td>0.05</td>
<td>Cation-transporting ATPase 13A4</td>
<td>P: response to dehydroepiandrosterone; F: gamma-catenin binding; C: extracellular space; P: response to 11-deoxycorticosterone; C: membrane; P: metabolic process; P: response to progesterone; P: response to estradiol; C: desmosome; F: hydrolase activity</td>
<td></td>
</tr>
<tr>
<td>9463</td>
<td>-6.16</td>
<td>0.04</td>
<td>Lysosomal-trafficking regulator</td>
<td>F: calcium ion binding; P: keratinization</td>
<td></td>
</tr>
<tr>
<td>Transcript ID</td>
<td>Fold change</td>
<td>padj</td>
<td>BLAST description</td>
<td>Length (bp)</td>
<td>GO Names list</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>15350</td>
<td>175.29</td>
<td>0.00</td>
<td>EF-hand calcium-binding domain-containing 10-like</td>
<td>436</td>
<td>F: calcium ion binding; C: cellular component</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P: single-organism cellular localization; P: cell cycle process; F: binding; P: organelle organization; C: intracellular organelle part; P: single-organism transport; P: establishment of protein localization; C: endosome; C: membrane-bounded vesicle; P: vesicle-mediated transport; P: regulation of cellular process; C: membrane; C: cytoplasmic vesicle; P: establishment of localization in cell; P: cellular protein localization; P: establishment of organelle localization</td>
</tr>
<tr>
<td>16951</td>
<td>95.46</td>
<td>0.01</td>
<td>S100 calcium binding protein Z</td>
<td>2161</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C: extracellular space; F: carboxypeptidase activity; P: bradykinin catabolic process; P: proteolysis; P: response to glucocorticoid</td>
</tr>
<tr>
<td>1216</td>
<td>63.08</td>
<td>0.03</td>
<td>Extracellular calcium-sensing receptor-like</td>
<td>1214</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P: ion transport; P: regulation of membrane potential; C: integral component of membrane; F: extracellular ligand-gated ion channel activity; P: synaptic transmission; C: synapse; F: binding; P: regulation of biological process; P: neurological system process; P: response to stimulus; C: plasma membrane part</td>
</tr>
<tr>
<td>18107</td>
<td>4239.03</td>
<td>0.00</td>
<td>Potassium Voltage-gated channel subfamily E member 3-like</td>
<td>1864</td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Description</td>
<td>Value 3</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>---------</td>
<td>----------------------------------------------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>16614</td>
<td>9.90</td>
<td>0.04</td>
<td>MGC114646 protein</td>
<td>1514</td>
<td></td>
</tr>
<tr>
<td>907</td>
<td>-6.46</td>
<td>0.01</td>
<td>Voltage-dependent anion channel 2</td>
<td>990</td>
<td></td>
</tr>
<tr>
<td>376</td>
<td>-10.41</td>
<td>0.00</td>
<td>Voltage-dependent anion channel</td>
<td>388</td>
<td></td>
</tr>
<tr>
<td>8966</td>
<td>102.76</td>
<td>0.01</td>
<td>Sodium channel, voltage-dependent, gamma subunit 1</td>
<td>640</td>
<td></td>
</tr>
</tbody>
</table>

- **F:** high voltage-gated calcium channel activity; **C:** protein C-terminus binding; **C:** voltage-gated calcium channel complex; **P:** calcium ion transmembrane transport; **C:** cytosol; **P:** positive regulation of cytosolic calcium ion concentration.
- **C:** sarcoplasmic reticulum; **C:** voltage-gated calcium channel complex; **F:** calcium channel regulator activity; **F:** voltage-gated calcium channel activity; **F:** mating pheromone activity; **P:** pheromone-dependent signal transduction involved in conjugation with cellular fusion; **C:** T-tubule; **C:** extracellular exosome; **F:** metal ion binding; **P:** mating; **P:** regulation of ion transmembrane transport; **P:** calcium ion transmembrane transport; **P:** regulation of calcium ion transport.
<table>
<thead>
<tr>
<th>Fold change</th>
<th>p adj</th>
<th>BLAST description</th>
<th>Length (bp)</th>
<th>GO Names list</th>
</tr>
</thead>
<tbody>
<tr>
<td>420.10</td>
<td>0.04</td>
<td>Cytochrome P450-like TBP</td>
<td>408</td>
<td></td>
</tr>
<tr>
<td>36.47</td>
<td>0.01</td>
<td>Cytochrome P450, family 4, subfamily F, polypeptide 22</td>
<td>458</td>
<td>F:oxygen binding; P:oxygen transport; C:voltage-gated potassium channel complex; F:delayed rectifier potassium channel activity; F:A-type (transient outward) potassium channel activity; P:protein homooligomerization; C:neuronal cell body; P:membrane repolarization; F:oxygen transporter activity; F:ion channel binding; F:heme binding; C:hemoglobin complex; C:dendrite; C: caveola; C:perinuclear endoplasmic reticulum; F:iron ion binding; P:cellular response to BMP stimulus; C: sarcolemma; P: regulation of ion transmembrane transport; P: potassium ion export</td>
</tr>
<tr>
<td>30.72</td>
<td>0.02</td>
<td>Cytochrome P450 2F2-like</td>
<td>1201</td>
<td></td>
</tr>
<tr>
<td>17.97</td>
<td>0.00</td>
<td>Cytochrome P450 2D26-like</td>
<td>1851</td>
<td>P:sucrose biosynthetic process; C: chromosome; F: binding</td>
</tr>
<tr>
<td>6.93</td>
<td>0.04</td>
<td>Cytochrome P450 2K1-like</td>
<td>429</td>
<td>C:nucleus; F: DNA binding; P: cytoskeleton organization; F: protein heterodimerization activity; F: calcium ion binding; C: nucleosome; F: calcium-dependent phospholipid binding</td>
</tr>
<tr>
<td>Fold change</td>
<td>Padj</td>
<td>BLAST description</td>
<td>Length (bp)</td>
<td>GO Names list</td>
</tr>
<tr>
<td>-------------</td>
<td>------</td>
<td>------------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td>273.95</td>
<td>0.00</td>
<td>Melanoregulin-like protein</td>
<td>1131</td>
<td>P:regulation of appetite; C:secretory granule; C:peroxisomal matrix; P:positive regulation of transcription from RNA polymerase II promoter; F:hormone activity; P:regulation of blood pressure; C:extracellular space; P:neuropeptide signaling pathway; P:glucose homeostasis; F:type 1 melanocortin receptor binding; P:cellular pigmentation; P:regulation of corticosterone secretion; F:type 3 melanocortin receptor binding; F:type 4 melanocortin receptor binding; P:regulation of glycogen metabolic process; P:negative regulation of tumor necrosis factor production</td>
</tr>
<tr>
<td>25.44</td>
<td>0.03</td>
<td>Melanocortin receptor binding NECAP endocytosis associated 2</td>
<td>2070</td>
<td>P:regulation of nucleobase-containing compound metabolic process; P:G-protein coupled receptor signaling pathway; P: single-multicellular organism process; F:peptide binding; F:melanocortin receptor activity; F:protein binding; P:cellular macromolecule metabolic process; F:carbohydrate derivative binding; C:intracellular membrane-bounded organelle; P:homeostatic process; C:membrane; P:regulation of gene expression; C:macromolecular complex</td>
</tr>
<tr>
<td>12.71</td>
<td>0.05</td>
<td>Melanocortin m-phase phosphoprotein 8</td>
<td>2732</td>
<td>P:protein binding; C:proton-transporting ATP synthase complex; P: single-organism cellular process; P:nucleobase-containing compound metabolic process; P:transport; C:mitochondrial inner membrane; F:hydrolase activity</td>
</tr>
<tr>
<td>13.52</td>
<td>0.00</td>
<td>Tyrosine kinase, non-receptor 2</td>
<td>1981</td>
<td>P:protein binding; C:proton-transporting ATP synthase complex; P: single-organism cellular process; P:nucleobase-containing compound metabolic process; P:transport; C:mitochondrial inner membrane; F:hydrolase activity</td>
</tr>
</tbody>
</table>
Supporting Figures

S1 Fig. Pie charts representing the number and proportion of contigs with significant BLAST hits (E<1.0E^-5) for each individual transcriptome from Oophaga species. Transcriptomes obtained using paired-ends RNA-seq libraries are represented in (A) and those obtained with single-end libraries in (A).
S2 Fig. A) Contig length distribution of the de-novo composite reference transcriptome of *Oophaga* species from skin tissue. B) Pie chart representing the number and proportion of contigs with significant BLAST hits (E<1.0E-5) in the reference *Oophaga* transcriptome. C) Gene ontology (GO) categories distribution (level II) for the annotated unigenes in the *Oophaga* reference transcriptome.
S3 Fig. Distribution plots of the effective number of reads (ER) from *Oophaga* RNA-seq experiments that mapped to the composite reference transcriptome contigs (n=31,498).
S4 Fig. a) Percentile plot of the estimated transcripts per million (TPM) in the composite reference transcriptome as calculated based on raw RNA reads from individual libraries. Color names are equivalent to those in tendency lines. b) Principal component analyses (PCA) plot of the reference contigs dataset using TPMs values as independent variables. Red dots represent the selected highly represented unigenes (2% of the total contigs; n=1,437) while dark dots represent the remaining ones (n=30,061).
S5 Fig. A) Pie chart representing the total number of differentially expressed unigenes and the proportion with significant BLAST hits (E<1.0E⁻⁵). B) BLAST hit species distribution of highly expressed unigenes. C) Gene ontology (GO) categories distribution (multilevel) for the annotated highly expressed unigenes.