

Transcriptome Profile of the Green Odorous Frog (*Odorrana margaretae*)

Liang Qiao^{1,2}, Weizhao Yang², Jinzhong Fu^{2,3*}, Zhaobin Song^{1,4*}

1 Sichuan Key Laboratory of Conservation Biology on Endangered Wildlife, College of Life Sciences, Sichuan University, Chengdu, China, **2** Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, China, **3** Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada, **4** Key Laboratory of Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, China

Abstract

Transcriptome profiles provide a practical and inexpensive alternative to explore genomic data in non-model organisms, particularly in amphibians where the genomes are very large and complex. The odorous frog *Odorrana margaretae* (Anura: Ranidae) is a dominant species in the mountain stream ecosystem of western China. Limited knowledge of its genetic background has hindered research on this species, despite its importance in the ecosystem and as biological resources. Here we report the transcriptome of *O. margaretae* in order to establish the foundation for genetic research. Using an Illumina sequencing platform, 62,321,166 raw reads were acquired. After a *de novo* assembly, 37,906 transcripts were obtained, and 18,933 transcripts were annotated to 14,628 genes. We functionally classified these transcripts by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). A total of 11,457 unique transcripts were assigned to 52 GO terms, and 1,438 transcripts were assigned to 128 KEGG pathways. Furthermore, we identified 27 potential antimicrobial peptides (AMPs), 50,351 single nucleotide polymorphism (SNP) sites, and 2,574 microsatellite DNA loci. The transcriptome profile of this species will shed more light on its genetic background and provide useful tools for future studies of this species, as well as other species in the genus *Odorrana*. It will also contribute to the accumulation of amphibian genomic data.

Citation: Qiao L, Yang W, Fu J, Song Z (2013) Transcriptome Profile of the Green Odorous Frog (*Odorrana margaretae*). PLoS ONE 8(9): e75211. doi: 10.1371/journal.pone.0075211

Editor: Marc Robinson-Rechavi, University of Lausanne, Switzerland

Received: July 3, 2013; **Accepted:** August 13, 2013; **Published:** September 20, 2013

Copyright: © 2013 Qiao et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work is supported by a National Natural Science Foundation of China (<http://www.nsf.gov.cn/Portal0/default152.htm>), grant number 31172061. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: jfu@uoguelph.ca (JF); zbsong@scu.edu.cn (ZS)

Introduction

Genomics has revolutionized many disciplines of biological science, and genomic data of non-model organisms are rapidly accumulating [1-3]. Among vertebrates, for example, a total of 61 annotated genomes are currently available from Ensembl database (June 21, 2013). Nevertheless, amphibians, as a major transitional group in the evolutionary history of vertebrates, are currently represented by only one species (*Xenopus tropicalis*) [4]. This is likely due to the fact that amphibians generally have very large and complex genomes [5,6]. As a small but important part of the genome, transcriptome includes most protein coding genes. With advances of the Next Generation Sequencing (NGS) technologies, transcriptome data offer an opportunity to deliver fast, inexpensive, and accurate genome information for genomic exploration in non-model organisms [7,8]. Transcriptome data are particularly useful for amphibians, because of their large genome sizes, which make obtaining whole genome data difficult. Currently, nine amphibian

transcriptomes have been reported, including *Xenopus tropicalis* [9], *Cyclorana alboguttata* [10], *Rana chensinensis* [11,12], *R. kukunoris* [11], *R. muscosa* [13], *R. sierra* [13], *Hyla arborea* [14], *Notophthalmus viridescens* [15,16], and *Ambystoma mexicanum* [17]. The accumulation of transcriptome data will provide a 'sneak peek' of amphibian genome evolution.

The green odorous frog *Odorrana margaretae* is a typical anuran from the family Ranidae, and is an important species both ecologically and in terms of biological resources. It is a dominant species in the stream ecosystem of the Hengduan Mountain, a global diversity hotspot [18], and is an excellent environmental indicator species. Odorous frog species are sensitive to environmental changes, and at least two of them have been observed to shift their ranges northward possibly as a response to global warming ([19], personal observation). Additionally, odorous frogs are reservoirs for antimicrobial peptides (AMPs), and may represent the most extreme AMPs diversity in nature [20]. AMPs are generally short peptides with potent antibacterial and antifungal activity [21]. So far, 728

different AMPs have been identified from nine odorous frog species, which account for approximately 30% of all AMPs discovered [20].

A transcriptome profile of the green odorous frog will provide new molecular markers for ecological research of this as well as other odorous frog species. Single nucleotide polymorphism (SNP) and microsatellite DNA loci are excellent genetic markers, and are commonly employed in population genetic and molecular ecological studies. They are abundant in transcriptomes. Furthermore, due to limitations of conventional biochemical isolation methods, some peptides cannot be isolated and purified. Transcriptome data will not suffer from these limitations and provide an important alternative way to identify potential AMPs.

In this study, we construct the transcriptome profile of *O. margaretae* using an Illumina sequencing platform. Multiple tissues types from multiple individuals were pooled to maximize the chance of revealing as many genes as possible. After *de novo* assembly, we implemented a functional annotation using bioinformatic analysis. In addition, we made a genome wide search for cDNA encoding AMPs, microsatellite DNA loci, and SNP sites from the transcriptome. Our data will serve as an important step forward to establish the foundation for genomic research of amphibians, as well as to provide new markers for molecular ecological studies.

Results

Illumina sequencing, *de novo* assembly, and gene annotation

Illumina sequencing of *O. margaretae* yielded a total of 62,321,166 raw reads. Among them, 3,504,433 were first filtered out before assembly as low quality sequences or potential contaminations. Thirty combinations of multiple K-mer lengths and coverage cut-off values were used to perform *de novo* assembly of the clean reads [22,23]. Finally, we merged the 30 raw assemblies by integrating sequence overlaps and eliminating redundancies. The final assembly included a total of 54.3 mega base pairs (Mb), and 37,906 transcripts were obtained with a N50 length of 1,870 base pairs (bps) and a mean length of 1,434 bps. The sequencing information is presented in Table 1 and the length distribution of all transcripts is shown in Figure 1. All sequence reads are deposited at NCBI (accession number SRA091981), and the final assembly is presented as supporting information (Sequence S1 and S2).

A reference dataset was constructed for gene annotation, which included protein data from seven species (*Anolis carolinensis*, *Danio rerio*, *Gallus gallus*, *Homo sapiens*, *Mus musculus*, *Oryzias latipes*, and *Xenopus tropicalis*), representing all major lineages of vertebrates. Transcripts were blasted against this dataset. In total, 18,933 transcripts were annotated to 14,628 genes, which comprised 49.95% of the total transcripts. E-value distribution showed that 73.89% of the annotated sequences had strong homology (E-value below $1E-50$), and similarity distribution showed that 72.50% of the annotated sequences had a similarity greater than 60% (Table 1 and Figure 2).

Table 1. Summary of transcriptome data for *Odorrana margaretae*.

Total number of raw reads	62,321,166
Total number of clean reads	58,816,733
Length of reads (bp)	101
Total length of clean reads	5.88G
Total length of assembly (bp)	54,362,822
Total number of transcripts	37,906
N50 length of assembly (bp)	1,870
Mean length of assembly (bp)	1,434
Median length of assembly (bp)	1,096
Transcripts annotated	18,933
Number of unique genes represented	14,628

bp = base pair

doi: 10.1371/journal.pone.0075211.t001

GO classification

Gene Ontology (GO) is widely used to standardize representation of genes across species and provides a set of structured and controlled vocabularies for annotating genes, gene products, and sequences [24]. In total, 11,457 unique transcripts were assigned to 52 level-2 GO terms, which were summarized under three main GO categories, including cellular component, molecular function, and biological process (Figure 3). Compared to the GO annotations of two other species of the same family, *R. chensinensis* and *R. kukunoris*, the GO category distributions of the transcripts for the three ranid frogs were highly similar (Figure S1). Within the GO category of cellular components, 14 level-2 categories were identified, and the terms cell, cell part, and organelle were the most abundant (>50%). Within the GO category of molecular function, 15 level-2 categories were identified, and the term binding was the most abundant (>50%). For biological process function, 23 level-2 categories were identified, and the terms of cellular process and metabolic process were the most abundant (>50%).

KEGG analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) [25] database was used to identify potential biological pathways represented in the *O. margaretae* transcriptome. A total of 1,438 transcripts were assigned to 128 KEGG pathways (Figure 4, Table S1). Among the pathways, purine metabolism, pyrimidine metabolism, phosphatidylinositol signaling system, and a few others were highly represented. These annotations provide a valuable resource for investigating specific processes, functions, and pathways in amphibian research.

Comparison to other amphibian transcriptomes

Overall, our data were similar to other reported amphibian transcriptomes from Illumina sequencing platform (Table 2). We had a slightly longer N50 length than most of the other transcriptomes, suggesting the quality of our assembly was high. Clearly, our strategy with multiple K-mer length and cut-off value combinations worked well. The transcriptomes of *R.*

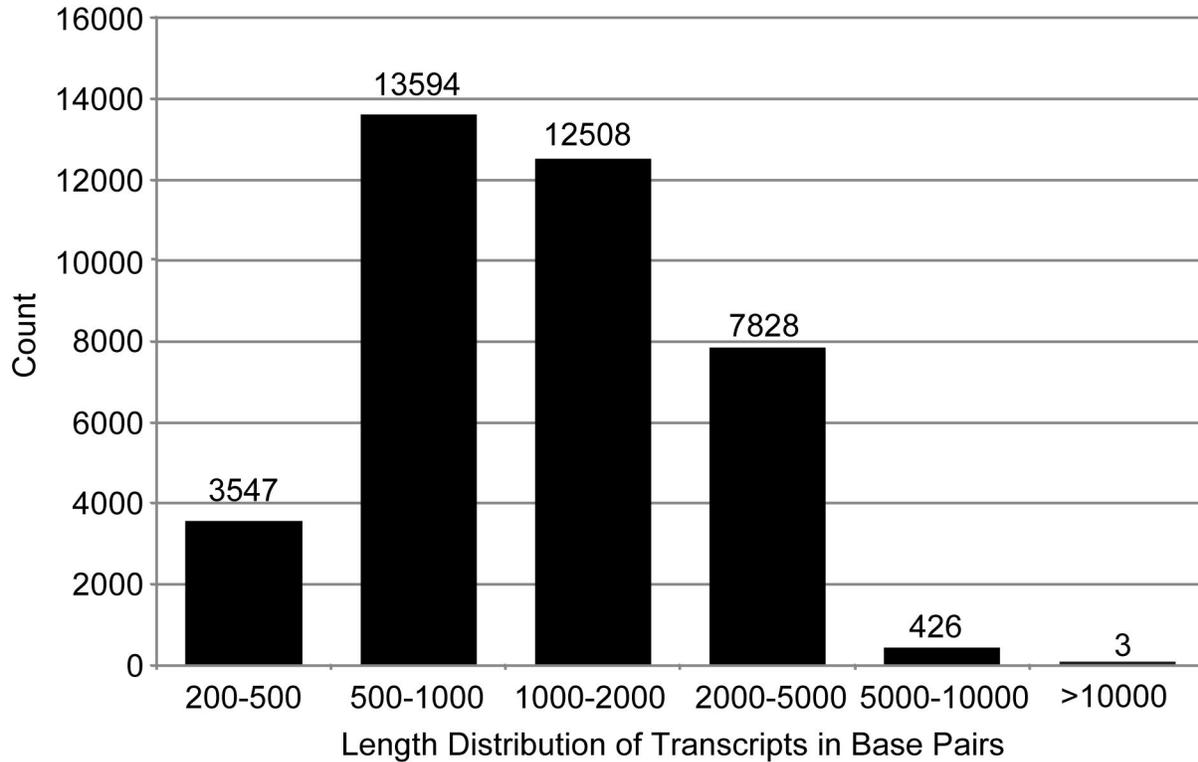


Figure 1. Length distribution of transcripts in base pairs. The numbers of transcripts are shown on top of each bar.

doi: 10.1371/journal.pone.0075211.g001

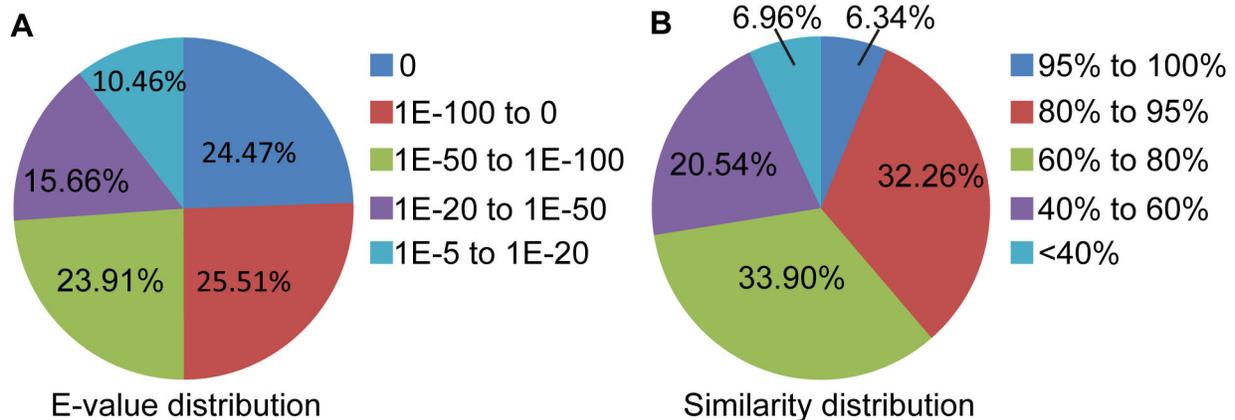


Figure 2. Characteristics of gene annotation of assembled transcripts against the reference dataset. (A) E-value distribution of BLASTX hits for transcript with a cut-off E-value of 1E-5. (B) Similarity distribution of BLASTX hits for transcript.

doi: 10.1371/journal.pone.0075211.g002

chensinensis [12] and *Hyla arborea* [14] had short N50 lengths and large numbers of transcripts, suggesting that the assembly generated many short transcripts and could be improved. Among the eight transcriptomes, *C. alboguttata* had the largest number of transcripts; this is not particularly surprising, considering that it also had the largest number of raw reads. In terms of transcript annotation rate, *O. margaretae* had

approximately 50% transcripts that were annotated, whereas the rates were approximately 33%, 40%, 44%, 42% and 32% for *C. alboguttata* [10], *R. chensinensis* [11], *R. chensinensis* [12], *R. kukunoris* [11], and *N. viridescens* [15], respectively.

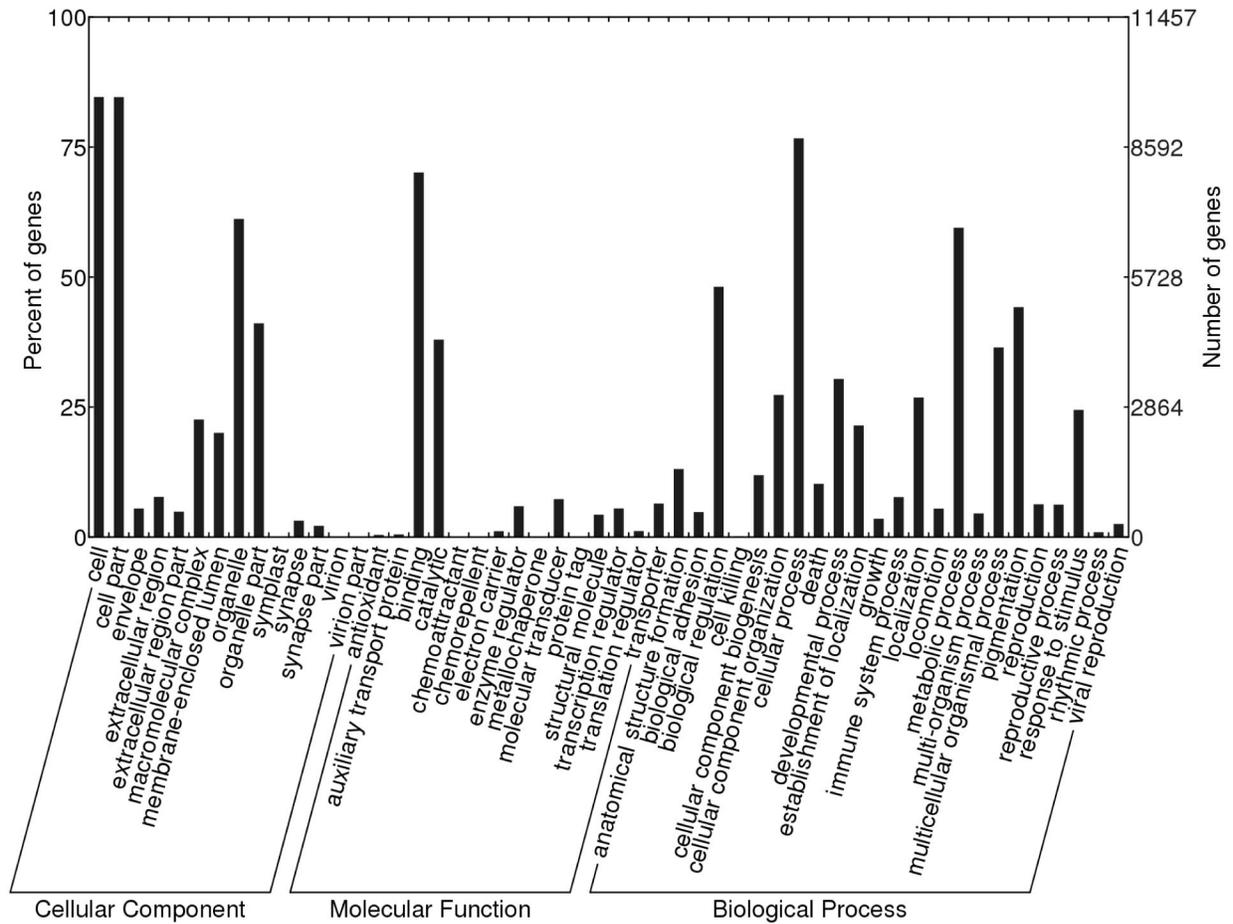


Figure 3. Distribution of Gene Ontology (GO) categories (level 2) of transcripts for *O. margaretae*. GO functional annotations are summarized in three main categories: cellular component, molecular function and biological process. doi: 10.1371/journal.pone.0075211.g003

AMPs

A total of 27 transcripts were identified as potential AMPs (Table S2). Among them, five transcripts matched AMPs that were previously identified in *O. margaratae*, including Esculentin-2-OMar1, Brevinin-2-OMar1, Odorrainin-A-Omar1, Esculentin-I-Omar4, and Margaratain-C1. The other 22 were new to *O. margaratae*. Although the exact functions of these peptides remain largely unknown, AMPs have received much attention lately [20,26], due to their anti-infective properties.

Molecular markers

A total of 50,351 single nucleotide polymorphism (SNP) sites were identified; 33,735 were transitions and 16,616 were transversions (Table S3). We also identified 2,574 microsatellite DNA loci, of which 78.59% were dinucleotide repeats, 19.54% were trinucleotide repeats, 1.40% were tetranucleotide repeats, 0.43% were pentanucleotide repeats, and 0.04% were hexanucleotide repeats (Table S4). These molecular markers will provide useful tools in population

genetic and molecular ecological studies of *O. margaretae*, and potentially for other odorous frogs as well.

Discussion

Available genomic data for amphibians are very limited and our transcriptome data of *O. margaretae* will certainly make a significant contribution to the understanding of genome evolution of amphibians. Currently, there is only one completed genome sequence for amphibians (i.e. *X. tropicalis*) available in public databases [4]. Amphibians often have large genome sizes (0.95-120.60 pg [6]), which makes genome sequencing and analysis difficult. With the current technological limitations, transcriptome data provide a viable alternative to whole genome sequencing. Although only representing a small portion of the genome, transcriptome includes most protein coding genes and arguably represent the most functional part of the genome. The accumulation of transcriptome data will not only offer us an opportunity to have a glance at the genome evolution of amphibians, but also establish foundation for gene expression level genomic studies. There are currently nine

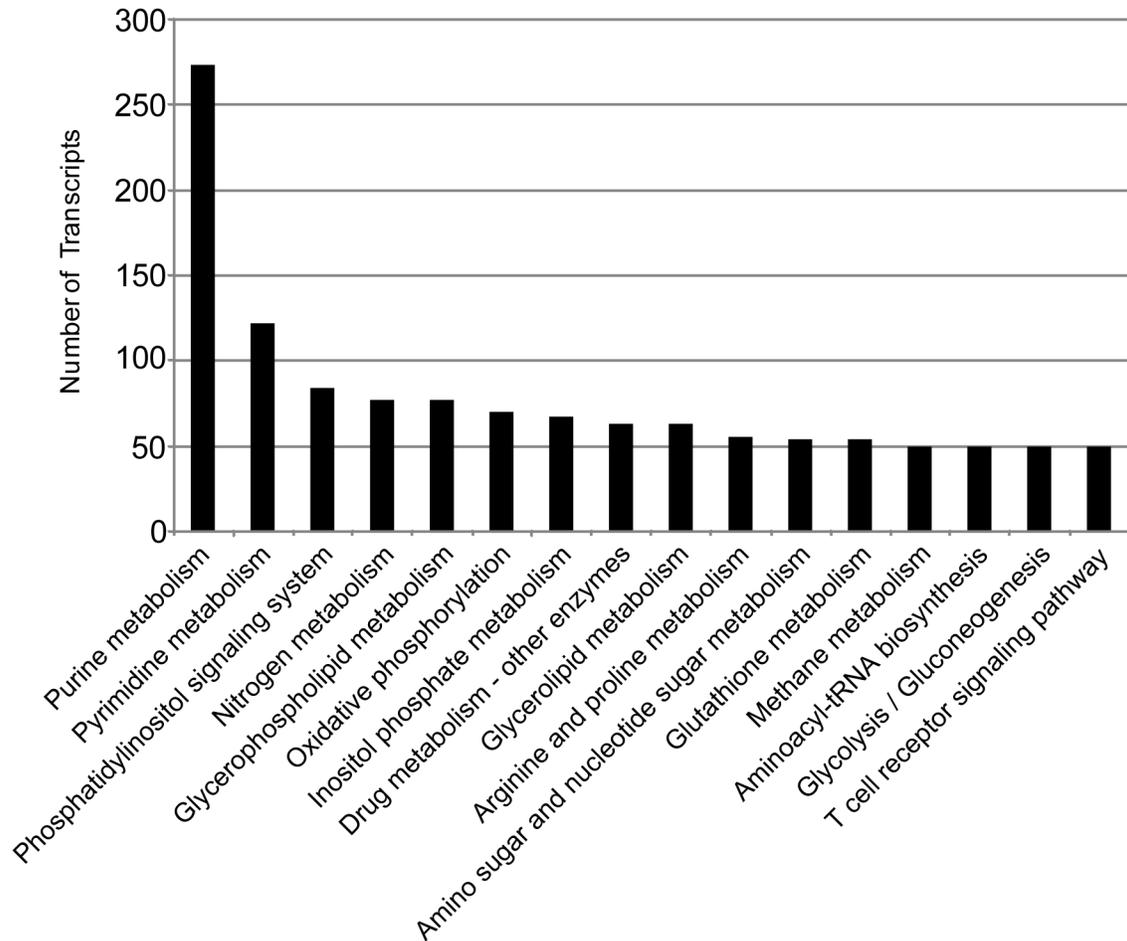


Figure 4. Distribution of *O. margaratae* transcripts among Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The top 16 most highly represented pathways are shown. Analysis was performed using Blast2GO and the KEGG database.

doi: 10.1371/journal.pone.0075211.g004

published amphibian transcriptomes [9-17]. More data are needed to establish patterns and formulate meaningful hypotheses.

We obtained a high quality *de novo* assembly by using a multiple K-mer lengths and cut-off values strategy. The N50 length of our assembly is higher than most other amphibian transcriptome data without a significant reduction of the overall number of unique transcripts (Table 2). N50 length is commonly used for assembly evaluation, and a higher number suggests high quality assembly [27]. In addition, our high transcripts annotation rate also indicates that the accuracy of assembled transcripts is greater than those assembled in other amphibians (Table 2). Clearly, our strategy worked well in this case.

We only identified a small number of AMPs. Previously, 72 mature AMPs have been identified in *O. margaratae* by peptidomic analysis [20], and we only found five homologous to these 72. This is most likely due to the fact that we did not include skin tissue samples in our RNA extraction, which is

often a rich resource for AMPs [21]. Nevertheless, our study demonstrated that AMPs not only exist in the amphibian skin, but also in other tissues, such as stomach and brain, which is consistent with several previous studies [28-30]. In addition, we identified 22 new AMPs for *O. margaratae*, and some of them have been identified in the skin of other *Odorrana* species, such as Brevinin-2-Omar1 and Odorrana-A-Omar1 [31]. This suggests that transcriptome is a valid alternative way for AMP discovery.

We identified a large number of microsatellite DNA and SNP loci for *O. margaratae*. In comparison to conventional methods for microsatellite DNA isolation (e.g. FIASCO protocol [32]), transcriptomes and NGS provide a fast, economical, and high-throughput alternative. It also selects multiple types of repeats simultaneously and is not limited by types of probes. Nevertheless, an initial characterization of these loci suggests that microsatellite DNA loci from the UTR regions generally have lower allelic diversity compared to conventionally selected loci (unpublished data). Molecular markers play an important

Table 2. Comparison of transcriptome data for six amphibian species.

	Number of raw reads	Number of transcripts	N50 length of assembly (bp)	Transcripts annotated
<i>Odorrana margaretae</i>	62,321,166	37,906	1,870	18,933
<i>Cyclorana alboguttata</i> [10]	400,032,568	68,947	1,586	22,695
<i>Rana chensinensis</i> [11]	67,676,712	41,858	1,333	16,738
<i>Rana chensinensis</i> [12]	39,300,002	78,117	413	34,706
<i>Rana kukunoris</i> [11]	66,476,534	39,293	1,485	16,549
<i>Hyla arborea</i> [14]	11,034,721	83,293	700	-
<i>Notophthalmus viridescens</i> [15]	-	120,922	975	38,384
<i>Notophthalmus viridescens</i> [16]	-	118,893	2,016	-

bp=base pair; - =data not available

doi: 10.1371/journal.pone.0075211.t002

role in contemporary biological research [33]. These microsatellite DNA loci and SNP sites will facilitate research of *O. margaretae*, as well as other amphibian species.

Conclusions

Using next generation sequencing technology, we produced a transcriptome profile for an odorous frog species (*O. margaretae*). The profile is similar to other published amphibian transcriptomes. In addition, we identified 27 potential AMPs, which confirmed that NGS can serve as an alternative way for AMP discovery. A large number of microsatellite DNA loci and SNP sites were also identified, which will facilitate studies on population genetics of odorous frogs. Perhaps most importantly, our data represent a significant contribution to the accumulation of genomic data of amphibians.

Materials and Methods

Sample collection

In order to recover as many expressed genes as possible, multiple tissues from three adult individuals (one male and two females) and one tadpole were used. For adults, seven types of tissues (cerebrum, eye, skeletal muscle, heart, liver, testicle, and *Ootheca*) were collected; for tadpole, the whole body was used after removing the guts. Samples of one male, one female, and the tadpole were collected from Mt. Emei (Sichuan, China; E103°38918', N29° 56418', 749m) in August 2012. The other female was collected from Xiaohogou Nature Reserve (Sichuan, China; E104°47257', N32° 53149', 1393m; approximately 345 km from the first site) in September 2012. All individuals were euthanized by immersion in MS-222 buffered solution (3g/L), and tissues were collected and stored

in Sample Protector Solution (TAKARA) immediately after euthanasia. All individuals were identified by both morphological and molecular (mitochondrial cytochrome *b* gene sequence) traits.

cDNA library construction and Illumina sequencing

Total RNA was extracted from each tissue sample individually using TRIzol® reagent (Life Technologies) and mixed with approximately same quantity. RNA integrity was assessed using the RNA 6000 Nano Assay Kit with a Bioanalyzer 2100 (Agilent Technologies) after checking the RNA purity and concentration. A single cDNA library was constructed. The mRNA was purified from total RNA using poly-T oligo-attached magnetic beads (Life Technologies). The first cDNA strand was synthesized using random oligonucleotides and M-MuLV Reverse Transcriptase (RNase H-). The second cDNA strand synthesis was subsequently performed using DNA Polymerase I and RNase H. The cDNA with an insert size of 200 bp were preferentially purified with AMPure XP beads system (Beckman Coulter) and sequenced on an Illumina HiSeq 2000 platform. 101 bp paired-end reads were generated and all raw sequence read data were stored in FastQ format. Both cDNA library construction and Illumina sequencing were performed by NovoGene (Beijing).

Data filtration and *de novo* assembly

We first filtered the raw reads by removing the adapter sequences, reads with unknown bases call (N) more than 5%, and low quality sequences (<Q20) using an in-house workflow of Novogene and Trimmomatic [34]. Then we removed reads that were likely derived from contaminants of human and *Escherichia coli* genomes using Bow tie [35]. *De novo* assembly of clean reads was carried out using a strategy of multiple K-mer lengths and coverage cut-off values. Five different K-mer lengths (21, 31, 41, 51, and 61) and six coverage cut-off values (2, 3, 6, 10, 15, and 20) were used to generate 30 raw assemblies by ABYSS [36]. The raw assemblies were merged to produce a combined assembly. Then CD-HIT-EST [37] was used to eliminate redundancies with sequence identity threshold of 1.0 and word length of 8. CAP3 [38] was used to integrate sequence overlaps with default parameters for three times. A final assembly was generated after removing contigs shorter than 200 bp. All raw sequence reads were mapped back to the final assembly to identify variable sites using Bow tie [35] and SAMtools pipeline [39]. The base call that was consistent with the most mapped reads at a variable site was chosen for the consensus sequences using an in-house Python script.

Gene annotation and GO/KEGG classification

We first constructed a reference dataset for gene annotation, because of the lack of genomic information of any *Odorrana* species. The dataset combined protein data of seven vertebrate species from the ENSEMBL Database [40], including *Anolis carolinensis*, *Danio rerio*, *Gallus gallus*, *Homo sapiens*, *Mus musculus*, *Oryzias latipes*, and *Xenopus tropicalis*. Assembled sequences were annotated to the

reference dataset based on BLAST similarity using BLASTX [41] with an E-value cut-off of 1E-05.

GO categories and KEGG pathways were used to classify the functions and metabolic pathways of the transcripts. In order to exclude the interference from alternative splicing of transcripts, we first clustered all transcripts that matched the same reference gene; then we removed redundant transcripts and only preserved the longest transcript from each cluster to represent a unique gene. GO and KEGG classification was performed using the Blast2GO [42] pipelines with the default parameters.

AMPs and molecular markers

In order to identify AMPs in the transcriptome of green odorous frog, we blasted the assembled transcripts against the known AMPs from Database of Anuran Defense Peptides (DADP) [43] using Blast-2.2.26+ [41] with the similarity cutoff of 80%. SNP sites were identified using SAMtools [39] pipeline after mapping all clean reads to the assembled transcripts using Bow tie with default parameters [35]. Microsatellite DNA loci were identified by QDD2 pipeline [44], and the same pipeline also automatically designed all associated primers.

Ethics Statement

The animal specimens were collected legally. All animal collection and utility protocols were approved by the Chengdu Institute of Biology Animal Use Ethics Committee.

Supporting Information

Figure S1. Distribution comparison of Gene Ontology (GO) categories (level 2) of transcripts among three ranid

frogs, *O. margaretae*, *R. chensinensis*, and *R. kukunoris*. (TIF)

Table S1. Distribution of *O. margaretae* transcripts among KEGG pathways. (XLSX)

Table S2. Putative AMPs in *O. margaretae* sequences. (XLSX)

Table S3. SNPs in *O. margaretae* sequences. (XLSX)

Table S4. Microsatellite DNA loci in *O. margaretae* sequences. (XLSX)

Sequence S1. The final assembly of *O. margaretae* transcriptome (part 1). (RAR)

Sequence S2. The final assembly of *O. margaretae* transcriptome (part 2). (RAR)

Acknowledgements

We thank Dr. B. Lu for assisting with RNA isolation, Dr. Y. Qi for help with sample collection, and Dr J. Liu for data analysis. Dr. B. Lu and Dr. Y. Qi kindly read and commented on an early version of this manuscript.

Author Contributions

Conceived and designed the experiments: JF ZS. Performed the experiments: LQ WY. Analyzed the data: LQ WY. Contributed reagents/materials/analysis tools: LQ WY. Wrote the manuscript: LQ WY JF.

References

- Schuster SC (2008) Next-generation sequencing transforms today's biology. *Nat Methods* 5: 16-18. PubMed: 18165802.
- Li R, Fan W, Tian G, Zhu H, He L et al. (2010) The sequence and *de novo* assembly of the giant panda genome. *Nature* 463: 311-317. doi: 10.1038/nature08696. PubMed: 20010809.
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM et al. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* 12: 499-510. doi:10.1038/nrg3012. PubMed: 21681211.
- Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J et al. (2010) The Genome of the western clawed frog *Xenopus tropicalis*. *Science* 328: 633-636. doi:10.1126/science.1183670. PubMed: 20431018.
- Gregory TR (2002) Genome size and developmental complexity. *Genetica* 115: 131-146. doi:10.1023/A:1016032400147. PubMed: 12188045.
- Gregory TR (2005) Animal Genome Size Database. Available: <http://www.genomesize.com>.
- Gibbons JG, Janson EM, Hittinger CT, Johnston M, Abbot P et al. (2009) Benchmarking next-generation transcriptome sequencing for functional and evolutionary genomics. *Mol Biol Evol* 26: 2731-2744. doi: 10.1093/molbev/msp188. PubMed: 19706727.
- Metzker ML (2010) Sequencing technologies - the next generation. *Nat Rev Genet* 11: 31-46. doi:10.1038/nrg2626. PubMed: 19997069.
- Tan MH, Au KF, Yablonovitch AL, Wills AE, Chuang J et al. (2013) RNA sequencing reveals a diverse and dynamic repertoire of the *Xenopus tropicalis* transcriptome over development. *Genome Res* 23: 201-216. doi:10.1101/gr.141424.112. PubMed: 22960373.
- Reilly BD, Schlipalius DI, Cramp RL, Ebert PR, Franklin CE (2013) Frogs and estivation: transcriptional insights into metabolism and cell survival in a natural model of extended muscle disuse. *Physiol Genomics* 45: 377-388. doi:10.1152/physiolgenomics.00163.2012. PubMed: 23548685.
- Yang W, Qi Y, Bi K, Fu J (2012) Toward understanding the genetic basis of adaptation to high-elevation life in poikilothermic species: a comparative transcriptomic analysis of two ranid frogs, *Rana chensinensis* and *R. kukunoris*. *BMC Genomics* 13: 588. doi: 10.1186/1471-2164-13-588. PubMed: 23116153.
- Zhang M, Li Y, Yao B, Sun M, Wang Z et al. (2013) transcriptome sequencing and *de novo* analysis for *Oviductus Ranae* of *Rana chensinensis* using Illumina RNA-seq technology. *J Genet Genomics* 40: 137-140. doi:10.1016/j.jgg.2013.01.004. PubMed: 23522386.
- Rosenblum EB, Poorten TJ, Settles M, Murdoch GK (2012) Only skin deep: shared genetic response to the deadly chytrid fungus in susceptible frog species. *Mol Ecol* 21: 3110-3120. doi:10.1111/j.1365-294X.2012.05481.x. PubMed: 22332717.
- Brelsford A, Stöck M, Betto-Colliard C, Dubey S, Dufresnes C et al. (2013) Homologous sex chromosomes in three deeply divergent anuran species. *Evolution* 67: 2434-2440. doi:10.1111/evo.12151. PubMed: 23888863.
- Looso M, Preussner J, Sousounis K, Bruckskotten M, Michel CS et al. (2013) A *de novo* assembly of the newt transcriptome combined with proteomic validation identifies new protein families expressed during tissue regeneration. *Genome Biol* 14: R16. doi:10.1186/gb-2013-14-2-r16. PubMed: 23425577.
- Abdullayev I, Kirkham M, Björklund AK, Simon A, Sandberg R (2013) A reference transcriptome and inferred proteome for the salamander *Notophthalmus viridescens*. *Exp Cell Res* 319: 1187-1197. doi:10.1016/j.yexcr.2013.02.013. PubMed: 23454602.
- Stewart R, Rascón CA, Tian S, Nie J, Barry C et al. (2013) Comparative RNA-seq analysis in the unsequenced axolotl: the

- oncogene burst highlights early gene expression in the blastema. *PLOS Comput Biol* 9: e1002936. PubMed: 23505351.
18. Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403: 853-858. doi:10.1038/35002501. PubMed: 10706275.
 19. Gao H, Qiao L, Yang W, Fu J (2013) Isolation and characterization of 13 microsatellite DNA loci for the odorous frog *Odorrana margaretae* and *O. graminea* (Anura: Ranidae). *Conserv Genet Resour* (Online). doi:10.1007/s12686-013-9936-2.
 20. Yang X, Lee WH, Zhang Y (2012) Extremely abundant antimicrobial peptides existed in the skins of nine kinds of Chinese odorous frogs. *J Proteome Res* 11: 306-319. doi:10.1021/pr200782u. PubMed: 22029824.
 21. Barra D, Simmaco M (1995) Amphibian skin: a promising resource for antimicrobial peptides. *Trends Biotechnol* 13: 205-209. doi:10.1016/S0167-7799(00)88947-7. PubMed: 7598843.
 22. Surget-Groba Y, Montoya-Burgos JI (2010) Optimization of *de novo* transcriptome assembly from next-generation sequencing data. *Genome Res* 20: 1432-1440. doi:10.1101/gr.103846.109. PubMed: 20693479.
 23. Gruenheit N, Deusch O, Esser C, Becker M, Voelckel C et al. (2012) Cutoffs and k-mers: implications from a transcriptome study in allopolyploid plants. *BMC Genomics* 13: 92. doi: 10.1186/1471-2164-13-92. PubMed: 22417298.
 24. GO Consortium (2008) The Gene Ontology project in 2008. *Nucleic Acids Res* 36: D440-D444. PubMed: 17984083.
 25. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28: 27-30. doi:10.1093/nar/28.7.e27. PubMed: 10592173.
 26. Liu J, Jiang J, Wu Z, Xie F (2012) Antimicrobial peptides from the skin of the Asian frog, *Odorrana jingdongensis*: *de novo* sequencing and analysis of tandem mass spectrometry data. *J Proteomics* 75: 5807-5821. doi:10.1016/j.jprot.2012.08.004. PubMed: 22917879.
 27. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921. doi:10.1038/35057062. PubMed: 11237011.
 28. Moore KS, Bevins CL, Brasseur MM, Tomassini N, Turner K et al. (1991) Antimicrobial peptides in the stomach of *Xenopus laevis*. *J Biol Chem* 266: 19851-19857. PubMed: 1717472.
 29. Cho JH, Sung BH, Kim SC (2009) Buforins: Histone H2A-derived antimicrobial peptides from toad stomach. *Biochim Biophys Acta* 1788: 1564-1569. doi:10.1016/j.bbame.2008.10.025. PubMed: 19041293.
 30. Liu R, Liu H, Ma Y, Wu J, Yang H et al. (2011) There are abundant antimicrobial peptides in brains of two kinds of *Bombina* toads. *J Proteome Res* 10: 1806-1815. doi:10.1021/pr101285n. PubMed: 21338048.
 31. Li J, Xu X, Xu C, Zhou W, Zhang K et al. (2007) Anti-infection peptidomics of amphibian skin. *Mol Cell Proteomics* 6: 882-894. doi: 10.1074/mcp.M600334-MCP200. PubMed: 17272268.
 32. Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol* 11: 1-16. doi:10.1046/j.0962-1083.2001.01418.x. PubMed: 11903900.
 33. Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* 4: 981-994. doi:10.1038/nrm1255. PubMed: 14631358.
 34. Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE et al. (2012) *RobiNA*: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res* 40: W622-W627. doi: 10.1093/nar/gks540. PubMed: 22684630.
 35. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25. doi:10.1186/gb-2009-10-3-r25. PubMed: 19261174.
 36. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM et al. (2009) ABySS: A parallel assembler for short read sequence data. *Genome Res* 19: 1117-1123. doi:10.1101/gr.089532.108. PubMed: 19251739.
 37. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658-1659. doi:10.1093/bioinformatics/btl158. PubMed: 16731699.
 38. Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9: 868-877. doi:10.1101/gr.9.9.868. PubMed: 10508846.
 39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078-2079. doi:10.1093/bioinformatics/btp352. PubMed: 19505943.
 40. Hubbard TJP, Aken BL, Beal K, Ballester B, Caccamo M et al. (2007) Ensembl 2007. *Nucleic Acids Res* 35: D610-D617. doi:10.1093/nar/gkl996. PubMed: 17148474.
 41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410. doi:10.1016/S0022-2836(05)80360-2. PubMed: 2231712.
 42. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676. doi: 10.1093/bioinformatics/bti610. PubMed: 16081474.
 43. Novković M, Simunić J, Bojović V, Tossi A, Juretić D (2012) DADP: the database of anuran defense peptides. *Bioinformatics* 28: 1406-1407. doi:10.1093/bioinformatics/bts141. PubMed: 22467909.
 44. Megléc Z, Costedoat C, Dubut V, Gilles A, Malausa T et al. (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* 26: 403-404. doi:10.1093/bioinformatics/btp670. PubMed: 20007741.