Genome Biology

CORRESPONDENCE

Are there unequivocal criteria to label a given protein as a toxin? Permissive versus conservative annotation processes

Yves Terrat[1]* and Frédéric Ducancel[2]

A response to **Novel venom gene discovery in the platypus** by Whittington CM, Papenfuss AT, Locke DP, Mardis ER, Wilson RK, Abubucker S, Mitreva M, Wong ESW, Hsu AL, Kuchel PW, Belov K, Warren WC. *Genome Biol* 2010, **11**:R95.

During the past decade, the field of toxin discovery has changed in many aspects, including the study of new phyla and the application of new discovery strategies. High-throughput screening at the cDNA level has been successful in characterizing a broad range of toxin sequences [1-7]. In an article published in *Genome Biology,* Whittington and colleagues [8] report the discovery of novel toxin genes in the platypus, expanding the small list of previously characterized venom compounds in this fascinating species [9-15]. At first glance, we were surprised to find the expression of sarafotoxin-like sequences as, to date, no similar sequences have been identified outside the small genus of burrowing snakes (*Atractaspis*). A preliminary analysis yielded strong evidence for the non-toxin function of these sarafotoxin-like sequences. Hence, we decided to re-evaluate the original venom-gene list, which includes 83 sequences. Here, we address issues that question the validity of the annotation method adopted by Whittington *et al.*, which is largely based on similarity searches using toxin-related genes and on a tissue-expression criterion that might lead to false positives.

Why is the similarity-based annotation process prone to false-positive toxin identifications?

As correctly mentioned by Whittington *et al.*, most proteins found in venoms are the result of toxin recruitment events in which an ordinary protein gene, typically one involved in a key regulatory process, is duplicated, and the new gene is selectively expressed in the venom gland. In many cases, such toxin genes are amplified to obtain multigene families with extensive neofunctionalization [10,16,17]. Alternatively, the similarities shared with non-venom-related genes could also originate from a process of convergent evolution. In both cases, the result is a high probability of retrieving non-venom-related genes through a basic local alignment search tool (BLAST) similarity search using the given toxin as the query. Based on these observations, Whittington and colleagues state that rejecting potential platypus venom genes on the basis of similarities with a non-venom gene is inappropriate. However much we agree with this point, we suggest that additional validation steps have to be performed to ensure that the toxin candidate is not in itself a non-venom gene, for the following reasons. First, the annotation process adopted by Whittington and co-workers is particularly sensitive to detection of false positives because it uses the Tox-Prot database enriched in vertebrate toxins [18] as a query to perform a TBLASTN search on the platypus cDNA database. Second, their conclusions were derived largely from top BLAST hit results, and that might increase the chance of retrieving cellular genes commonly expressed in vertebrate cells. Third, strong similarities to non-venom-related genes are discussed for some families, but no clear procedure is used to evaluate their impact. Finally, the likelihood of each candidate to be a toxin is evaluated through tissue-expression criteria, an approach for which the false detection rate needs further investigation.

Why is the tissue-expression criterion too permissive?

First, in order to validate the putative venom function of BLAST-annotated candidates, the authors chose to screen EST databases of different platypus tissues. Candidates expressed in at least three non-venomous tissues were removed. We believe that this criterion is not stringent enough for several reasons. First, as also

*Correspondence: yves.terrat@umontreal.ca
[1]Montréal University/Institute in Plant Biology, Montreal Botanical Garden, 4101 Sherbrooke East, Montreal, Québec, Canada, H1X 2B2
Full list of author information is available at the end of the article

mentioned by the authors, venom compounds could be expressed at a basal level in various tissues, much in the same way as non-venom-related genes could be expressed in venomous tissues. This could lead to both the rejection of true candidates and the annotation of a false-positive as toxin. From a statistical point of view, this 'switch-on / switch-off' approach lacks both power and sensitivity.

The second issue of this validation step is that, depending on the size of the different tissue databases, some compounds could simply be missing owing to sampling effects. For instance, the authors used an EST database of fibroblast cell lines that includes 9,699 EST sequences available in GenBank. In this database, we identified 8,813 ESTs from 'true' fibroblast cell lines. The remainder (886 sequences) include both liver and spleen cDNA of low quality (mean length of 176 bp, standard error of 94 bp). Unfortunately, we were unable to evaluate the size and quality of the bill, brain, liver, spleen and testis databases generated for the platypus genome because we were unable to locate them. The original 2008 *Nature* journal publication of the platypus genome only mentioned the fibroblast cell line database, which was used to add experimental support to the *in silico* gene prediction [15]. In Whittington and colleagues' study, only a small number of transcripts were generated (probably the 8,813 ESTs mentioned above).

Our final point is that, if one looks closely at the Illumina read counts for the toxin candidates, more than one-third of the putative candidates match fewer than 50 Illumina reads (about 3.10×10^{-6} of the total reads), and three of them were not even detected in the Illumina reads but in the 454 reads only. To us, it seems rather contradictory to try to validate a very low transcribed toxin through tissue expression criteria - it lacks at least some statistical support.

For all these reasons, we believe that the tissue expression criteria will not be valid until applied, with caution, in combination with databases generated through high-throughput means.

Are there any sarafotoxin-like sequences in the platypus venom?

The starting point that led us to reconsider the toxin list was the discovery of a sarafotoxin sequence (SRTX) in the platypus transcriptome. SRTXs are highly toxic components isolated from the venom of scarce *Atractaspis* snakes. SRTXs are similar to the endothelins (ETs) of vertebrates, which function as potent vasoconstrictors and act by means of identical receptors [19-23]. However, the similarity between SRTX and ET sequences is restricted to the endothelin region that corresponds to the mature peptide (approximately 21 amino acids in length; Figure 1a). The SRTX sequence presented by Whittington *et al*. (named edn1_O.ananitus in our study) is clearly labeled as a 'previously unknown toxin' and seems not to be expressed (or at least detected) in non-venom tissues. Because of its ranking in the list, the authors classified this unknown toxin in the top 33 list of 'probable (likely)' platypus venom genes. In our opinion, this sequence could be excluded from the original list for many reasons, as summarized below.

First, the transcript presents a typical ET-type and not a SRTX-type organization (Figure 1a). SRTXs and ETs contain a common core of approximately 21 amino acids (containing the endothelin peptide) but have different precursor organizations: SRTXs from members of the *Atractaspis* genus present various repetitions of the ET motif, whereas the precursors of ETs contain a single and complete 21 amino acid motif followed by one peptide of 16 residues that displays a highly divergent amino acid composition, except for the conserved cysteine residues, a so-called 'ET-like' core motif (Figure 1a).

Second, the putative venom gene identified by Whittington and colleagues shares more similarities to ET sequences (71% identity and 82% similarity with endothelin-1 from *Mus musculus* on the full-length sequence of 202 amino acids) than to any SRTX sequences (the best score for *Atractaspis microlepidota* SRTX with 15 of 21 residues identical is restricted to the endothelin region; see alignment in Figure 1a). Indeed, it is nearly perfectly identical to the endothelin-1 annotation suggested by the Ensembl genebuild.

In order to establish the phylogenetic position of the putative toxin, we performed a maximum likelihood (ML) analysis using the complete endothelin platypus repertory (edn1, edn2 and edn3 identified by Braasch and colleagues [24]), a subset of orthologous vertebrate sequences (edn1, edn2 and edn3 but not edn4 fish-specific genes) and SRTX sequences. We were unable to identify full-length edn2 and edn3 sequences owing to low conservation outside the 'big-endothelin' region and the ambiguous sequenced regions of genomic contigs. Thus, the nucleotide phylogeny of endothelin genes was reconstructed using 108 bp within the 'big-endothelin' domain. The endothelin phylogeny (Figure 1b) confirms that the 'new sarafotoxin-like gene' groups within the edn1 clade. These results are in agreement with the phylogenetic analysis performed by Brassch and collaborators that included numerous ETs and SRTXs.

Finally, we performed synteny analysis of the Ultracontig 474, which contains the gene encoding edn-1. We found that the syntenic relationship of the edn-1 block among vertebrates is also conserved for the platypus (data not shown). Combining gene organization, similarity/identity evidence, phylogenetic analysis and syntenic conservation of the edn-1 block, we believe that this sequence should not be labeled as a toxin but probably be

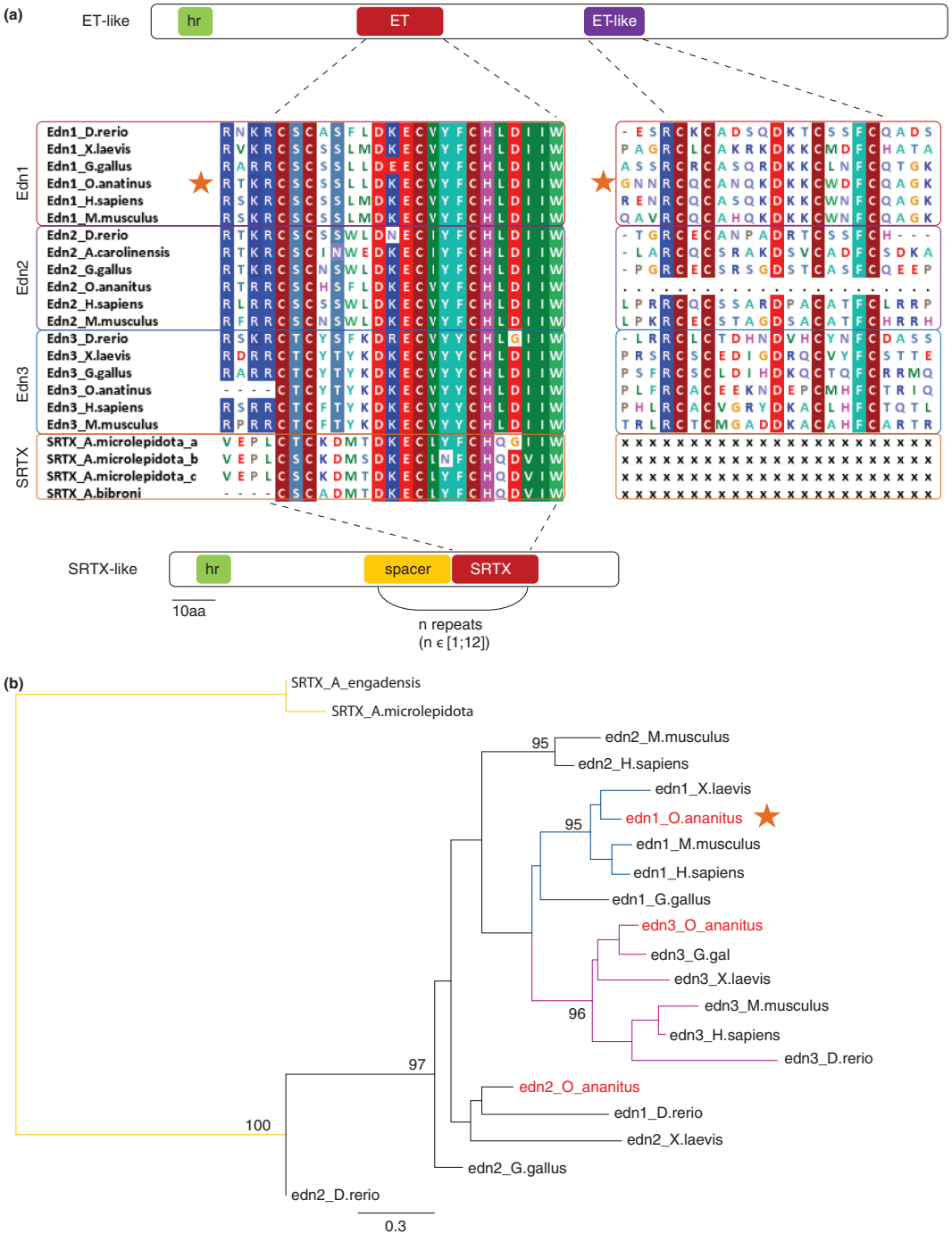

Figure 1. Schematic representation, protein sequence alignment and maximum likelihood (ML) nucleotidic phylogeny of endothelin and sarafotoxin sequences. The novel toxin gene identified by Whittington and colleagues (named edn1_O.ananitus) is identified by an orange star. For each endothelin (ET) group, we used two mammalian sequences (*Homo sapiens* and *Mus musculus*), one archosaurian sequence (*Gallus gallus*), one amphibian sequence (*Xenopus laevis*) and one actinopterygian sequence (*Danio rerio*). Only one copy of the fish-specific-duplicated edn-2 and edn-3 gene products was used in the alignment and in the phylogenetic analysis for *D. rerio*. Sequences from the fish-specific edn4 group have not been included. **(a)** The ET-like structure is represented at the top of the figure; the SRTX-like structure is represented at the bottom. The 'X' used in the alignment for Sarafotoxin sequences means that no homologous regions are identified. Three different sarafotoxin isoforms isolated from *Atractaspis microlepidota* and encoded on the same precursor are used for multiple alignment. The dots used in the alignment for edn2 of the *Ornithorhynchus ananitus* sequence indicate missing data. The threshold for identity/similarity shading is 80%. **(b)** ML was performed on the ET/SRTX region of 108 bp using the generalized time-reversible plus invariant plus gamma (GTR+I+G) model. Topology is rooted with sarafotoxin sequences. Bootstrap values greater than 90% are indicated. The identified edn-1, edn-2 and edn-3 platypus sequences are indicated in red.

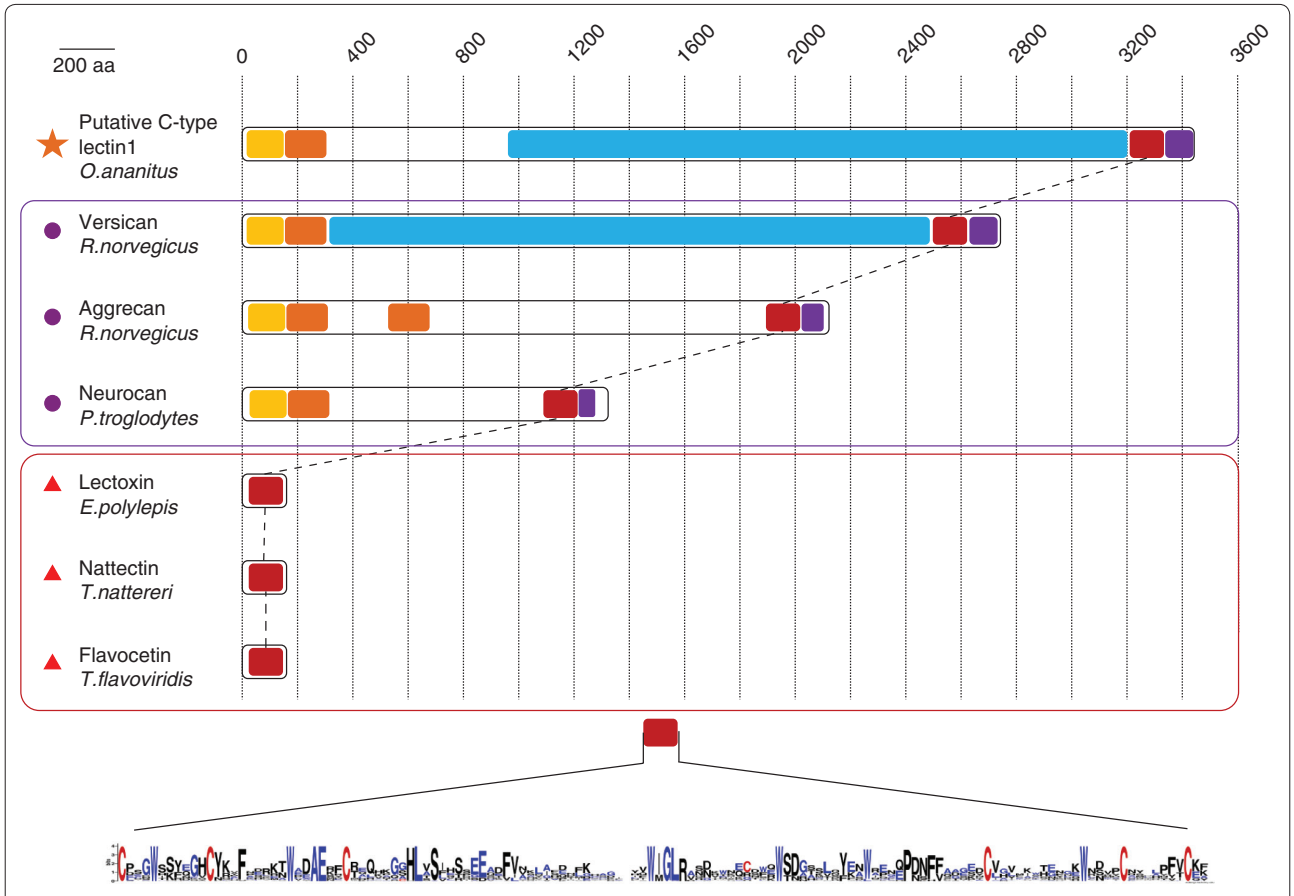

Figure 2. Comparative analysis of the domain organization of the C-type lectin-domain-containing protein. The dataset includes representative sequences of each protein group similar to the putative C-type lectin toxin. The different groups were identified after phylogenetic analysis of BLASTp hits retrieved on Tox-Prot and Uniprot databases associated with an e-value below 1^{E}-05. To identify domains similar to those of platypus sequence parts, an alignment of the putative toxin and similar hits was performed using BLASTp. Domains associated with an e-value less than 1^{E}-05 are indicated by colored boxes. The only shared motif is the C-type lectin domain. A sequence logo of the multiple alignment of C-type-lectin domains is represented above. Compared are the rat *Rattus norvegicus*, the chimpanzee *Pan troglodytes*, the water snake *Enhydris polylepsis*, the toadfish *Thalassophryne natterei* and the pit viper *Trimeresurus flavoviridis*.

considered as possessing an endogenous function, comparable to that of edn-1 sequences found in most vertebrates. Of course, with a pure bioinformatics approach, it is not possible to exclude the possibility of a toxin nature for this sequence. However, our analysis produces the most parsimonious hypothesis.

Are there other ambiguous attributions?

Indeed, ambiguous annotations are suspected for most of the new platypus toxin candidates. For instance, the putative C-type lectin 1 identified in the platypus transcriptome matches over nearly 3,500 amino acids in length on Uniprot sequences and less than 200 amino acids on the most similar Tox-prot sequences. Focusing on the matching domain (Figure 2), we observed that the C-type lectin domain (which is the only functional domain of relevant toxins) represents a small carboxy-terminal part of the full-length platypus toxin candidates (less than 4% of the full platypus protein sequence). The domain organization is indeed more similar to mammalian versican proteins, which also present a C-type lectin domain [25]. Based on a phylogenetic analysis, the C-type lectin domain itself is more related to non-venom lacticans (versican/aggrecan and neurocan protein; data not shown). Thus, it is more parsimonious to propose that this putative C-type lectin 1 sequence is involved in an endogenous function. Similar results were found for other toxin candidates, such as the Kunitz-type protease inhibitor 19, latrotoxin 27, latrotoxin 116 and the 'no-hits 93' and 'no-hits 96' sequences (no similar sequences on the Tox-Prot database).

The rejected venom genes mostly concern Kunitz-type proteases and latrotoxins. These protein families contain ankyrin and Kunitz-type domains, which are widespread among non-venom-related gene products. Consequently, they are good 'seeds' to retrieve many sequenced genes

through BLAST search analyses. Most of the new toxin candidates from these families contain such domains, but, as is the case in the above-mentioned examples, similarity to previously characterized toxins is restricted to a single domain.

Are there unequivocal criteria to label a given protein as a toxin?

Acceptance or rejection of a toxin candidate is quite complex, at the bioinformatic level as well as through benchwork analyses. From our point of view, unequivocal criteria do not exist. It is certain, however, that, when adopting a bioinformatics approach, prudent choice of a database resolves some issues. The dataset used as a query should include toxin as well as non-toxin-related genes. We will briefly touch upon the remaining issues. To start with, there is no reason to assume that toxin candidates evolved from identical paralogous copies in different taxa - that is, toxins of the same family might not cluster together within a monophyletic group. Moreover, we agree with Whittington and colleagues that, following a very recent recruitment of a common protein after a duplication event, the new toxin is hypothesized to be very similar to the original 'proof'. However, the application of a more conservative validation criterion leads to the rejection of most 'novel venom genes' presented in this study. Hence, we believe that the annotation process used by Whittington and colleagues is too permissive. Their tissue-expression criterion is only valid when performed on high-throughput-generated datasets and should also include statistically sound comparative arguments to validate candidates associated with a broad range of transcription levels. Finally, we agree with the authors that functional analyses have to be performed to confirm or refute the toxic activity of a given sequence. Nevertheless, *in vitro* and *in vivo* validation of *in silico* candidates will always need to be approached with the utmost caution. After all, any compound, depending on dosage, could potentially act as a toxin.

Camilla M Whittington, Anthony T Papenfuss, Katherine Belov and Wesley C Warren respond:

The correspondence by Terrat and Ducancel raises some valid points about the limitations of our annotation of the platypus venom gland transcriptome [8]. While we agree that our annotation methodology was permissive (and have stated as much in our manuscript, referring to the platypus venom genes as 'putative'), we believe that it represents the best approach, given the data and methods available at the time.

While Terrat and Ducancel do provide some interesting new analytical techniques, in particular refuting our annotation of a sarafotoxin-like sequence and a C-type lectin as putative venom genes, which we concede might be the case, it is worth noting that approximately 50% of members of families singled out as false positives by Terrat and Ducancel are not included in our list of likely venom genes passing our more stringent criteria. For our more permissive dataset (83 genes), our choice of screening out putative venom genes based on expression in 50% of the available non-venom platypus EST libraries represents an effort to strike a balance between avoiding false positives and excluding true venom genes from further analysis, given that we have previously shown that platypus venom genes can be expressed in non-venom tissues [26]. We used more stringent criteria to identify a subset of likely venom genes (33 genes) that were not expressed in any non-venom tissues.

In addition, Terrat and Ducancel have misunderstood parts of our annotation process. Their statements about our annotation using TBLASTN to search Tox-Prot proteins against the platypus cDNA database miss the point that we utilized BLAST searches of Tox-Prot sequences against the platypus genome. The authors also refer to the use of our expression data in a quantitative way, stating that more than one-third of the putative candidates match fewer than 50 Illumina reads. We do not believe that this is valid, owing to the fact that the library was normalized before sequencing. As evidence, we cite the platypus genome manuscript where, in the supplementary notes, a description of the library preparation method is found. In brief, the optimally cycled cDNA product is boiled and allowed to re-anneal, and, during this time, the high-copy molecules re-anneal, whereas the low-copy molecules maintain the single-stranded state, achieving normalization. This would account for the low copy number of potential venom genes. While it is possible to conclude, as we have, that high read counts in normalized libraries probably indicate high expression *in situ*, it is not valid to perform the reverse comparison and conclude that low read counts in normalized libraries indicate low expression *in situ*. Their argument that some of our putative toxins are probably non-toxins given that they are more related to non-venom proteins than to venom sequences is flawed as one might expect this given the degree of divergence between platypuses and other venomous species, and recruitment of specific domains might result in only partial homology to known toxins. We also disagree with the use of synteny data to infer non-venom function as we have previously shown that genes with venom function are conserved in a syntenic block with related non-toxin genes [27].

We clearly state in our article that our analysis is a first-pass analysis of the platypus venom transcriptome, and we discuss the limitations of our approach. While we agree with Terrat and Ducancel that our criteria were

permissive, we stand by the use of these permissive criteria, as we did not want to exclude any potential venom genes from further study. We discuss in our manuscript the fact that further research, including functional testing, is required before any of the putative toxins can be definitively classified as novel platypus venom toxins. This fact is also recognized by Terrat and Ducancel, who concede that they are unable to refute definitively our designation of putative platypus venom toxins without functional testing.

A first-pass analysis of the transcriptome is a necessary step towards identifying new venom toxins, and the results we obtained are valid with our methodology. As with many *in silico* analyses, a change in methodology will affect the results. This is inevitable with any re-analysis of genomic data, as bioinformatic methods evolve over time. We have thus undertaken follow-up studies to address some of the concerns raised in both articles. These include improved phylogenetic analyses, with an investigation of genes under positive selection [28], as well as sequencing of in-season and out-of-breeding-season venom glands in order to identify differentially expressed venom genes, which we combined with proteomics work [29].

In conclusion, although we agree that Terrat and Ducancel raise some valid points, we disagree with several of their arguments and have addressed other issues in our further research, some of which is already published.

Correspondence should be sent to Wesley C Warren: The Genome Institute, Washington University School of Medicine, Campus Box 8501, 4444 Forest Park Avenue, St Louis, MO 63108, USA. Email: wwarren@genome.wustl.edu

Abbreviations

bp, base pair; BLAST, basic local alignment search tool; EST, expressed sequence tag; ET, endothelin; ML, maximum likelihood ; SRTX, sarafotoxin sequence.

Author details

[1]Montréal University/Institute in Plant Biology, Montreal Botanical Garden, 4101 Sherbrooke East, Montreal, Québec, Canada, H1X 2B2. [2]Antibody Engineering for Health Laboratory (LIAS), CEA/iBiTecS/SPI/LIAS, 91191, Gif-sur-Yvette Cedex, France.

Published: 3 September 2013

References

1. Ruiming Z, Yibao M, Yawen H, Zhiyong D, Yingliang W, Zhijian C, Wenxin L: **Comparative venom gland transcriptome analysis of the scorpion *Lychas mucronatus* reveals intraspecific toxic gene diversity and new venomous components.** *BMC Genomics* 2010, **11:**452.
2. Cardoso KC, Da Silva MJ, Costa GGL, Torres TT, Del Bem LEV, Vidal RO, Menossi M, Hyslop S: **A transcriptomic analysis of gene expression in the venom gland of the snake *Bothrops alternatus* (urutu).** *BMC Genomics* 2010, **11:**605.
3. Chatrath ST, Chapeaurouge A, Lin Q, Lim TK, Dunstan N, Mirtschin P, Kumar PP, Kini RM: **Identification of novel proteins from the venom of a cryptic snake *Drysdalia coronoides* by a combined transcriptomics and proteomics approach.** *J Proteome Res* 2011, **10:**739-750.
4. Hu H, Bandyopadhyay PK, Olivera BM, Yandell M: **Characterization of the *Conus bullatus* genome and its venom-duct transcriptome.** *BMC Genomics* 2011, **12:**60.
5. Jiang Y, Li Y, Lee W, Xu X, Zhang Y, Zhao R, Zhang Y, Wang W: **Venom gland transcriptomes of two elapid snakes (*Bungarus multicinctus* and *Naja atra*) and evolution of toxin genes.** *BMC Genomics* 2011, **12:**1.
6. Morgenstern D, Rohde BH, King GF, Tal T, Sher D, Zlotkin E: **The tale of a resting gland: Transcriptome of a replete venom gland from the scorpion *Hottentotta judaicus*.** *Toxicon* 2011, **57:**695-703.
7. Rokyta DR, Wray KP, Lemmon AR, Lemmon EM, Caudle SB: **A high-throughput venom-gland transcriptome for the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*) and evidence for pervasive positive selection across toxin classes.** *Toxicon* 2011, **57:**657-671.
8. Whittington CM, Papenfuss AT, Locke DP, Mardis ER, Wilson RK, Abubucker S, Mitreva M, Wong ESW, Hsu AL, Kuchel PW, Belov K, Warren WC: **Novel venom gene discovery in the platypus.** *Genome Biol* 2010, **11:**R95.
9. de Plater G, Martin RL, Milburn PJ: **A pharmacological and biochemical investigation of the venom from the platypus (*Ornithorhynchus anatinus*).** *Toxicon* 1995, **33:**157-169.
10. Kordis D, Gubensek F: **Adaptive evolution of animal toxin multigene families.** *Gene* 2000, **261:**43-52.
11. Torres AM, Menz I, Alewood PF, Bansal P, Lahnstein J, Gallagher CH, Kuchel PW: **D-Amino acid residue in the C-type natriuretic peptide from the venom of the mammal, *Ornithorhynchus anatinus*, the Australian platypus.** *FEBS Lett* 2002, **524:**172-176.
12. Torres AM, Alewood D, Alewood PF, Gallagher CH, Kuchel PW: **Conformations of platypus venom C-type natriuretic peptide in aqueous solution and sodium dodecyl sulfate micelles.** *Toxicon* 2002, **40:**711-719.
13. Torres AM, Wang X, Fletcher JI, Alewood D, Alewood PF, Smith R, Simpson RJ, Nicholson GM, Sutherland SK, Gallagher CH, King GF, Kuchel PW: **Solution structure of a defensin-like peptide from platypus venom.** *Biochem J* 1999, **341:**785-794.
14. Torres AM, de Plater GM, Doverskog M, Birinyi-Strachan LC, Nicholson GM, Gallagher CH, Kuchel PW: **Defensin-like peptide-2 from platypus venom: member of a class of peptides with a distinct structural fold.** *Biochem J* 2000, **348:**649-656.
15. Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grützner F, Belov K, Miller W, Clarke L, Chinwalla AT, Yang SP, Heger A, Locke DP, Miethke P, Waters PD, Veyrunes F, Fulton L, Fulton B, Graves T, Wallis J, Puente XS, López-Otín C, Ordóñez GR, Eichler EE, Chen L, Cheng Z, Deakin JE, Alsop A, Thompson K, Kirby P, *et al.*: **Genome analysis of the platypus reveals unique signatures of evolution.** *Nature* 2008, **453:**175-183.
16. Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS, Renjifo C, de la Vega RCR: **The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms.** *Annu Rev Genomics Hum Genet* 2009, **10:**483-511.
17. Fry BG: **From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins.** *Genome Res* 2005, **15:**403-420.
18. Jungo F, Bairoch A: **Tox-Prot, the toxin protein annotation program of the Swiss-Prot protein knowledgebase.** *Toxicon* 2005, **45:**293-301.
19. Landan G, Bdolah A, Wollberg Z, Kochva E, Graur D: **Evolution of the sarafotoxin/endothelin superfamily of proteins.** *Toxicon* 1991, **29:**237-244.
20. Kochva E, Viljoen CC, Botes DP: **A new type of toxin in the venom of snakes of the genus *Atractaspis* (Atractaspidinae).** *Toxicon* 1982, **20:**581-592.
21. Kloog Y, Ambar I, Sokolovsky M, Kochva E, Wollberg Z, Bdolah A: **Sarafotoxin, a novel vasoconstrictor peptide: phosphoinositide hydrolysis in rat heart and brain.** *Science* 1988, **242:**268-270.
22. Ducancel F: **Endothelin-like peptides.** *Cell Mol Life Sci* 2005, **62:**2828-2839.
23. Ducancel F: **The sarafotoxins.** *Toxicon* 2002, **40:**1541-1545.
24. Braasch I, Volff J-N, Schartl M: **The endothelin system: evolution of vertebrate-specific ligand–receptor interactions by three rounds of genome duplication.** *Mol Biol Evol* 2009, **26:**783-799.
25. Yamaguchi Y: **Lecticans: organizers of the brain extracellular matrix.** *Cell Mol Life Sci* 2000, **57:**276-289.
26. Whittington CM, Belov K: **Platypus venom genes expressed in non-venom tissues.** *Aus J Zool* 2009, **57:**199-202.
27. Whittington CM, Papenfuss AT, Bansal P, Torres AM, Wong ES, Deakin JE, Graves T, Alsop A, Schatzkamer K, Ponting CP, Temple-Smith P, Warren WC, Kuchel PW, Belov K: **Defensins and the convergent evolution of platypus and reptile venom genes.** *Genome Res* 2008, **18:**986-994.

28. Wong ES, Papenfuss AT, Whittington CM, Warren WC, Belov K: **A limited role for gene duplications in the evolution of platypus venom.** *Mol Biol Evol* 2012, **29:**167-177.
29. Wong ES, Morgenstern D, Mofiz E, Gombert S, Morris KM, Temple-Smith P, Renfree MB, Whittington CM, King GF, Warren WC, Papenfuss AT, Belov K: **Proteomics and deep sequencing comparison of seasonally active venom glands in the platypus reveals novel venom peptides and distinct expression profiles.** *Mol Cell Proteomics* 2012, **11:**1354-1364.

doi:10.1186/gb-2013-14-9-406
Cite this article as: Terrat Y, Ducancel F: Are there unequivocal criteria to label a given protein as a toxin? Permissive versus conservative annotation processes. *Genome Biology* 2013, **14:**406.