

# RH: SPECIES DELIMITATION IN HARLEQUIN POISON FROGS

Ecology, molecules and colour: Multivariate species delimitation and conservation of Harlequin poison frogs

Andres Posso-Terranova<sup>1-2\*</sup>, Jose A. Andres<sup>1</sup>

<sup>1</sup>*Department of Biology. University of Saskatchewan. 112 Science Pl. Saskatoon. SK. Canada*

<sup>2</sup>*Universidad Nacional de Colombia sede Palmira. A.A. 237. Palmira. Colombia*

*\*Correspondence to be sent to: 112 Science Pl. Saskatoon. SK. Canada. S7N5E2; E-mail: andres.posso@usask.ca*

## ABSTRACT

We propose a iterative protocol for delimiting species under the generalized lineage concept (GLC) based on the multivariate clustering of morphological, ecological, and genetic data. Our rationale is that the resulting groups should correspond to evolutionarily independent metapopulation lineages because they reflect the common signal of different secondary defining properties (ecological and genetic distinctiveness, morphological diagnosability, etc.), implying the existence of barriers preventing or limiting gene exchange. We applied this method to study a group of highly endangered poison frogs, the *Oophaga histrionica* complex. In our study case, we use next

generation targeted amplicon sequencing to obtain a robust genetic dataset that we then combined with patterns of morphological and ecological divergence. Our analyses revealed the existence of at least five different species in the *histrionica* complex (three of them new to science) occurring in very small isolated populations outside any protected areas. More broadly, our study exemplifies how transcriptome-based reduction of genomic complexity and multivariate statistical techniques can be integrated to successfully identify species and their boundaries.

Keywords: Integrative taxonomy, Gaussian clustering, geographic information systems, total-evidence, phylogenetics, morphology.

#### *IN MEMORIAM*

*“I propose that each species has a distinctive life history, which include a series of stages that correspond to some of the named species concepts”*

*Richard G. Harrison*

*1945-2016*

#### INTRODUCTION

The delimitation of species boundaries is often ambiguous and we continuously debate the taxonomic status of many threatened organisms. In many cases, conflicting species concepts based on operational properties (e.g. reproductive isolation, Mayr 1942; ecological distinctiveness, Van Valen 1976; morphological diagnosability, Helbig et al.

2002; genealogical exclusivity, Baum and Donoghue 1995; etc) have led to confusing situations in which particular populations are considered to be species –or not– under alternative species concepts. For instance, the Florida panther, thought to be a unique (sub)species based on its morphological and ecological distinctiveness (Hedrick 1995) is genetically considered an allopatric population of the North American cougar (*Puma c. cougar*; Culver et al. 2000).

While the adoption by many taxonomists of the General Lineage Concept (GLC) (de Queiroz 2005a, b) has helped to ameliorate some taxonomic controversies, the delimitation of species is still far from trivial. Distinguishing between local adaptation, isolation by distance and species boundaries can be difficult in highly mobile organisms (Brown et al. 2007). Similarly, inter- and intra-specific variation can also be very difficult to differentiate in those taxa characterized by allopatric lineages or low levels of morphological diversity (Barley et al. 2013). As a result, in the last decade we have witnessed the rapid succession of methodological approaches for identifying species boundaries. Thanks to the advent of sophisticated genomics tools and increasing computational power, new methods have been developed to delimit species based exclusively on genetic data (Heled and Drummond 2010; Leaché and Fujita 2010; Huelsenbeck et al. 2011; Grummer et al. 2013; Leaché et al. 2014; Ruane et al. 2014; Zhang et al. 2014; Yang 2015). Similarly, powerful statistical approaches have been proposed to use gaps in morphological variation or ecological discontinuities as the criteria for species delimitation (e.g. Varcarel and Vargas 2010; Zapata and Jimenez 2012). Despite the widespread use of the above approaches, it is arguable whether species delimitation should be based on just a single line of evidence (Carstens et al.

2013) Instead, an integrative approach should be able to provide the best inferences about species boundaries (Padial and De la Riva 2006; Padial and De La Riva 2010; Padial et al. 2010).

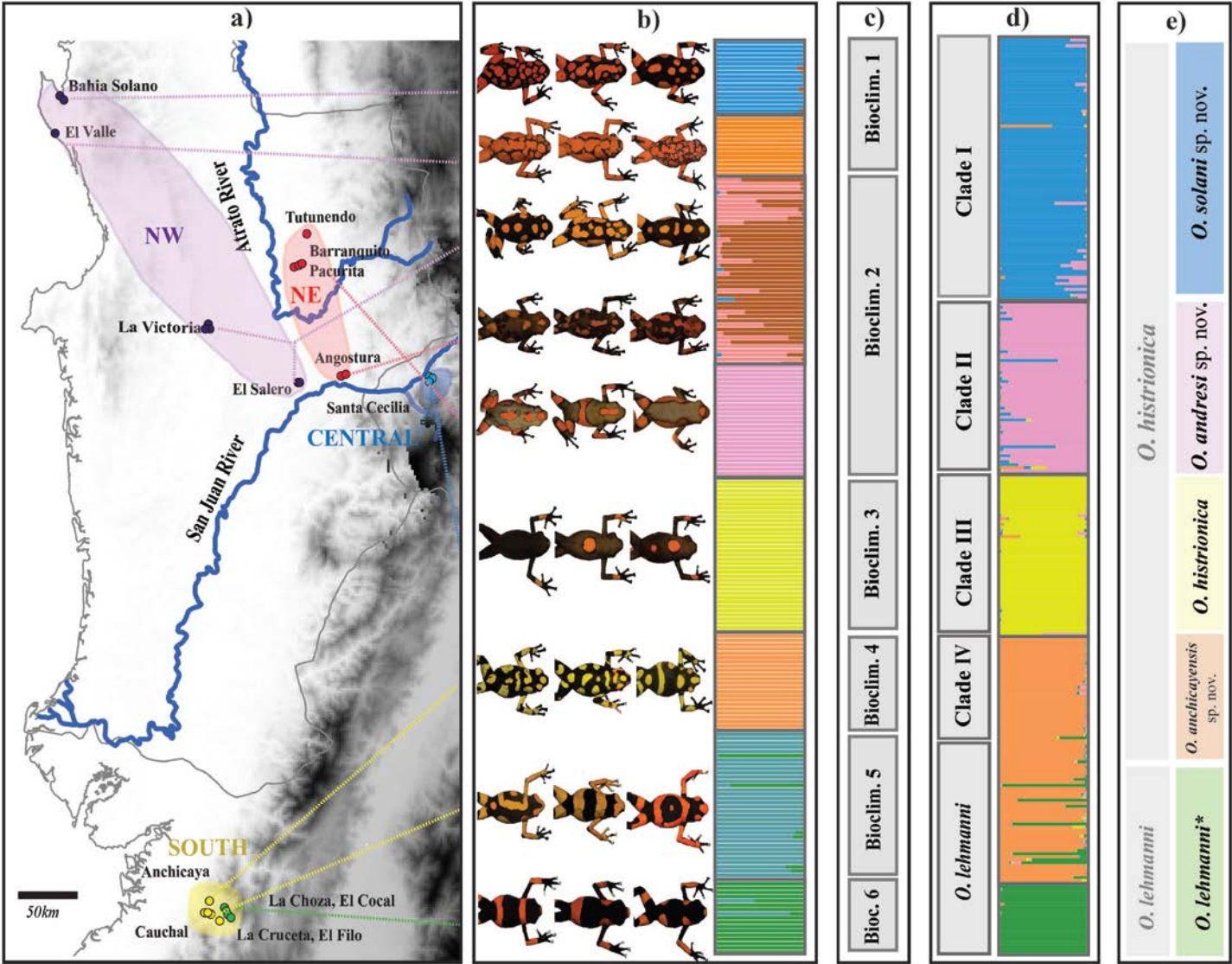
Integrative taxonomy (Dayrat 2005), combines several lines of evidence to support or reject the hypothesis that populations are independently evolving lineages (*i.e.* species; Padial and De la Riba 2009). Under this framework, different lines of evidence can either disagree over the number of species, or agree over the number of species but disagree over the assignment of individuals into species. So, how much agreement must different taxonomic characters show to consider a population (or a group of populations) a separate species? A prevalent strategy has been deriving species hypotheses by consensus, congruence or cumulation (*sensu* Padial et al. 2010) of separately analysed datasets (e.g. Cardoso et al. 2009; Vieites et al. 2009 ; DeJaco 2012; Yeates et al. 2011). One limitation of these methods is that they cannot extract the common signal that may emerge from the simultaneous analysis of all taxonomic characters. A second limitation is that the resulting species boundaries are likely to be affected by order in which different characters (criteria) are introduced in the analyses. These methodological shortcomings, common in many taxonomic studies, need to be addressed.

One of our main objectives is to provide a total-evidence procedure to analyse morphological, ecological, and genetic variation that can serve to assign specimens into diagnosable species. For this purpose, we performed a comprehensive study of the Harlequin poison frog (*Oophaga histrionica* Myers and Daly 1975) species complex. We have two reasons to focus on this system. First, polytypic taxa present important challenges for species delimitation. Second, specimens of these frogs are actively

sought in the wildlife pet trade. Intense illegal trafficking and habitat loss associated with the proliferation of illegal crops have pushed these frogs to near extinction. Thus, it is critically important to delimit conservation units (at or below the species level) that can guide the implementation of the appropriate taxon-based conservation legislature (e.g. CITES trade treaty; Lynch and Arroyo 2009).

Harlequin poison frogs inhabit the lowland Pacific rainforests of the Colombian Choco. Individuals from different populations can either be striped, spotted or relatively homogenous, and their colours range from bright red and yellow to dull green and blue (Figure 1). This phenotypic variability combined with their patchy distribution and strong population structure suggests that *O. histrionica* may, in fact, be a complex of several different species. However, early taxonomic studies (Berthold 1846; Myers and Daly 1976) and preliminary genetic data from geographically scattered individuals (Grant et al. 2006; Brown et al. 2011) have failed to recognize the morphological diversity and potential ecological divergence in this group (Lötters 1992; Lötters et al. 1999; Medina et al. 2013; Posso-Terranova and Andres 2016). To date, only two nominal species are recognized: *O. histrionica* (*sensu lato*) and *O. lehmanni* (Myers and Daly 1976). While the former taxon has a widespread distribution, *O. lehmanni* is a micro-endemic species restricted to a few small submontane wards of the central Choco (Figure 1a).

Figure 1.



In this study, we used next generation sequencing (NGS) to obtain haplotypes that are then analysed using multivariate clustering approaches. This approach, similar to the species delimiting method of Hausdorf and Hennig (2010) and the species assignment method of Edwards and Knowles (2014), combines transcriptome polymorphisms with mtDNA and microsatellites variation to generate a comprehensive genetic dataset that is then combined with patterns of morphological and ecological divergence. The resulting common signal is used to establish species delimitations. To evaluate the power of our approach to identify independent evolving lineages (i.e. species), we compared its results to those obtained by other “traditional” integrative methods (see Edwards and Knowles 2014). Our analyses revealed the existence of at least five different species in the *histrionica* complex (three of them new to science) occurring in very small isolated populations outside any protected areas. More broadly, our study exemplifies how targeted amplicon sequencing and multivariate statistical techniques can be integrated to successfully identify species and their boundaries.

## MATERIALS AND METHODS

### *Specimen sampling*

We sampled 310 individuals in 15 different geographic locations across the known geographic distribution of the *O. histrionica* complex (Myers and Daly 1976; Lötters 1992; Lötters et al. 1999) (Fig. 1a). For each of the sampled individuals, we recorded its geographic location using a handheld GPS unit. High-resolution images and tissue samples were then collected as follows: raw digital pictures (Nikkon Electronic

Format: NEF) were taken using a Nikon D610 digital camera against a Spectralon grey reflectance standard (Labsphere, Congleton, UK). Images were then linearized with respect to light intensity (Stevens et al. 2007), and stored as TIFF images (lossless compression) to preserve fidelity for further analysis. Tissue samples were obtained by toe-clipping (Phillott et al. 2011). Toe-clips were conserved in 95% ethanol until DNA extraction. Genomic DNA was extracted using the MasterPure™ DNA purification kit (Epicentre, Madison, WI, USA) following the manufacturer's recommendations.

All samples were obtained in compliance with the Ethic Committee of the Universidad Nacional de Colombia (sesión del 22 de Julio, Acta No.03, July 22<sup>nd</sup> 2015), under the research permits granted by Autoridad Nacional de Licencias Ambientales ANLA, Resolución 0255 del 12 de Marzo de 2014 and Resolución 1482 del 20 de Noviembre de 2015.

### *Morphological clustering*

Species are commonly delimited on the basis of gaps in patterns of morphological variation (Zapata and Jiménez 2012). Hence, we estimated the number of distinct morphological clusters (species) using hierarchical and discriminant grouping methods. Dorsal and ventral pictures of each specimen were scaled using ImageJ v. 1.48 (Sheffield 2007) and a known metric reference. For each specimen, we then measured a total of 49 variables ( $n_{\text{dorsal}} = 25$ ;  $n_{\text{ventral}} = 24$ ) related to morphology, pattern of coloration and coloration intensity (Table S1 in Appendix 1). All variables were standardized prior to subsequent analyses. First, we performed a hierarchical cluster analysis as implemented in PAST v. 3.08 with bootstrap- and PERMANOVA-

based significance tests (Hammer et al. 2001). In addition, we also performed a discriminant analysis of principal components (DAPC) as implemented in the R package *Adegenet* v 2.0.0. (Jombart et al. 2010). This method has the advantage that allows for a probabilistic assignment of specimens to each group. To determine the optimal number of morphological clusters ( $K_M$ ), we ran the sequential  $K$ -means (*find.clusters*,  $K=1-19$ ) algorithm and identified one showing the lowest Bayesian Information Criterion (BIC) (Jombart et al. 2010). To avoid overfitting, the individual assignment of specimens was based on 11 axes (*optim.a.score*) representing 86% of the variation in the dataset.

### *Ecological clustering*

To identify ecological discontinuities and determine the number of environmental clusters ( $K_E$ ) we performed a non-metric multi-dimensional scaling (NMDS) analysis as proposed by Edwards and Knowles (2014). Briefly, for each sampled individual, we extracted high resolution (30 arc-seconds) environmental data from 19 bioclimatic variables ([www.worldclim.org](http://www.worldclim.org)) along with altitude data (STRM90 <http://srtm.csi.cgiar.org/>) (Hijmans et al. 2005). Bioclimatic data were standardized and reduced to PC-scores along 4 axes (explaining > 90% of the variation) using the *dudi.pca* function in the Vegan R-package (Oksanen et al. 2013). Next, geographic distances among sampled specimens (function: *Distance*, Diva-Gis v. 7.5; <http://www.diva-gis.org/>) (Hijmans et al. 2001) were converted to continuous variables (function: *pcnm*, Vegan R-package) and combined with the bioclimatic scores before generating the final Gower-distance environmental dataset. Then, we used non-metric

multi-dimensional scaling (NMDS) and one-way PERMANOVA to estimate  $K_E$ .

### *Genetic clustering*

*Mitochondrial DNA (mtDNA)*: Mitochondrial genetic subdivision is often considered a leading criterion of interspecific differentiation. To determine the number of mitochondrial clusters ( $K_{MT}$ ), we applied phylogenetic and multivariate procedures. For a subsample of 107 individuals (including *O. sylvatica* (n=6) and *O. occultator* (n=4) as “*bona fide*” control species) we amplified and Sanger-sequenced a 523 bp of the COI gene (see Hauswaldt et al. 2011; Brusa et al. 2013; Medina et al. 2013 for detailed molecular protocols). Aligned sequences (*Clustal-W*) (Thompson et al. 1994) were used to estimate the model of nucleotide substitution using the corrected Akaike Information Criterion as implemented in jModelTest v2.1.4 (Posada 2008). MtDNA phylogenetic trees were constructed under different Bayesian and maximum likelihood optimality criteria and a TCS haplotype network was also plotted (Appendix 2). Individual pairwise genetic distances were calculated using MEGA6 (Tamura et al. 2013), and a Principal Coordinate analysis (PCoA) was carried out upon the similarity matrix (Reeves and Richards 2011). Finally, to delimit independent evolving lineages we also used a generalized mixed- Yule-coalescent (GMYC) method capable of fitting within- and between-species branching models (Zhang et al. 2013) (Appendix 2).

*Microsatellites and transcriptome-derived haplotypes*: Twelve microsatellite loci were isolated from a di-, tri- and tetra-mer enriched library constructed from a single *O. histrionica* specimen using pyrosequencing technology (454GS-FLX, Roche Applied Science) as described in Andres and Bogdanowicz (2011) (Table S2 in

Appendix 1). All individuals ( $n = 310$ ,  $n_{\text{localities}} = 15$ ) were genotyped using multiplex-PCR and fluorescently-labelled primers as described in Appendix 2.

To further identify polymorphic nuclear loci for multiplex PCR we used a transcriptome-based approach. *De novo* skin transcriptome sequencing and assembly (Table S3 in Appendix 1) from a single *O. lehmanni* individual was carried out as described in Appendix 2. The resulting transcriptome was utilized to design primers for 50 nuclear markers (250-350 bp), each of them representing a single randomly selected unigene. To reduce ascertainment bias, we selected eight individuals from a broad geographic range to screen each locus for robust PCR using genomic DNA as template. After stringent selection, 20 remaining loci were used to genotype all 310 specimens in two multiplex PCR groups using the QIAGEN Multiplex PCR Kit. Individual specimens and loci were barcoded using Illumina<sup>TM</sup> S5 and N7 Nextera primers following the manufacturer's conditions. Amplicons were paired-end sequenced (2X 250 bp) on a Illumina MiSeq at the Cornell University's BioResource Center.

Post-sequencing data processing required extracting reads from the Miseq run and assigning them to the appropriate specimen and locus with a custom *Perl* script that discarded low quality reads, trimmed adapters, overlapped paired-reads, identified reads corresponding to each locus, collapsed identical reads for each individual, and identified the two most frequent haplotypes for each individual at all loci (quality score: Q12, minimum overlap: 20, mismatch rate: 0.05). At any given locus, an individual was scored as heterozygous if the minor allele frequency was  $\geq 20\%$ . Individuals that fail to amplify in more than 80% of the loci were excluded from the dataset, and any sample

by locus cells with fewer than 10 total reads were considered missing data.

The number of genetic clusters based on microsatellite loci ( $K_{SAT}$ ), transcriptome-derived haplotypes ( $K_{TRN}$ ), and both datasets combined ( $K_{NUC}$ ) was estimated using three allele frequency-based methods (STRUCTURE, STRUCTURAMA, and PCoA). While the first two methods estimate the number of clusters by minimizing the deviations from Hardy-Weinberg equilibrium, the latter one finds clusters based on the Nei's distance (Pritchard et al. 2000; Hubisz et al. 2009; Peakall and Smouse 2012). Detailed information on the different models, parameters and run conditions is described in Appendix 2.

### *Total-evidence species delimitation*

The multivariate species delimitation approach described in this paper assumes the GLC species concept, and is based on the conceptual and methodological frameworks of Hausdorf and Henning (2010) and Edwards and Knowles (2014). First, a provisional species hypothesis is formulated based on a small balanced geographical sampling of the focal system and the assumption that species can be defined as distinguishable groups (i.e. multivariate clusters) of individuals with similar phenotypic, genotypic and ecological characteristics (see Discussion). Second, we weighted the strength of the evidence supporting the proposed species hypothesis by evaluating our ability to detect and assign new specimens to the provisional species. Because this multivariate approach does not necessarily take into account historical patterns of diversification, we compared our results to those obtained by using a Generalized Mixed Yule Coalescent (GMYC) model capable of fitting within- and between-species

branching models (Zhang et al. 2013).

Accordingly, for a first sample of individuals (n=86, 4-12 ind./population), we generated a new set of standardized orthogonal variables containing the same information as the original datasets using multivariate methods (morphology: PCA, environmental and genetic: PCoA). We then calculated distance matrices for each data-type (morphology and environmental: Gower-distances; mtDNA: TN93-distance; nuclear loci: Nei's distance) and standardized them using NMDS (*isoMDS*, MASS R-package). To determine the provisional number of species we followed Edwards and Knowles (2014) and retained four dimensions for each data-type and determined the number of Gaussian clusters present in the subsample using the R-package *mclust* and the Bayesian Information Criterion (Fraley and Raftery 2002). Additionally, we also performed hierarchical clustering (average method) with 1000 bootstrap repetitions as implemented in PAST v. 3.08 (Hammer et al. 2001).

To assess the strength of our provisional species delimitation hypothesis, we conducted an identical series of analyses on an independent, much larger dataset (n=214 individuals) and assessed our ability to recover the same clusters (species) with high individual assignment probabilities. All stress values for NMDS analyses were below 10%, with an absolute maximum of 6%.

### *Nomenclature acts*

All new species names resulting from this study conform to the requirements of the amended International Code of Zoological Nomenclature (ICZN) and they have been registered in ZooBank (<http://Zoobank.org/>). All Relevant information as well as the Life Science Identifiers (LSIDs) can be found under the account [DACAA96D-](#)

[2E69-4718-83DD-EB907A83BC95.](#)

## RESULTS

### *Morphological and environmental clustering*

The morphological hierarchical clustering applied over individuals of the *O. histrionica* complex revealed the presence of nine clearly defined morphological clusters with high node support (>82%; PERMANOVA Bonferroni-adjusted  $P < 0.003$ ; Fig. 1b and Fig. S1 in Appendix 3). The optimal number of morphologic clusters ( $K_M$ ) as defined by DAPC (BIC) was also nine. As a result, both clustering methods were highly congruent in their individual assignments (Fig. 1a). Posterior assignment probabilities (PP) for  $K_M=9$  indicated that 93% of individuals were correctly assigned to their morphologic clusters ( $PP > 0.9$ ); 4.7% were considered as ambiguously assigned ( $PP < 0.9$ ) and only 2.3% were incorrectly assigned (*i.e.* assigned to a different cluster with  $PP > 0.9$ ).

Gaussian clustering based on a NMDS analysis of bioclimatic distances partitioned the sampled geographic range of the *O. histrionica* complex into six distinct ecological clusters ( $K_E=6$ , stress=0.04) with statistical significance after Bonferroni correction ( $P < 0.0001$ ; Fig. 1c): while *O. histrionica* (*sensu lato*) encompasses four potential bioclimatic niches, putative *O. lehmanni* specimens form two different clusters that correspond to two different morphological clusters (Fig. 1c and Fig. S2 in Appendix 3).

# *Genetic clustering*

Bayesian (BI) and Maximum Likelihood (ML) COI gene genealogies showed identical topologies and revealed the presence of seven distinct mtDNA clades representing *O. sylvatica*, *O. occultator*, *O. lehmanni* (defined by phenotype and sampling location), and 4 monophyletic, geographically structured clades (I-IV) within *O. histrionica* (*sensu lato*, Fig. 1d and Fig. S3 in Appendix 3): Clade I (NW-populations), Clade II (NE-populations), Clade III (Central populations), and Clade IV (South populations) –See also haplotype network provided in Figure S4 in Appendix 3 –. As expected, on average, the *p*-distances between the “*bona fide* species” (*O. occultator* and *O. sylvatica*), and the *O. histrionica* complex clades were greater (range: 4-6%) than those observed among the four *histrionica* clades (range: 2.2%-2.9%). The single-locus species delimitation analysis by GMYC strongly supported the hypothesis that each of these mtDNA clades represent an independent species (PS>0.96 in all cases).

Clustering analyses of the *O. histrionica* complex based on the two nuclear datasets yield consistent partitions that, in turn, were very similar to that based on mtDNA. Although the PCoA results were mostly concordant, the microsatellite dataset supported a five-cluster partition ( $K_{SAT}= 5$ , Fig. S5a in Appendix 3) while transcriptome-derived dataset supported six clusters ( $K_{TRN}= 6$ , Fig. S5b in Appendix 3). The source of discordance is the adscription of the individuals from the localities of La Choza and El Cocal. The analysis of the transcriptome-derived haplotypes suggested that these populations represent an intermediate cluster between the Southern populations of *O. histrionica* (*sensu lato*) -Anchicaya and El Cauchal- and the *O.*

*lehmanni* populations of La Cruceta and El Filo (Fig. 1a). In contrast, the microsatellite-based analysis clustered these individuals with either the Southern *O. histrionica* or the *O. lehmanni* cluster. These results are consistent with La Choza and El Cocal being admixed populations. As expected under this admixture hypothesis, La Choza-Cocal specimens have intermediate phenotypes between those of *O. lehmanni* and Southern *O. histrionica* (Fig. 1b), and either Clade IV or *O. lehmanni* mtDNA haplotypes (Fig. 1d).

Bayesian clustering (STRUCTURE) analyses for the combined data set (microsatellites + transcriptome haplotypes) showed that the estimated log probability [ $\ln P(D)$ ] plateaued at  $K_{\text{NUC}} = 6$  with the second-order rate of change of  $K$  ( $\Delta K$ ) reaching its maximum at  $K_{\text{NUC}} = 5$ , suggested the existence of five (potentially six) distinct genetic clusters. Assuming  $K_{\text{NUC}} = 5$ , the first cluster grouped all *O. lehmanni* specimens, while the other four clusters corresponded to the mtDNA clades I-IV (Fig. 1d). There were 37 individuals exhibiting admixed ancestry ( $q = 0.1-0.9$ ), and 73% of them were found in four populations (Angostura and El Salero  $n = 14$ , Clade I  $\times$  Clade II; La Choza and El Cocal,  $n = 13$ , Clade IV  $\times$  *O. lehmanni*). The geographic location of these admixed populations (Figure 1a) suggests that these localities represent two different contact zones, one between the NW and the NE demes, and the second one between the South and the *O. lehmanni* ones. The results of the STRUCTURAMA analyses also supported the partition of  $K_{\text{NUC}} = 5$  with a posterior probability of 74% (Table 1).

Table 1. Results from STRUCTURAMA analyses at a range of priors for shape and scale using the *O. histrionica* species complex genetic dataset from microsatellite

(n=6) and transcriptome-based markers (n=13). Optimal number of inferred populations is shown in light grey.

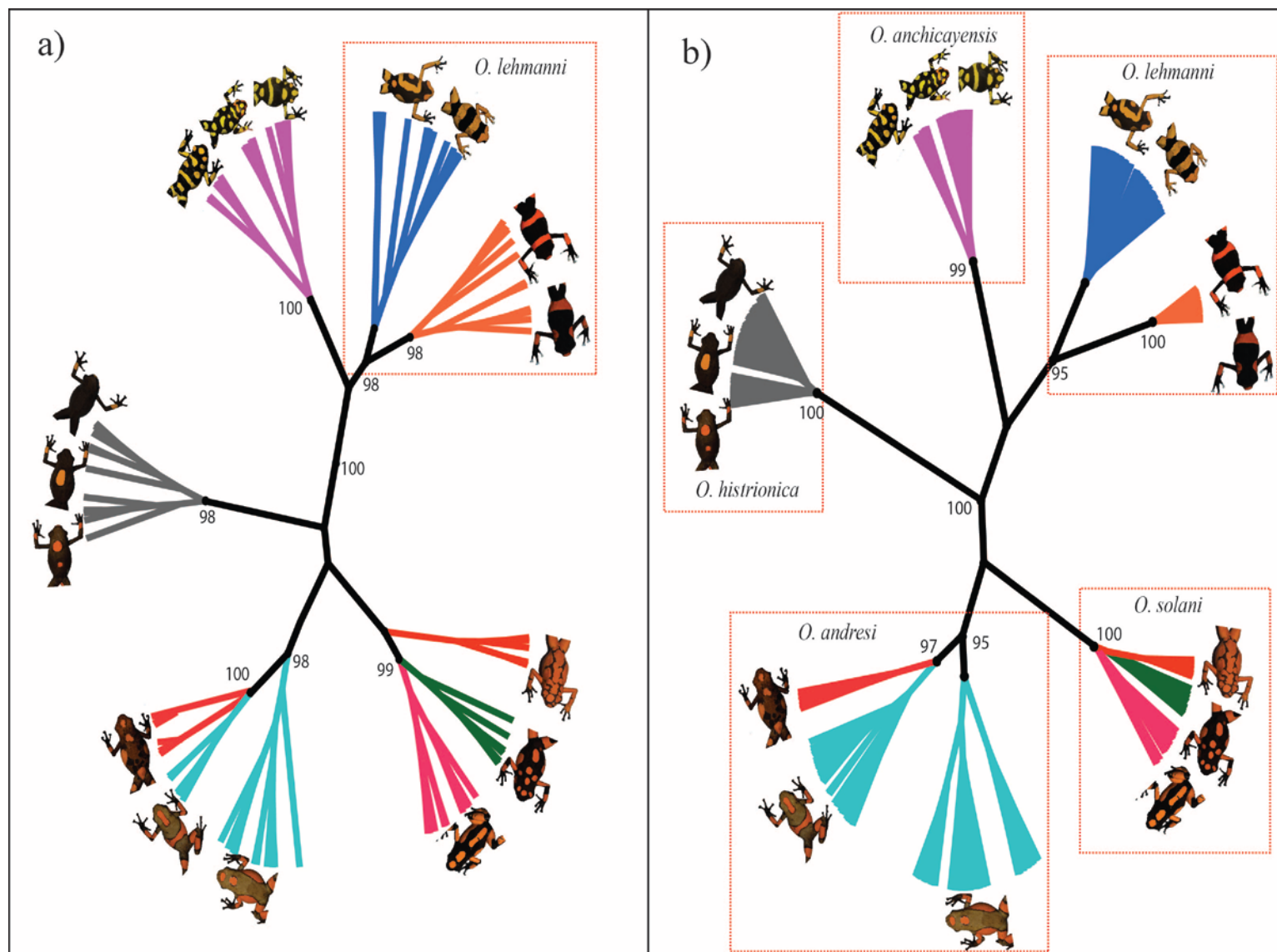
Priors used in STRUCTURAMA analyses						
Number of populations inferred	Pr [K =i   X]	Pr [K =i   X]	Pr [K =i   X]	Pr [K =i   X]	Pr [K =i   X]	Pr [K =i   X]
	Shape = 2.5	Shape = 2.5	Shape = 5	Shape = 10	Shape = 10	Shape = 1
	Scale = 0.52	Scale = 1	Scale = 2	Scale = 1	Scale = 0.1	Scale = 1
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0.36	0.39	0.34	0.19	0.25	0.48
5	0.6	0.58	0.62	0.72	0.74	0.5
6	0.04	0.03	0.04	0.09	0	0.02
7	0	0	0	0	0	0

### *Species delimitation by integrative analysis*

To first delimit the number of species, we used an integrated dataset composed of 12 multivariate dimensions (see Methods) and two different clustering methods: first, we used a model-based Gaussian clustering. The best model corresponded to an equal-covariance and diagonal distribution model (VEI) (Fraley and Raftery 2002) with the BIC plateaued at K= 5 (Fig. S6 in Appendix 3). Bootstrap likelihood ratio tests (*MclustBootstrap*, n= 999) for K = 1-7 supported the existence of five lineages representing *O. lehmanni* (including the genetically admixed populations of La Choza –

El Cocal), and the S, Central, NE, and NW lineages of “*histrionica*”. Such taxonomic partitioning resulted in 100% assigned individuals ( $PP \geq 0.95$ ). The second-best model ( $K=6$ ) split *O. lehmanni* into ‘pure’ (El Filo –La Cruceta) and admixed (La Choza –El Cocal) populations. Secondly, our hierarchical clustering approach was fully concordant with the Gaussian approach and showed five highly supported clusters ( $BS > 95\%$ ) with identical composition of specimens (Fig. 2a). Given these results we subdivided the *histrionica* complex into five provisional species. To test the robustness of these provisional species we performed a second round of clustering analyses in an independent set of specimens ( $n= 214$ ) and  $K= 5$ . Under such delimitation, all individuals were correctly assigned to the expected lineage with  $PP \geq 0.95$  (Fig. 2b). Clustering analyses excluding mtDNA yield identical assignment results.

389



## DISCUSSION

### *Conceptual basis*

Edwards and Knowles (2014) developed a statistical approach to evaluate hypotheses of species limits based on multivariate analyses and Gaussian clustering of total evidence (morphological, ecological and genetic) datasets. Here, we extend their approach and propose that the iterative implementation of such methodological framework can be used to delimit species under the GLC without any *a priori* species boundaries. Our rationale here is that these multivariate clusters should correspond to evolutionarily independent metapopulation lineages because they reflect the common signal of different secondary defining properties (ecological and genetic distinctiveness, morphological diagnosability, etc.), implying the existence of barriers preventing or limiting gene exchange. It is important to recognize that because of the hierarchical nature of evolution, we expect clustering to occur at many levels. That is, we should observe a continuum of increasing degree of cluster overlap as the levels of genetic, morphological and ecological divergence decreases between species to variation across populations. Thus, depending on the focal system, the inferred clusters might represent inter-population variation, local host races, subspecies, species, etc. So, where along the divergence continuum separate lineages (clusters) should be recognized as distinct species?

To try to answer this question, one may rely on “arbitrary” statistical thresholds. For example, it has been argued that 70% correct multivariate assignment is a reliable threshold for species delimitation in complex groups (Valcárcel and Vargas 2010). However, as noted by Coyne and Orr (2004, pg. 449) species do not have a greater “objective” reality than lower or higher taxa and we do not see a fundamental distinction between species and any other grouping

level. A better alternative is to contrast the multivariate clusters with those resulting from the application of multispecies coalescent methods that infer species status from a genealogical perspective.

In the case of the *histrionica* complex studied here, the GMYC analysis showed strong correspondence between the inferred multivariate clusters and the branching pattern of the mtDNA species tree. The only source of discordance was that the integrative analysis grouped a few *O. lehmanni* specimens together with individuals from the ‘*histrionica*’ mtDNA cluster IV (Fig. S3 and S4 in Appendix 3). Multivariate and Bayesian analyses showed that all ‘misclassified’ individuals are of admixed origin (Fig. 1b and 1d), a result that is concordant with previous studies suggesting hybridization between *O. histrionica* (*sensu lato*) and *O. lehmanni* (Medina et al. 2013).

Despite its limitations, we believe that this approach has some practical advantages over other delimitation methods often reliant on genetic data. First, assuming that there is enough variation, this method provides a combination of intrinsic characters (attributes) that can be used to recognize and describe the newly discovered species. Second, because this approach relies in the cumulative effect of independent characters it is likely to be useful to unravel species in early differentiation stages where varying levels of differentiation across genetic, phenotypic and ecological axes are expected (Padial and De La Riva 2010; Padial et al. 2010). Third, this method can be useful for analysing large datasets. Most coalescent based delimitation methods present serious computational challenges when analysing significant numbers of loci, putative species and/or individuals. In contrast, Gaussian clustering requires little computational power and can be easily scalable to include numerous genetic loci as well as large numbers of specimens.

# *Taxonomic implications*

A key question in our study was to determine whether the extensive phenotypic diversity within the *O. histrionica* complex reflects the existence of several independent taxonomic units (*i.e.* species) or just one (possibly two) highly polytypic species (Myers and Daly 1976; Grant et al. 2006). The application of our integrative multivariate procedure confirmed the taxonomic status of *O. lehmanni* and provided statistical support for the existence of five distinct species of *Oophaga*, which we propose should be formally recognized.

Though the specific status of *O. lehmanni* has casted doubt among taxonomist during the last decades (Lötters 1992; Lötters et al. 1999; Medina et al. 2013), analyses indicated a considerable amount of genetic, ecological and morphological differentiation between the two current nominal species *O. lehmanni* and *O. histrionica* (Myers and Daly, 1976; Fig. 1e). Thus, we recommend maintaining the taxonomic status of *O. lehmanni*. Our study also provides further support for existence of a geographically isolated hybrid lineage (*i.e.* cluster) of *O. lehmanni* (Medina et al. 2013). Overall, the pattern of genetic additivity of parental marker alleles (Fig. 1d), along with ecological (Fig. 1c) and morphological separation (Fig. 1b) suggests that hybridization between these two lineages may have resulted in the successful establishment of an ecologically distinct lineage. Further genetic studies are required to test this hypothesis.

We also propose the existence of four species within *O. histrionica* (*sensu lato*). These species correspond to each of the four ‘*histrionica*’ clusters revealed by our integrative analyses.

Here, we provide the proposed names and descriptions of the newly delimited species. Although detailed morphological descriptions have been previously proposed (Berthold 1846; Silverstone 1975; Myers and Daly 1976; Daly et al. 1978; Lötters 1992; Lötters et al. 1999; Grant et al. 2006), we have included a summarized description of each species with respect to body measurements and a synthesized description of coloration patterns based on our multivariate morphologic analysis. Distributional ranges are based on our own sampling.

We would like to highlight that if we accept species as evolutionary units, we cannot expect or demand immutable species hypothesis, but only the most corroborated ones (Padial and De La Riva 2010). Several lines of evidence and different methodological approaches support the species described in this paper. Therefore, they can be regarded as having high potential of further corroboration. However, the inclusion of other lines of evidence (e.g. acoustic, behavioral, etc.) over an extended geography may result in new, revised hypotheses.

***Oophaga solani* sp. nov.** [Mitochondrial clade I]

We designate this species as *O. solani* sp. nov. (syn. *O. histrionica*) based on the mtDNA (PS=1) and nDNA species delimitation support and the general conservative agreement of the integrative analysis (BS=99%). We named this new species after the municipality of Bahia Solano, department of Choco, Colombia where we started part of our field work and found the first populations of this species. Museum specimens AMNH 86979-86985 correspond to this species (Myers and Daly 1976).

**Range:** Based on our own sampling plus data collected from different authors (Silverstone 1975; Myers and Daly 1976; Mendez-Narvaez and Amezcuita 2014), this species inhabits the coastal lowlands at the west banks of the San Juan and Atrato rivers which, due

their width, presumably represent a major barrier to gene flow between the eastern and western *Oophaga* species (Figure 1a). Populations of this species are distributed in a relatively wider geographic extension of dense rainforest with known populations from Mecana, Bahia Solano (Lat. 6.29640, Lon.-77.37869) and Serranía del Baudó (Silverstone 1973) throughout La Victoria, (Lat. 5.51074, Lon. -76.86979), El Salero, (Lat. 5.36412, Lon.-76.64629), Nuquí, (Lat.5.70652, Lon.-77.26079) until Quebrada Docordo, department of Chocó (Lat. 4.56265, Lon.-77.01399). In contrast to *O. histrionica*, this species inhabits lower altitudes (0 to 324 m.a.s.l) with an annual mean temperature of 25.9 °C and annual mean precipitation of 6279 mm.

**Diagnosis:** Body size in these frogs ranges between 53-33 mm ( $36.2 \pm 1.7$  mm,  $n=75$ ). This species represents three different morphologic clusters (morphotypes) detected in this study with high bootstrap support ( $BS>82\%$ ) (Fig. S3 in Appendix 3). Individual frogs are black with bright orange or reddish orange spots that are highly variable in size and number ( $n=3-30$ ); however, the coloured dorsal area remains relatively constant (40-70% of total dorsal area). Colour variation is exemplified by Myers and Daly (1976) (Fig. 10, pp. 209).

***Oophaga andresi* sp. nov.** [Mitochondrial clade II]

We designate this species as *O. andresi* sp. nov. (syn. *O. histrionica*) based on the genetic species delimitation (mtDNA and nDNA,  $PS=1$ ) and the general conservative agreement of the integrative analysis ( $BS=98\%$ ). We named this new species after our beloved friend Andres Orrego-Ortiz, who passed away at a very young age in Palmira, Valle del Cauca, Colombia, while we were still collecting and analysing data there. Museum specimens representing this species are AMNH 85186-85189 and AMNH 86957-86960 (Myers and Daly 1976)

*Range:* This species is distributed at the east banks of the Atrato river which probably represent a geographic barrier between *O. solani* and *O. andresi* (Fig. 1a). After confirmation of the presence of this species in those geographic locations originally visited by Myers and Daly (1976) plus our own field records, we consider that populations of *O. andresi* are distributed from the surrounding areas of Tutunendo, Choco (Lat.5.75219, Lon.-76.53634) throughout Pacurita, Choco (Lat 5.646565, Lon.-76.56655), Bagado, Choco (Lat.5.38121, Lon.-76.42759) until Vereda La Angostura, Playa de Oro, Choco (Lat. 5.32194, Lon.-76.43090). Similar to *O. solani*, this species inhabits the relatively warmer lowlands (annual mean temperature of 26.6 °C; altitude 55 to 168 m.a.s.l). However, *O. andresi* exhibits the higher requirements of humidity, inhabiting locations with annual mean precipitation >7500 mm.

*Diagnosis:* Frogs of this species are similar in body size to *O. histrionica* ( $34.9 \pm 2.4$  mm, n=43) and are represented by two distinct morphotypes detected in this study (BS>91%) (Fig. S1 in Appendix 3). The basic and most common coloration pattern is represented by one to three large orange-reddish-yellow spots and bracelets on a background ranging from light to dark brown, but never black in colour. Detailed description of variation in spot sizes and distribution, as well as a representation of morphologic variation is presented in Myers and Daly (1976) as “*Playa de Oro*” frogs (Fig. 8-9, pp.206-208).

#### ***Oophaga histrionica* Berthold 1846. [Mitochondrial clade III]**

*Oophaga histrionica* (syn. *Dendrobates histrionicus*) was originally described as a species (Berthold 1846) based on descriptions of an unknown holotype (reproduced in Fig. 6, pp. 200, Myers & Daly, 1976) phenotypically similar to frogs occurring on the upper Rio San

Juan, in the Department of Risaralda, Colombia. In this study, *O. histrionica* was highly supported as a unique independent evolutionary lineage by all individual lines of evidence and the integrative analysis (Fig. 1e). A detailed description of coloration and pattern variation of this species is presented in Myers & Daly (1976) as “*Santa Cecilia*” frogs (pp. 205). Museum specimens that represent this species are deposited at the American Museum of Natural History (AMNH) under collection numbers AMNH 85159-85161 (Myers and Daly 1976).

*Range:* Based on our evidence, and contrary to the previously described range, this species is distributed in a narrow geographic range at an altitude between 342-410 m (mean 394 m). Populations of this species are geographically restricted to less than 30 km<sup>2</sup> in the surrounding areas of the municipality of Santa Cecilia, Colombia (Lat. 5.30336, Lon. - 76.21662) with annual mean temperature of 25.6 °C and annual mean precipitation of 5267 mm.

*Diagnosis:* Frogs of this species are significantly smaller in body size than other species of the complex ( $32.9 \pm 1.7$  mm, n=28). This species represent a clearly defined and unique morphologic cluster (BS=82%) (Fig. S1 in Appendix 3) where individual frogs have light brown to black bodies with usually one or two large orange or red-orange spots on the back. In rare cases, the dorsal marking is absent or difficult to notice. Morphological variation is shown in Myers and Daly (1976) (Fig. 7, pp. 205).

#### ***Oophaga anchicayensis* sp. nov.** [Mitochondrial clade IV]

We designate this species as *O. anchicayensis* sp. nov. (syn. *O. histrionica*) based on the genetic species delimitation (mtDNA and nDNA, PS=1) and the general conservative agreement of the integrative analysis (BS=100%). We named this new species after the type

locality Anchicaya, Buenaventura, Colombia. Museum specimens representing this species are deposited at the Universidad del Valle, Cali (UVC 12431-52) (Lötters et al. 1999), at the Institut Royal des Sciences Naturelles de Belgique (IRSNB) (MRHN 1038) and AMNH 10613-15 (Silverstone 1975).

*Range:* This species is located at southern localities (Fig. 1a) and its geographic range correspond to <170 km<sup>2</sup> at Valle del Cauca department, Colombia, at higher elevations when compared with other species of the genus (616-791 m.a.s.l). Populations of this species has been reported at Delfina, Buenaventura (Lat.3.81364, Lon.-76.81305), Cisneros (Lat.3.79955, Lon.-76.80305), Rio Zabaletas, Buenaventura (Lat.3.72003, Lon.-76.91802), Bajo Anchicaya, Dagua (Lat.3.63621, Lon.-76.94299), El Danubio, Dagua (Lat.3.62210, Lon.-76.90199) and vereda El Cauchal, Dagua (Lat.3.61712, Lon.-76.865). In contrast to *O. solani* and *O. andresi*, this species inhabits cooler and dryer geographic regions (annual mean temperature=24.4 °C; annual mean precipitation=2419 mm).

*Diagnosis:* These frogs are significantly bigger in body size than other species of the complex (41.2 ±1.3 mm, n=29). Background coloration is always completely black with dorsal yellow, red-orange or orange spots highly variable in number (3 to 20) and size. Dorsal pattern sometimes includes median bracelets as *O. lehmanni* (Myers and Daly 1976) that are incompletely formed ventrally in most cases. Some populations as “*Delfina*” and “*Cisneros*” (Silverstone 1975; Medina et al. 2013) present an increased reddish coloration towards the head and mouth (first frontal quarter of size length). This pattern is locally known as “*red head*” frogs and is shown in Lötters et al. (1999) (Fig. 3, pp. 27) However, this species is represented by a single morphological cluster with high statistical support (BS=95%) (Fig. S1 in Appendix 3).

# *Conservation implications*

This paper also illustrates how understanding lineage diversification can inform wildlife management and has relevance to the conservation efforts of South American poison frogs. The new integrative taxonomy proposed here has very important conservation implications as it revealed that some of the species should be considered amongst the most vulnerable species of Neotropical frogs: of the five species (*O. histrionica*, *O. lehmanni*, *O. solani*, *O. andresi*, and *O. anchicayensis*) two of them (*O. histrionica* and *O. lehmanni*) are single-locality endemics and should be both considered critically endangered. The inclusion of *O. lehmanni* as a critically endangered species more than 10 years ago led to the proposal of several conservation programs and regional legal policies aiming to preserve the remaining populations and their habitat (Bolívar et al. 2004; Valencia-Zuleta et al. 2014). Our study reveals that the distribution of *O. histrionica* is restricted to a very small area (< 30 km<sup>2</sup>), and the only 3 known populations of this microendemic species should probably be subject of similar actions. At the very least, *O. histrionica* should be listed as Critically Endangered in the UICN list and the resolution N<sup>o</sup> 0192-2014 (Ministerio de Ambiente y Desarrollo Sostenible) that establishes the Colombian list of endangered species. Also, given that these two species are highly sought-after in the pet trade market, it will be desirable to include them in the (Appendix I) of the CITES treaty to prevent the commercial trade of wild-caught specimens.

The conservation status of the newly described species (*O. anchicayensis*, *O. andresi* and *O. solani*) is less clear as they all have relatively wide, discontinuous, distributions. Although their effective population sizes are likely to be large, preliminary data suggest that populations are strongly structured. Therefore, several management (MUs) and evolutionary significant

(ESUs) units (Funk and Richardson 2002) should probably be considered. In the absence of any better information regarding species population ecology, density or abundance, we consider that these species should be temporarily categorized as vulnerable to extinction. Regardless of their conservation status none of the *Oophaga* species described in this paper occur in protected areas. Ideally, once included in the appropriate endangered-species list, these charismatic frogs will serve as umbrella species for the future conservation of entire ecosystems.

#### FUNDING

This work was supported through a 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.

#### ACKNOWLEDGEMENTS

We dedicate this paper to the memory of Richard G. Harrison, beloved friend and inspirational mentor. We would like to thank the members of his lab at Cornell University for their comments on an early version of this manuscript and Biologists Pablo Palacios and Ramon Ramirez for their assistance during the initial field work.

# REFERENCES

- Andrés J.A. , Bogdanowicz S.M. 2011. Isolating Microsatellite Loci: Looking Back, Looking Ahead. Molecular Methods for Evolutionary Genetics (ed. by V. Orgogozo and V.M. Rockman), pp. 211-232. Humana Press, Totowa, NJ.
- Barley A.J., White J., Diesmos A.C. , Brown R.M. 2013. The challenge of species delimitation at the extremes: diversification without morphological change in Philippine sun skinks. Evolution. 67: 3556-3572.
- Baum D.A. , Donoghue M.J. 1995. Choosing among Alternative "Phylogenetic" Species Concepts. Syst. Bot. 20: 560-573.
- Berthold A.A. 1846. Ueber verschiedene neue oder seltene Reptilien aus Neu-Granada und Crustaceen aus China. Nachr. Ges. Wiss. Göttingen. 3: 43.
- Bolivar W., Castro F. , Lötters S. 2004. *Oophaga lehmanni*. The IUCN Red List of Threatened Species. e.T55190A11255085.
- Brown D., Brenneman R., Koepfli K.-P., Pollinger J., Mila B., Georgiadis N., Louis E., Grether G., Jacobs D. , Wayne R. 2007. Extensive population genetic structure in the giraffe. BMC Biol. 5: 57-70.
- Brown J.L., Twomey E., Amézquita A., Souza M.B.d., Caldwell J.p., Lötters S., May R.v., Melo-Sampaio P.R., Mejía-Vargas D., Perez-Peña P., Pepper M., Poelman E.h., Sanchez-Rodriguez M. , Summers K. 2011. A taxonomic revision of the Neotropical poison frog genus *Ranitomeya* (Amphibia: Dendrobatidae). Zootaxa. 3083: 1-120.

636 Brusa O., Bellati A., Meuche I., Mundy N.I. , Pröhl H. 2013. Divergent evolution in the  
637 polymorphic granular poison-dart frog, *Oophaga granulifera*: genetics, coloration,  
638 advertisement calls and morphology. J. Biogeogr. 40: 394-408.

639 Cardoso A., Serrano A. , Vogler A.P. 2009. Morphological and molecular variation in tiger  
640 beetles of the *Cicindela hybrida* complex: is an 'integrative taxonomy' possible? Mol.  
641 Ecol. 18: 648-664.

642 Carstens B.C., Pelletier T.A., Reid N.M. , Satler J.D. 2013. How to fail at species delimitation.  
643 Mol. Ecol. 22: 4369-4383.

644 Coyne J.A. , Orr H.A. 2004. Speciation. Sinauer Associates Sunderland, MA.

645 Culver M., Johnson W., Pecon-Slaterry J. , O'Brien S. 2000. Genomic ancestry of the American  
646 puma (*Puma concolor*). J. Hered. 91: 186-197.

647 Daly J.W., Brown G.B., Mensah-Dwumah M. , Myers C.W. 1978. Classification of skin  
648 alkaloids from neotropical poison-dart frogs (dendrobatidae). Toxicon. 16: 163-188.

649 Dayrat B. 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. 85: 407-415.

650 de Queiroz K. 2005a. Ernst Mayr and the modern concept of species. Proc. Natl. Acad. Sci. U. S.  
651 A. 102: 6600-6607.

652 de Queiroz K. 2005b. Different species problems and their resolution. Bioessays. 27: 1263-1269.

653 Dejaco T., Arthofer W., Sheets H.D., Moder K., Thaler-Knoflach B., Christian E., Mendes L.F.,  
654 Schlick-Steiner B.C. , Steiner F.M. 2012. A toolbox for integrative species delimitation in  
655 Machilis jumping bristletails (Microcoryphia: Machilidae). Zoologischer Anzeiger - A  
656 Journal of Comparative Zoology. 251: 307-316.

657 Edwards D.L. , Knowles L.L. 2014. Species detection and individual assignment in species  
658 delimitation: can integrative data increase efficacy? *Proc. R. Soc. Lond., Ser. B: Biol.*  
659 *Sci.* 281: 20132765.

660 Fraley C. , Raftery A.E. 2002. Model-Based Clustering, Discriminant Analysis, and Density  
661 Estimation. *J Am Stat Assoc.* 97: 611-631.

662 Funk V.A. , Richardson K.S. 2002. Systematic Data in Biodiversity Studies: Use It or Lose It.  
663 *Syst. Biol.* 51: 303-316.

664 Grant T., Frost D.R., Caldwell J.P., Gagliardo R.O.N., Haddad C.F.B., Kok P.J.R., Means D.B.,  
665 Noonan B.P., Schargel W.E. , Wheeler W.C. 2006. Phylogenetic systematics of dart-  
666 poison frogs and their relatives (Amphibia: Athesphatanura: Dendrobatidae). *Bull Am*  
667 *Mus Nat Hist.* 1-262.

668 Grummer J.A., Bryson R.W. , Reeder T.W. 2013. Species Delimitation Using Bayes Factors:  
669 Simulations and Application to the *Sceloporus scalaris* Species Group (Squamata:  
670 *Phrynosomatidae*). *Syst. Biol.*

671 Hammer Ø., Harper D. , Ryan P. 2001. PAST: Paleontological Statistics Software Package for  
672 education and data analysis. *Palaeontologia Electronica.* 4(1): 9.

673 Hausdorf B. , Hennig C. 2010. Species Delimitation Using Dominant and Codominant  
674 Multilocus Markers. *Syst. Biol.*

675 Hauswaldt J.S., Ludewig A.-K., Vences M. , Pröhl H. 2011. Widespread co-occurrence of  
676 divergent mitochondrial haplotype lineages in a Central American species of poison frog  
677 (*Oophaga pumilio*). *J. Biogeogr.* 38: 711-726.

678 Hedrick P.W. 1995. Gene Flow and Genetic Restoration: The Florida Panther as a Case Study.  
679 *Conserv. Biol.* 9: 996-1007.

680 Helbig A.J., Knox A.G., Parkin D.T., Sangster G. , Collinson M. 2002. Guidelines for assigning  
681 species rank†. Ibis. 144: 518-525.

682 Heled J. , Drummond A.J. 2010. Bayesian inference of species trees from multilocus data. Mol.  
683 Biol. Evol. 27: 570-580.

684 Hijmans R., Guarino L., Cruz M. , Rojas E. 2001. Computer tools for spatial analysis of plant  
685 genetic resources data: 1. DIVA-GIS. Plant Genetic Resources Newsletter. 15-19.

686 Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G. , Jarvis A. 2005. Very high resolution  
687 interpolated climate surfaces for global land areas. Int J Climatol. 25: 1965-1978.

688 Hubisz M.J., Falush D., Stephens M. , Pritchard J.K. 2009. Inferring weak population structure  
689 with the assistance of sample group information. Mol. Ecol. Resour. 9: 1322-1332.

690 Huelsenbeck J.P., Andolfatto P. , Huelsenbeck E.T. 2011. Structurama: Bayesian Inference of  
691 Population Structure. Evol Bioinform Online. 7: 55-59.

692 Jombart T., Devillard S. , Balloux F. 2010. Discriminant analysis of principal components: a new  
693 method for the analysis of genetically structured populations. BMC Genet. 11: 94.

694 Leaché A.D. , Fujita M.K. 2010. Bayesian species delimitation in West African forest geckos  
695 (*Hemidactylus fasciatus*). Proc. R. Soc. Lond., Ser. B: Biol. Sci.

696 Leaché A.D., Fujita M.K., Minin V.N. , Bouckaert R.R. 2014. Species Delimitation using  
697 Genome-Wide SNP Data. Syst. Biol. 63: 534-542.

698 Lötters S. 1992. Zur Validität von *Dendrobates lehmani* Myers & Daly, 1976 aufgrund zweier  
699 neuer Farbformen von *Dendrobates histrionicus* Berthold, 1845. Salamandra. 15: 138-  
700 144.

701 Lötters S., Glaw F., Kohler F. , Castro F. 1999. On the geographic variation of the advertisement  
702 call of *Dendrobates histrionicus* Berthold, 1845 and related forms from north-western  
703 South America. *Herpetozoa*. 12: 23-38.

704 Lynch J.D. , Arroyo S.B. 2009. Risks to Colombian Amphibian Fauna from Cultivation of Coca  
705 (*Erythroxylum coca*): A Geographical Analysis. *J. Toxicol. Environ. Health, A*. 72: 974-  
706 985.

707 Mayr E. 1942. Systematics and the origin of species, from the viewpoint of a zoologist. Harvard  
708 University Press.

709 Medina I., Wang I.J., Salazar C. , Amézquita A. 2013. Hybridization promotes color  
710 polymorphism in the aposematic harlequin poison frog, *Oophaga histrionica*. *Ecol. Evol.*  
711 3: 4388-4400.

712 Mendez-Narvaez J. , Amezquita A. 2014. Physical combat in the poison-arrow frog, Kokoé-pá  
713 (*Oophaga histrionica*) from Arusi, Choco, Colombia. *Herpetology Notes*. 7: 1-2.

714 Myers C.W. , Daly J.W. 1976. Preliminary evaluation of skin toxins and vocalizations in  
715 taxonomic and evolutionary studies of poison-dart frogs (Dendrobatidae). *Bull Am Mus*  
716 *Nat Hist*. 175-262.

717 Oksanen J., Blanchet F., Kindt R., Legendre P. , Minchin P. 2013. vegan: Community Ecology  
718 Package. R package version 2.0-6. Available at:  
719 <http://mirror.bjtu.edu.cn/cran/web/packages/vegan/>.

720 Padial J.M. , De la Riva I. 2006. Taxonomic Inflation and the Stability of Species Lists: The  
721 Perils of Ostrich's Behavior. *Syst. Biol.* 55: 859-867.

722 Padial J.M. , De La Riva I. 2009. Integrative taxonomy reveals cryptic Amazonian species of  
723 *Pristimantis* (Anura: Strabomantidae). *Zool. J. Linn. Soc.* 155: 97-122.

724 Padial J.M. , De La Riva I. 2010. A response to recent proposals for integrative taxonomy. Biol.  
725 J. Linn. Soc. 101: 747-756.

726 Padial J.M., Miralles A., De la Riva I. , Vences M. 2010. The integrative future of taxonomy.  
727 Front Zool. 7: 1-14.

728 Peakall R. , Smouse P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic  
729 software for teaching and research—an update. Bioinformatics. 28: 2537-2539.

730 Phillott A.D., McDonald K.R. , Skerratt L.F. 2011. Inflammation in digits of unmarked and toe-  
731 tipped wild hylids. Wildl. Res. 38: 204-207.

732 Posada D. 2008. jModelTest: Phylogenetic Model Averaging. Mol. Biol. Evol. 25: 1253-1256.

733 Posso-Terranova A. , Andres J. 2016. Complex niche divergence underlies lineage  
734 diversification in *Oophaga* poison frogs. J. Biogeogr. Accepted

735 Pritchard J.K., Stephens M. , Donnelly P. 2000. Inference of Population Structure Using  
736 Multilocus Genotype Data. Genetics. 155: 945-959.

737 Reeves P.A. , Richards C.M. 2011. Species delimitation under the general lineage concept: an  
738 empirical example using wild North American hops (Cannabaceae: *Humulus lupulus*).  
739 Syst. Biol. 60: 45-59.

740 Ruane S., Bryson R.W., Pyron R.A. , Burbrink F.T. 2014. Coalescent Species Delimitation in  
741 Milksnakes (Genus *Lampropeltis*) and Impacts on Phylogenetic Comparative Analyses.  
742 Syst. Biol. 63: 231-250.

743 Sheffield J. 2007. ImageJ: A useful tool for biological image processing and analysis. Microsc.  
744 Microanal. 13: 200-201.

745 Silverstone P.A. 1973. Observations on the Behavior and Ecology of a Colombian Poison-Arrow  
746 Frog, the Kōkoé-Pá (*Dendrobates histrionicus* Berthold). Herpetologica. 29: 295-301.

747 Silverstone P.A. 1975. A revision of the poison-arrow frogs of the genus *Dendrobates* Wagler.  
748 Natural History Museum of Los Angeles County. 21: 1-55.

749 Stevens M., Parraga C.A., Cuthill I.C., Partridge J.C. , Troscianko T.S. 2007. Using digital  
750 photography to study animal coloration. Biol. J. Linn. Soc. 90: 211-237.

751 Tamura K., Stecher G., Peterson D., Filipski A. , Kumar S. 2013. MEGA6: Molecular  
752 Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30: 2725-2729.

753 Thompson J.D., Higgins D.G. , Gibson T.J. 1994. CLUSTAL W: improving the sensitivity of  
754 progressive multiple sequence alignment through sequence weighting, position-specific  
755 gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673-4680.

756 Valcárcel V. , Vargas P. 2010. Quantitative morphology and species delimitation under the  
757 general lineage concept: Optimization for Hedera (Araliaceae). Am. J. Bot. 97: 1555-  
758 1573.

759 Valencia-Zuleta A., Jaramillo-Martínez A., Echeverry-Bocanegra A., Viáfara-Vega R.,  
760 Hernández-Córdoba O., Cardona-Botero V., Gutiérrez-Zúñiga J. , Castro-Herrera F.  
761 2014. Conservation status of the herpetofauna, protected areas, and current problems in  
762 Valle del Cauca, Colombia. Amphibian & Reptile Conservation. 8: 1-18.

763 Van Valen L. 1976. Ecological species, multispecies, and oaks. Taxon. 233-239.

764 Vieites D.R., Wollenberg K.C., Andreone F., Köhler J., Glaw F. , Vences M. 2009. Vast  
765 underestimation of Madagascar's biodiversity evidenced by an integrative amphibian  
766 inventory. Proc Natl Acad Sci U S A Biol Sci. 106: 8267-8272.

767 Yang Z. 2015. The BPP program for species tree estimation and species delimitation. Current  
768 Zoology. 61(5): 1-16.

769 Yeates D.K., Seago A., Nelson L., Cameron S.L., Joseph L.E.O. , Trueman J.W.H. 2011.  
770 Integrative taxonomy, or iterative taxonomy? Syst. Entomol. 36: 209-217.  
771 Zapata F. , Jiménez I. 2012. Species Delimitation: Inferring Gaps in Morphology across  
772 Geography. Syst. Biol. 61: 179-194.  
773 Zhang C., Rannala B. , Yang Z. 2014. Bayesian Species Delimitation Can Be Robust to Guide-  
774 Tree Inference Errors. Syst. Biol. 63: 993-1004.  
775 Zhang J., Kapli P., Pavlidis P. , Stamatakis A. 2013. A general species delimitation method with  
776 applications to phylogenetic placements. Bioinformatics. 29: 2869-2876.  
777  
778

## Figure captions

Figure 1. Left to right. a) Map of the sampling localities. b) Representative examples of phenotypic variation found in the “*histrionica*” complex and posterior assignment probabilities to each of the nine distinct morphological groups inferred by the DAPC analysis of 50 morphological variables (see Fig. S1a in Appendix 3 provided in Supplementary Material). c) Schematic representation of the different bioclimatic niches occupied by different populations as inferred by the NMDS analysis of 19 bioclimatic variables and altitude. The multivariate representation of the Gaussian bioclimatic clusters is provided in Figure S2 in Appendix 3 provided in Supplementary Material. d) Grey: Monophyletic mitochondrial clades (I to IV and *O. lehmanni*) as inferred by Bayesian and ML methods. A detailed topology of the phylogenetic tree including *O. sylvatica* and *O. occultator* as outgroups is shown in Figure S3 in Appendix 3 provided in Supplementary Material. The corresponding parsimony haplotype network is available in Figure S4 in Appendix 3 provided in Supplementary Material. Color: Posterior probability of assignment to five genetic clusters based on Bayesian analysis of variation at microsatellite and transcriptome-derived haplotypes. Graphical representation of the corresponding PCoAs are provided in Figures S5a and S5b in Appendix 3 in Supplementary Material. e) In grey, species delimitation based on Myers and Daly (1976) and Funkhouser (1956). Coloured-boxes represent the new delimitation resulting from our multivariate integrative analyses. *sp. nov.* indicate the new species proposed, and described, in this paper.

Figure 2. Integrative species delimitation based on the hierarchical clustering standardized orthogonal variables (genetic, ecological and morphological). a) Provisional

802 delimitation resulting from the analysis of 84 specimens. b) Extended clustering analysis of 214  
803 individuals. The colored boxes enclose individual specimens belonging to the same Gaussian  
804 cluster,  $PP \geq 0.95$ ). Only nodes with support  $>95\%$  are indicated.

805

806

# APPENDIX 1: SUPPLEMENTARY TABLES

Ecology, molecules and colour: Multivariate species delimitation and conservation of  
Harlequin poison frogs

Andres Posso-Terranova<sup>1-2\*</sup>, Jose A. Andres<sup>1</sup>

<sup>1</sup>*Department of Biology. University of Saskatchewan. 112 Science Pl. Saskatoon. SK. Canada*

<sup>2</sup>*Universidad Nacional de Colombia sede Palmira. A.A. 237. Palmira. Colombia*

*\*Correspondence to be sent to: 112 Science Pl. Saskatoon. SK. Canada. S7N5E2; E-mail:*

[andres.posso@usask.ca](mailto:andres.posso@usask.ca)

823 Table S1. Description of 49 morphologic variables used for the estimation of morphological clusters and combinatory species  
 824 delimitation.

825

826

Type	Position	Variable	Description of measurement	Units/Type
Morphometric	Dorsal/ventral	<b>Snout-cloaca length</b> (av_lenght)	Mean value of dorsal/ventral distance between the tip of the nose and the cloaca.	mm
Morphometric	Dorsal	<b>Inter-orbital width</b> (eye)	Width of the skull between the ocular ridges	mm
Morphometric	Dorsal	<b>Dorsal area</b> (d_area)	Dorsal area without including limbs	mm <sup>2</sup>
Morphometric	Dorsal	<b>Head Depth</b> (snout)	depth of the head at its deepest point immediately posterior to the orbital cavities	mm
Coloration pattern	Dorsal	<b>Number of colored dorsal spots</b> (nspots_d)	Count of spots with bright coloration in dorsal surface	count
Coloration pattern	Dorsal	<b>Median dorsal color spot area</b> (d_spots_median)	Median value for the area of dorsal spots/individual	mm <sup>2</sup>
Coloration pattern	Dorsal	<b>Average of dorsal color spot area</b> (d_spots_mean)	Mean value for the area of dorsal spots/individual	mm <sup>2</sup>
Coloration pattern	Dorsal	<b>Standard deviation of dorsal color spot area</b> (d_spots_std)	Standard deviation value for the area of dorsal spots/individual	mm <sup>2</sup>
Coloration pattern	Dorsal	<b>Percentage of dorsal coloration</b> (%_col_d)	Summatory of individual dorsal color spot areas divided by the total dorsal area	%
Coloration pattern	Dorsal	<b>Dorsal colored spots longitudinal centrality</b> (xmean_c)	Mean distance from inter-pupilar Y axes to the centroid of each colored spot in the X axes.	mm

		<b>Standard deviation longitudinal centrality</b>	Standard deviation value for Dorsal colored spots longitudinal centrality	mm
Coloration pattern	Dorsal	(xcenter_std)		
		<b>Dorsal colored spots transversal centrality</b>	Mean distance from inter-pupilar Y axes to the centroid of each colored spot in the Y axes.	mm
Coloration pattern	Dorsal	(ymean_c)		
		<b>Standard deviation transversal centrality</b>	Standard deviation value for Dorsal colored spots transversal centrality	mm
Coloration pattern	Dorsal	(ycenter_std)		
Coloration pattern	Dorsal	<b>Mean dorsal circularity</b> (cd_mean)	Mean value for dorsal colored spots circularity ( $4*\pi*Area/Perimeter^2$ )	no units
		<b>Standard deviation dorsal spots circularity</b>	Standard deviation value for dorsal colored spots circularity	no units
Coloration pattern	Dorsal	(circ_d_std)		
Coloration pattern	Dorsal	<b>Mean dorsal roundness</b> (rd_mean)	Mean value for dorsal colored spots roundness ( $4*\pi*Area/Feret^2$ )**	no units
		<b>Standard deviation dorsal spots roundness</b>	Standard deviation value for dorsal colored spots roundness	no units
Coloration pattern	Dorsal	(roun_d_std)		
Coloration pattern	Ventral	<b>Number of colored ventral spots</b> (nspots_v)	Count of spots with bright coloration in ventral surface	count
Coloration pattern	Ventral	<b>Ventral area</b> (v_area)	Ventral area without including limbs	mm <sup>2</sup>
Coloration pattern	Ventral	<b>Median ventral color spot area</b> (v_spots_median)	Median value for the area of ventral spots/individual	mm <sup>2</sup>
		<b>Average of ventral color spot area</b>	Mean value for the area of ventral spots/individual	mm <sup>2</sup>
Coloration pattern	Ventral	(v_spots_mean)		
		<b>Standard deviation of ventral color spot area</b>	Standard deviation value for the area of ventral spots/individual	mm <sup>2</sup>
Coloration pattern	Ventral	(v_spots_std)		
		<b>Percentage of ventral coloration</b> (%_col_v)	Summatory of individual ventral color spot areas divided by the total dorsal area	%
Coloration pattern	Ventral			

Coloration pattern	Ventral	<b>Ventral colored spots longitudinal centrality</b>	Mean distance from inter-pupilar Y axes to the centroid of each colored spot in the X axes (Ventral)	mm
		(xvmean_c)		
Coloration pattern	Ventral	<b>Standard deviation ventral longitudinal centrality</b>	Standard deviation value for Dorsal colored spots longitudinal centrality (Ventral)	mm
		(xvcenter_std)		
Coloration pattern	Ventral	<b>Ventral colored spots transversal centrality</b>	Mean distance from inter-pupilar Y axes to the centroid of each colored spot in the Y axes (Ventral)	mm
		(yvmean_c)		
Coloration pattern	Ventral	<b>Standard deviation ventral transversal centrality</b>	Standard deviation value for ventral colored spots transversal centrality (Ventral)	mm
		(yvcenter_std)		
Coloration pattern	Ventral	<b>Mean ventral circularity</b> (cv_mean)	Mean value for ventral colored spots circularity ( $4*\pi*\text{Area}/\text{Perimeter}^2$ )	no units
Coloration pattern	Ventral	<b>Standard deviation dorsal spots circularity</b>	Standard deviation value for ventral colored spots circularity	no units
		(circ_v_std)		
Coloration pattern	Ventral	<b>Mean ventral roundness</b> (rv_mean)	Mean value for dorsal colored spots roundness ( $4*\pi*\text{Area}/\text{Feret}^2$ )**	no units
Coloration pattern	Ventral	<b>Standard deviation dorsal spots roundness</b>	Standard deviation value for ventral colored spots roundness	no units
		(round_v_std)		
Coloration intensity	Dorsal	<b>Dorsal Background red intensity</b> (red_b_d)	R value (RGB component ) dorsal background coloration	RGB Units
Coloration intensity	Dorsal	<b>Dorsal Background green intensity</b> (green_b_d)	G value (RGB component ) dorsal background coloration	RGB Units
Coloration intensity	Dorsal	<b>Dorsal Background blue intensity</b> (blue_b_d)	B value (RGB component ) dorsal background coloration	RGB Units
Coloration intensity	Dorsal	<b>Dorsal Background Brightness Index</b>	Dorsal background coloration (brightness index calculated as $0.299R+0.587G+0.114B$ )	RGB Units
		(indexbrigh_ind_b_d)		

Coloration intensity	Dorsal	<b>Dorsal colored red intensity</b> (red_d_d)	R value (RGB component ) dorsal colored coloration	RGB Units
Coloration intensity	Dorsal	<b>Dorsal colored green intensity</b> (green_d_d)	G value (RGB component ) dorsal colored coloration	RGB Units
Coloration intensity	Dorsal	<b>Dorsal colored blye intensity</b> (blue_d_d)	B value (RGB component ) dorsal colored coloration	RGB Units
Coloration intensity	Dorsal	<b>Dorsal colored Brightness Index</b> (brigh_ind_d_d)	Dorsal bright coloration (brightness index calculated as 0.299R+0.587G+0.114B))	RGB Units
Coloration intensity	Dorsal	<b>Dorsal Brightness difference</b> (brigh_diff_d)	Dorsal Brightness difference (brigh_ind_d_d) - (brigh_ind_b_d)	RGB Units
Coloration intensity	Ventral	<b>Ventral Background red intensity</b> (red_b_d)	R value (RGB component ) ventral background coloration	RGB Units
Coloration intensity	Ventral	<b>Ventral Background green intensity</b> (green_b_d)	G value (RGB component ) ventral background coloration	RGB Units
Coloration intensity	Ventral	<b>Ventral Background blue intensity</b> (blue_b_d)	B value (RGB component ) ventral background coloration	RGB Units
Coloration intensity	Ventral	<b>Ventral Background Brightness Index</b> (indexbrigh_ind_b_d)	Ventral background coloration (brightness index)	RGB Units
Coloration intensity	Ventral	<b>Ventral colored red intensity</b> (red_d_d)	R value (RGB component ) ventral colored coloration	RGB Units
Coloration intensity	Ventral	<b>Ventral colored green intensity</b> (green_d_d)	G value (RGB component ) ventral colored coloration	RGB Units

Coloration intensity	Ventral	<b>Ventral colored blye intensity</b> (blue_d_d)	B value (RGB component ) ventral colored coloration	RGB Units
Coloration intensity	Ventral	<b>Ventral colored Brightness Index</b> (brigh_ind_d_d)	Ventral colored coloration (brightness index)	RGB Units
Coloration intensity	Ventral	<b>Ventral Brightness difference</b> (brigh_diff_d)	Ventral Brightness difference (brigh_ind_d_d) - (brigh_ind_b_d)	RGB Units

Table S2. Description of nuclear markers implemented in this study.

Loci	Forward Sequence (5'-3')	Reverse Sequence (5'-3')	Type	PCR amplicon	Estimated Size
4_hist	TAGCAATGTCGCTGGATGAC	ACAAAGCACCGCACTACAT	microsatellite	Yes	156-170
13_hist	TCCAGTTTGTGAGGTAGCAAAA	AAAACGGGAAAGATTTATTCTGA	microsatellite	Yes	160-190
44_hist	TCACTGCTTCATTAATTGTT	GAATCCAGCAGCCTCTGTAAG	microsatellite	Yes	196-290
63_hist	CCCAGGGTAGAGTTGCTTTTT	TTCGCTTTTCTTCCCTTCA	microsatellite	Yes	190-240
88_hist	CGGTCCACCTCTTCCCTGT	ATGCAAGCTTCCTGCTCTGT	microsatellite	No	205-233
147_hist	ACCTTATACATCTGCACGCTTTC	TCATTATCCAACGACATGCTTCC	microsatellite	No	240-285
109_hist	TGTGTTTGTATGCGGTACG	TGCCATGGTGTGTGATGAG	microsatellite	No	400-418
110_hist	TCACCTTTCATCCAGTTCA	TGGTATTGTGCTGTTCCACTTT	microsatellite	Yes	198-260
1696_hist	GGACTATCCGAATTTGCAT	CCATTTTGAGGGAGTGAGGA	microsatellite	No	160-190
3024_hist	TGTGAGAAATATTTGGAGGCTGG	GTCTGAATACTTTCCGTACCCAC	microsatellite	No	150-198
5133_hist	ATTAAAGCAATATAAGCCGGCCC	TCCGCAATTTGTCATGGTTCTAG	microsatellite	No	280-330
6868_hist	ACGGGCTAAGACTGCGATATAG	CACGCTATGGGTTCAACTG	microsatellite	Yes	165-190
comp324	GCTCTTCCGATCTCTGTGTT	TGTTTTCCAGAAACAAACCA	tbn-DNA	Yes	400-500
comp328	TGAACTCACCAGTAAACCTTTG	AATAGCCCTTGGTTGTATGG	tbn-DNA	No	400-500
comp332	CAAGTGCAGTCATGAAAACC	CCTCACATGTCCTGTGTCTC	tbn-DNA	Yes	400-500
comp344	TCACAAGACCTCCAGCAGTA	AGACAGCAGTTCCAGTGTGA	tbn-DNA	No	400-500
comp350	AAGCACTTCAAACCACAAGAC	AGCCAGAGGATGTGTAAACG	tbn-DNA	Yes	400-500
comp365	GGGGAGGAGAGAAAAATCAT	CACATTTCCCATTTCTGTACA	tbn-DNA	No	400-500
comp378	TGTAGTTTTCCTATATTTCTCTCAC	TAGTGAATGGATCCCTCTGG	tbn-DNA	Yes	400-500
comp380	GTGAACAGTCATTTTGTGCTG	CCCATCCACTACAGATCACA	tbn-DNA	No	400-500
comp396	ACAGGGCCTAATGCTGAG	CTGACTGCACTTTGGTTCTG	tbn-DNA	Yes	400-500
comp406	GTGGTGTCTTGGGAAGC	ATTTGGTTCAATTGTGTGTC	tbn-DNA	No	400-500
comp425	GGAATGATCTTTTGGTTCGT	GCCTCCCATCTAAACAGAAT	tbn-DNA	Yes	400-500
comp430	TCGTTCTCAAGAAACCCTGT	AGCATAGCAACCGAGTACCT	tbn-DNA	No	400-500
comp438	CATCCTTCTCTTCAGGCACT	TAAAGGCCATTTTAGGGACA	tbn-DNA	No	400-500
comp_462	AAAACAAGCTCCTGTCTGCT	AGAAGCTTTGGTTCCATCTG	tbn-DNA	No	400-500
comp444	TCATCCATCAGCTTAACTCCT	TCCATTCCATTAGATTGATCTTC	tbn-DNA	Yes	400-500
comp480	TTGGAACATAGTCATTTCGTCA	CACGTAAAACGGTAGGCTGT	tbn-DNA	Yes	400-500
comp484	TGACCTTATATAGGCAGTGCAG	TCACCCAGATTCCTTTATCC	tbn-DNA	No	400-500
comp493	GGGTGCAGTGATCTGTGAG	TCAACCACAAGAAGAACCTC	tbn-DNA	Yes	400-500

comp511	CAGCCTTACACAGGGATTG	CAACATCAAGAAATCGGCTA	tbn-DNA	Yes	400-500
comp534	CATCTTATCTTACCCAAATTTCC	GTGCTCTTCCGATCTGTTGT	tbn-DNA	No	400-500
comp539	TCAAGATTCAGTGGTGCAAG	CCTCTTCTGGACATGGAGAT	tbn-DNA	No	400-500
comp540	TATGTAAGAATGAGGTGGAGACA	CGTTCACCTATATATGCACACTT	tbn-DNA	No	400-500
comp547	AATTACCAGCCAGACTCCAG	GATGTTAAGCGTCACAGCAT	tbn-DNA	No	400-500
comp548	AGGCGTATAAGTGGATTTGC	CAGGAGCTGGTCTTGTAAACC	tbn-DNA	No	400-500
comp553	ACTTTGATGCAGTCTCACACA	CACATAGCAGTGCTTCAAAAA	tbn-DNA	No	400-500
comp562	GGACTCTACCTTGGGCAGTA	GTGCAAAAGTGTCACCAAAA	tbn-DNA	Yes	400-500
comp583	CGAAGAGACACCTCCCATTA	CGAGAGCTGCACTCACTTC	tbn-DNA	Yes	400-500
comp594	GTAAACCCGCAGGTGAACT	TAAAGTGACACGCTTTGGTG	tbn-DNA	Yes	400-500

Table S3. Quality descriptors of the assembled transcriptome used for nuclear loci markers design.

Species	<i>O. lehmanni</i>
<b>File size (Mb)</b>	77.3
<b># sequences (Contigs)</b>	117,082
<b>Total bases (b)</b>	66,367,430
<b>Mean sequence length (b)</b>	566.8
<b>Minimum length (b)</b>	201
<b>Maximum length (b)</b>	12,906
<b>N50 contig size (b)</b>	809
<b>N75 contig size (b)</b>	353

## APPENDIX 2: SUPPLEMENTARY METHODS

Ecology, molecules and colour: Multivariate species delimitation and conservation of  
Harlequin poison frogs

Andres Posso-Terranova<sup>1-2\*</sup>, Jose A. Andres<sup>1</sup>

<sup>1</sup>Department of Biology. University of Saskatchewan. 112 Science Pl. Saskatoon. SK.  
Canada

<sup>2</sup>Universidad Nacional de Colombia sede Palmira. A.A. 237. Palmira. Colombia

\*Correspondence to be sent to: 112 Science Pl. Saskatoon. SK. Canada. S7N5E2; E-  
mail: andres.posso@usask.ca

### *Mitochondrial genetic clustering*

To determine the number of mitochondrial clusters (KMT), phylogenetic trees were constructed under different Bayesian (BI) and maximum likelihood (ML) optimality criteria. First, using MEGA6 (Tamura et al. 2013), we obtained an ML genealogy using extensive (level 5) Subtree Pruning and Regrafting (SPR) heuristic searches. The relative support for each node was estimated by generating 1,000 bootstrap replicates. Second, we constructed a BI-based tree using BEAST v. 2.1.2 (Drummond et al. 2012; Bouckaert et al. 2014) with a relaxed molecular clock and an uncorrelated log-normal (UCLN) model of molecular rate heterogeneity. We ran three chains for 10 million generations sampled every 100 steps. The

resulting trees and log files of the three independent runs were combined using LOGCOMBINER 2.1.2. and resampled at a lower frequency (i.e. 400 steps or ¼) to reduce the number of final trees to 52,500 (Bouckaert et al. 2014). For each estimated parameter, convergence was assessed using TRACER 1.6 (<http://tree.bio.ed.ac.uk/software/tracer/>) and effective sample sizes (ESS) were calculated to ensure adequate mixing (ESS>320, after 30% burn in). We summarized the posterior probability density of the combined tree and log files as a maximum clade credibility (MCC) tree using TREEANNOTATOR 2.1.2 (Bouckaert et al. 2014). We visualized all trees using FIGTREE 1.4.0 (<http://tree.bio.ed.ac.uk/software/figtree/>). The phylogenetic relationship among sampling locations was plotted onto a TCS haplotype network with a connection limit of 95% (Clement et al. 2000) implemented in the program PopART (Population Analysis with reticulate trees) available at <http://popart.otago.ac.nz/>. Finally, to delimit independent evolving lineages we performed a Bayesian implementation of the PTP model for species delimitation (Zhang et al. 2013) at the bTPT online server (<http://species.h-its.org/>) with MCMC generations = 1,000,000, thinning = 1000 and Burn-in = 0.1.

### *Microsatellite molecular markers*

Microsatellite markers were generated as follow: genomic DNA from a single individual of the known *O. histrionica* nominal species was isolated as described in the Methods section of the main manuscript. The genomic sample was sequenced on one-eighth of a plate using 454GS-FLX technology (Roche Applied Science) at the Cornell University's BioResource Center. We obtained 14,992 fragment reads with lengths ranging from 150 to 825 bases. We searched for di-, tri- and tetra-nucleotide tandem repeats from the dataset using

MSATCOMMANDER (Faircloth 2008; Pino et al. 2012). The online version of PRIMER3 (Rozen and Skaletsky 2000) was used to access the suitability for annealing temperature and single primer-pair for each loci were chosen based on the highest PRIMER3 assigned score. Primer pairs were tested using conventional PCR in twelve individuals from two geographic localities and those loci that were successfully amplified (Table S3 in Appendix 1) were used to genotype 310 individuals from 15 different geographic locations (Table S1 in Appendix 1). We followed previously described protocols for multiplex PCR conditions, allele detection and allele calling (Anttila et al. 2014).

### *Nuclear loci for multiplex PCR*

First, we generated a skin transcriptome by Trizol (Invitrogen™) extraction of total RNA from skin portions of one known individual of *O. lehmanni*. Paired-end libraries (150-mer, x2; insert lengths ~300 bp) were synthesized using the Genomic Sample Prep kit (Illumina™) according to manufacturer's instructions. Library quality was assessed using a 2100 Bioanalyzer (Agilent). Libraries were sequenced on an Illumina Hi-Seq 2000 with a paired-end module. Recognition, sorting, trimming of tags and quality control of the raw DNA-sequence reads were performed in FLEXBAR (Dodt et al. 2012). The remaining reads (~128 million) and the TRINITY platform (Haas et al. 2013) were used for the generation of a de novo transcriptome. Further information about transcriptome composition and quality is presented in Table S4 in Appendix 1.

### *Allele-frequency genetic clustering*

In order to assign individuals of the *O. histrionica* complex to genetic clusters, we

implemented Bayesian clustering algorithms using a combination of allele frequency data from microsatellite (n=6) and transcriptome-derived haplotypes (n=13) in STRUCTURE v. 2.3.4 (Pritchard et al. 2000; Hubisz et al. 2009). A model assuming admixture and correlated allele frequencies was implemented in 10 independent runs with one million MCMC iterations (burn-in = 100000) for a series of clusters (KNUC) ranging from 1 (panmixia) to 15 (maximum number of localities sampled). Then, we used the statistic  $\Delta K$  to select the value of K with the uppermost hierarchical level of population structure in our data (Evanno et al. 2005). For detecting species clusters and individual assignment while treating KNUC as a random variable, we also implemented STRUCTURAMA v. 1.0 (Huelsenbeck et al. 2011). The models were run for 10 million generations, sampling every 100th cycle and a burn-in period of 10000 samples. A different set of prior values was used for the shape and scale of the gamma distribution of KNUC (shape: scale = 2.5:0.5, 2.5:1, 5:2, 10:1) (Edwards and Knowles 2014). We chose the KNUC with the highest probability and individuals were assigned to species cluster by using the posterior probability of cluster assignment. Finally, for the multivariate method the combined allele frequency data was converted to synthetic variables by PCoA of a Nei's genetic distance matrix as implemented in GenAlex v. 6.5 (Peakall and Smouse 2012) and clusters were visually inspected from PCoA dispersion plots.

# REFERENCES

- Anttila K., Couturier C.S., Øverli Ø., Johnsen A., Marthinsen G., Nilsson G.E. , Farrell A.P. 2014. Atlantic salmon show capability for cardiac acclimation to warm temperatures. *Nat Commun.* 5
- Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H., Xie D., Suchard M.A., Rambaut A. , Drummond A.J. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* 10: e1003537.
- Clement M., Posada D. , Crandall K.A. 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9: 1657-1659.
- Dodt M., Roehr J., Ahmed R. , Dieterich C. 2012. FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology.* 1: 895-905.
- Drummond A.J., Suchard M.A., Xie D. , Rambaut A. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29: 1969-1973.
- Edwards D.L. , Knowles L.L. 2014. Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proc. R. Soc. Lond., Ser. B: Biol. Sci.* 281: 20132765.
- Evanno G., Regnaut S. , Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14: 2611-2620.
- Faircloth B.C. 2008. msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol. Ecol. Resour.* 8: 92-94.
- Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., Eccles D., Li B., Lieber M., MacManes M.D., Ott M., Orvis J., Pochet N., Strozzi F.,

Weeks N., Westerman R., William T., Dewey C.N., Henschel R., LeDuc R.D., Friedman N. ,  
Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity  
platform for reference generation and analysis. *Nat. Protocols*. 8: 1494-1512.

Hubisz M.J., Falush D., Stephens M. , Pritchard J.K. 2009. Inferring weak population  
structure with the assistance of sample group information. *Mol. Ecol. Resour*. 9: 1322-1332.

Huelsenbeck J.P., Andolfatto P. , Huelsenbeck E.T. 2011. Structurama: Bayesian  
Inference of Population Structure. *Evol Bioinform Online*. 7: 55-59.

Peakall R. , Smouse P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population  
genetic software for teaching and research—an update. *Bioinformatics*. 28: 2537-2539.

Pino J., Ascunce M., Reed D. , Hugot J.-P. 2012. Development and characterization of  
microsatellite markers for the endangered Laotian rock-rat (*Laonastes aenigmamus*) using 454-  
sequencing technology. *Conservation Genetics Resources*. 4: 999-1002.

Pritchard J.K., Stephens M. , Donnelly P. 2000. Inference of Population Structure Using  
Multilocus Genotype Data. *Genetics*. 155: 945-959.

Rozen S. , Skaletsky H. 2000. Primer3 on the World Wide Web for general users and  
for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular  
Biology*. Humana Press, Totow, NJ.

Tamura K., Stecher G., Peterson D., Filipski A. , Kumar S. 2013. MEGA6: Molecular  
Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol*. 30: 2725-2729.

Zhang J., Kapli P., Pavlidis P. , Stamatakis A. 2013. A general species delimitation  
method with applications to phylogenetic placements. *Bioinformatics*. 29: 2869-2876.

# APPENDIX 3: SUPPLEMENTARY FIGURES

## *Ecology, molecules and colour: Multivariate species delimitation and conservation of Harlequin poison frogs*

Andres Posso-Terranova<sup>1-2\*</sup>, Jose A. Andres<sup>1</sup>

<sup>1</sup>*Department of Biology. University of Saskatchewan. 112 Science Pl. Saskatoon. SK. Canada*

<sup>2</sup>*Universidad Nacional de Colombia sede Palmira. A.A. 237. Palmira. Colombia*

*\*Correspondence to be sent to: 112 Science Pl. Saskatoon. SK. Canada. S7N5E2; E-mail:*

[andres.posso@usask.ca](mailto:andres.posso@usask.ca)

Figure S1. a) Morphological hierarchical clustering of individuals of the *O. histrionica* complex. Clusters with node support >82% are shown in red boxes. b) Individual morphologic variation within clusters and their respective sampling location. c) Posterior assignment probabilities (PP) for each individual based on DAPC analysis of 50 morphologic variables.

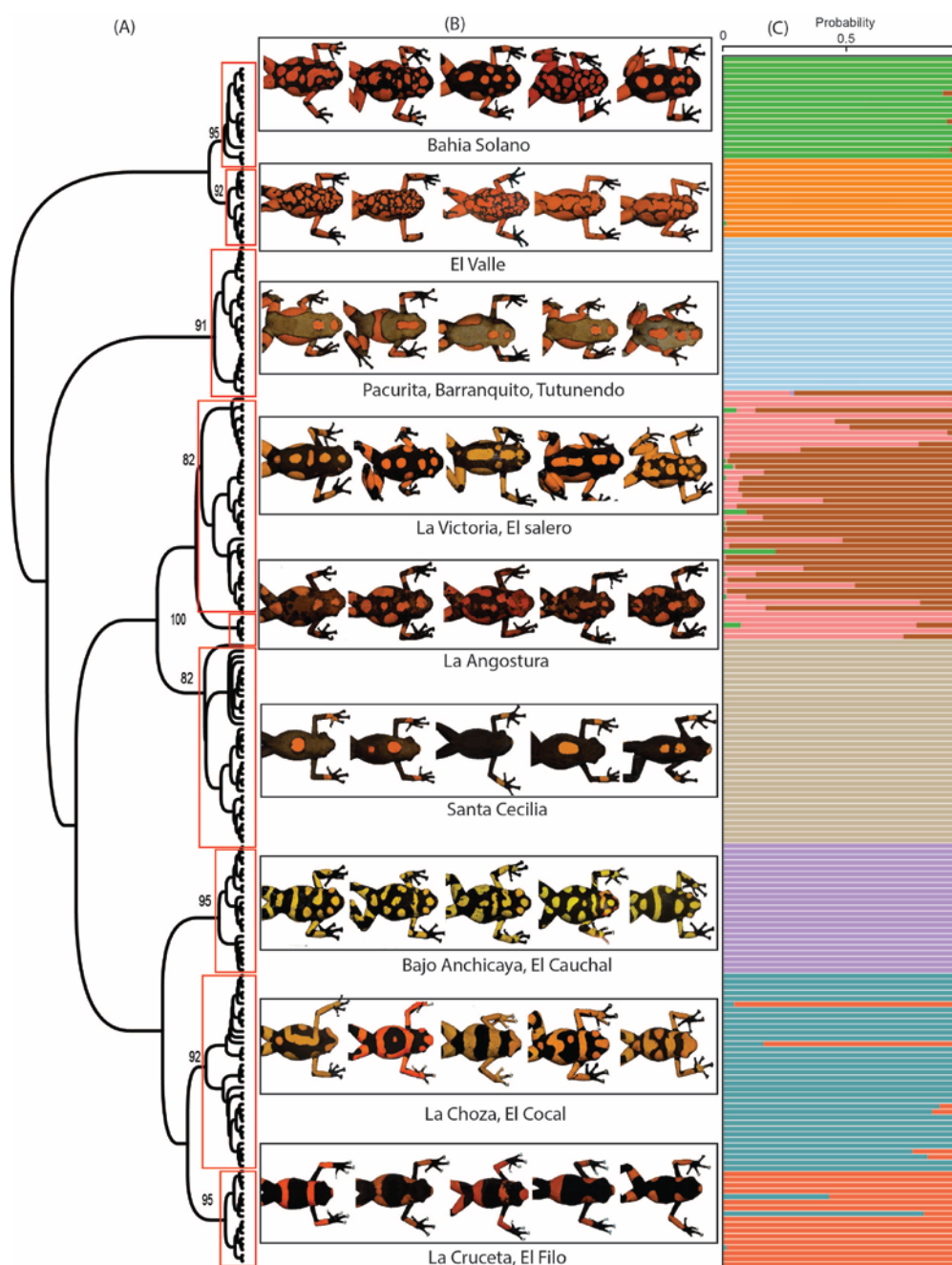


Figure S2. Environmental clusters by NMDS of climate data. Six potential climatic niches were found that correlates with geographic distribution. Clusters within coloured ellipses were significantly different ( $P < 0.0001$  after Bonferroni correction) and include 95% of the data. Clusters within the shadowed ellipse correspond to *O. lehmanni* populations.

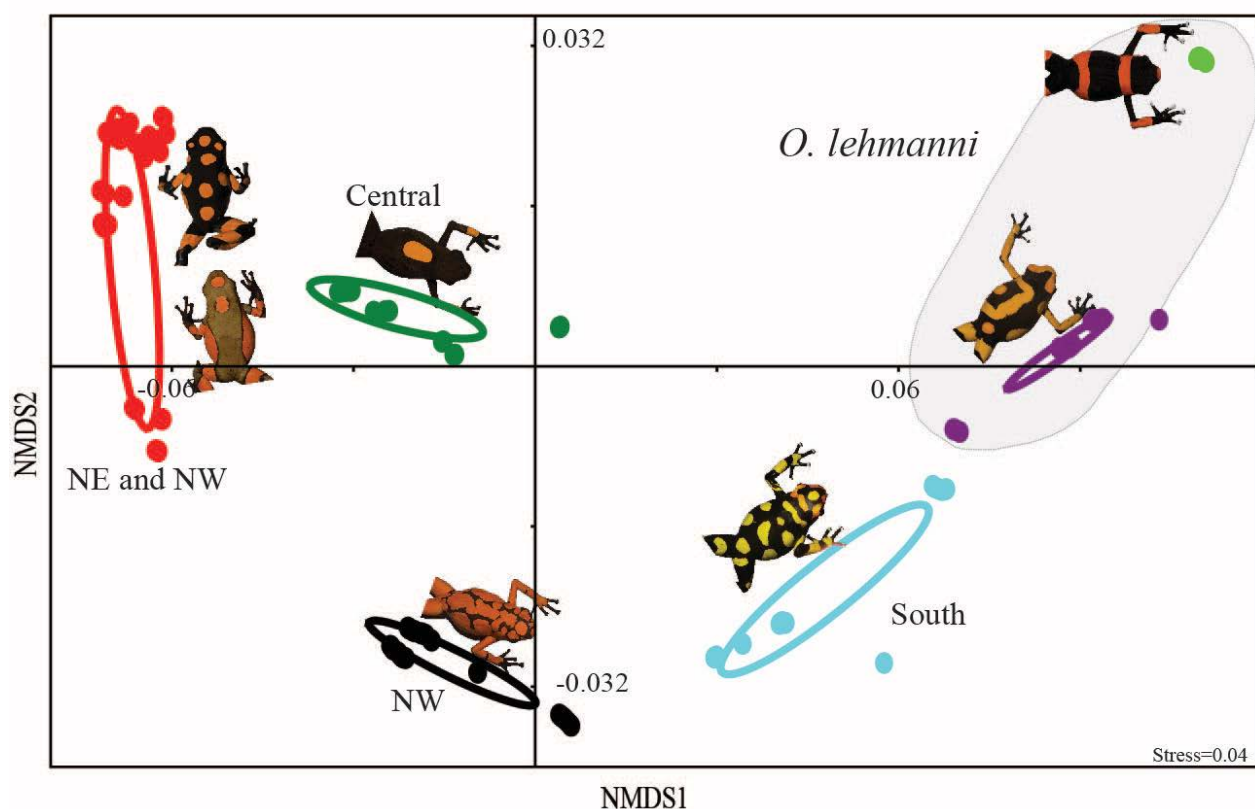
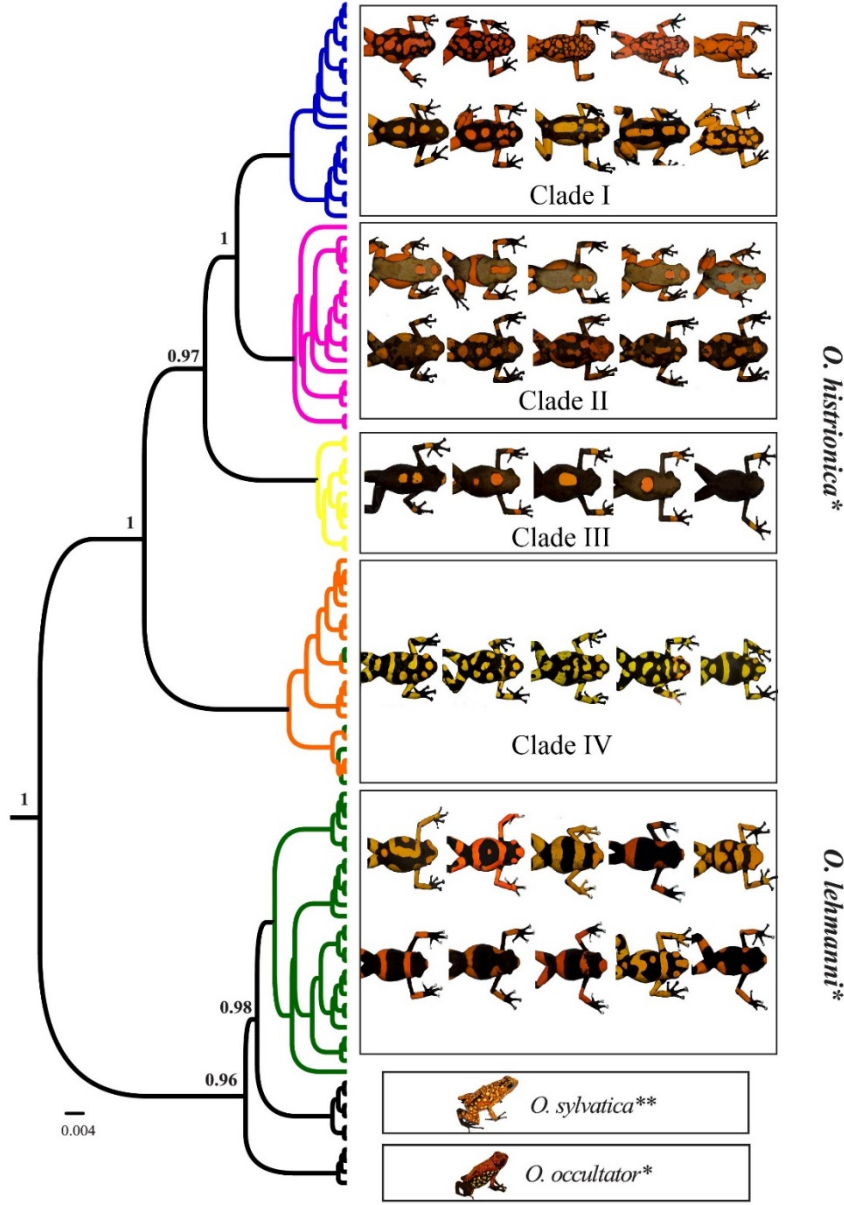


Figure S3. Mitochondrial species delimitation of the *O. histrionica* species complex. Numbers in nodes represent Bayesian posterior probabilities (speciation probability, SP) and only values higher than 0.95 are shown. The currently recognized *O. sylvatica* and *O. occultator* nominal species are included as control species. Species names in the figure correspond to the current taxonomic classification (\*: *sensu lato* Myers and Daly 1976. \*\*: Funkhouser 1956).



1035 Figure S4. Mitochondrial (COI) parsimony haplotype network of individuals of the *O. histrionica* species complex. Clade I-IV  
 1036 correspond to those mitochondrial clades of Figure S3 in Appendix 3. *O. sylvatica* and *O. occultator* are included as control species.

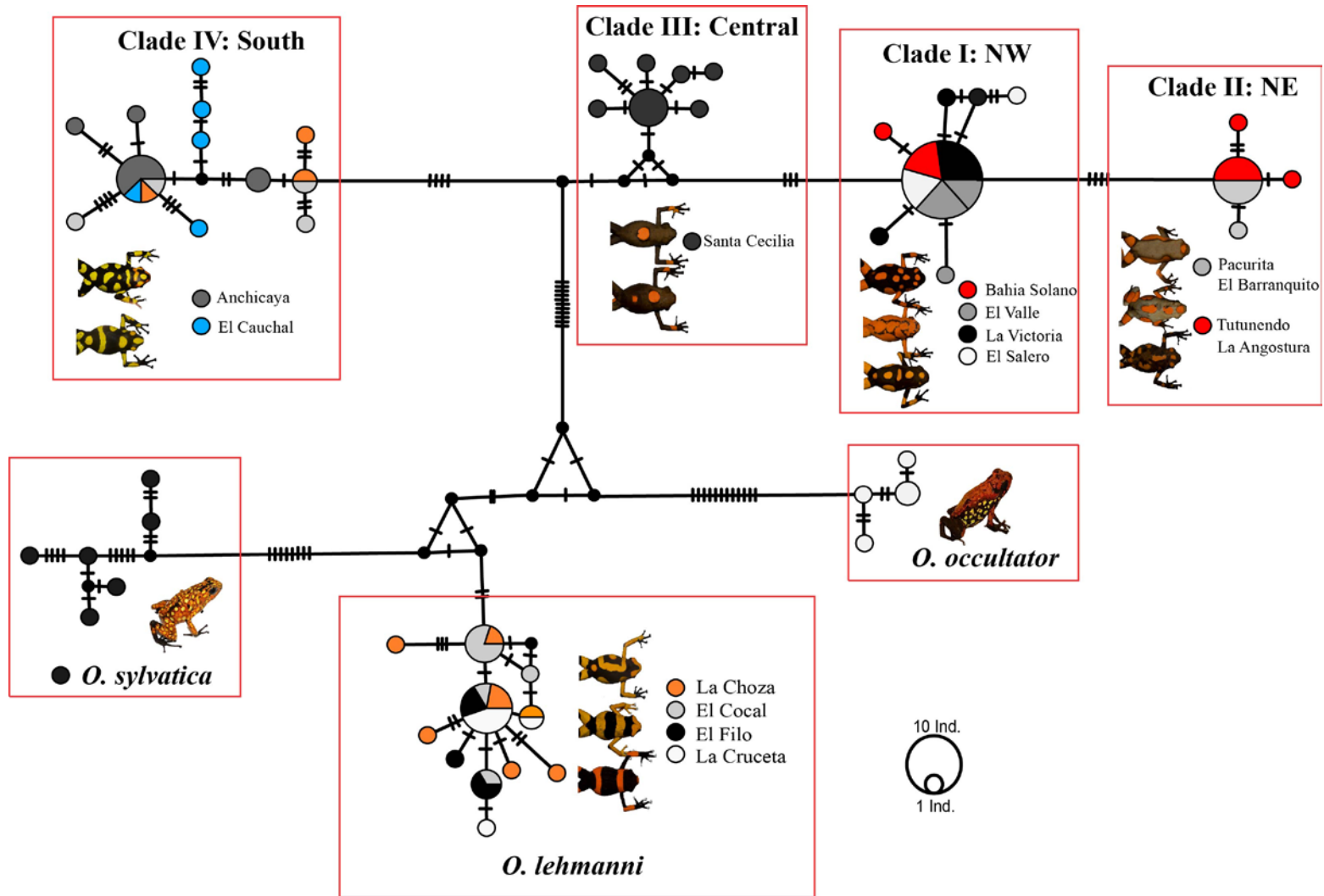


Figure S5. Genetic clusters by PCoA of allele frequency data from microsatellites (a) and transcriptome-based markers (b). Five cluster were detected by microsatellites markers while next generation sequencing data suggested the formation of six distinct clusters.

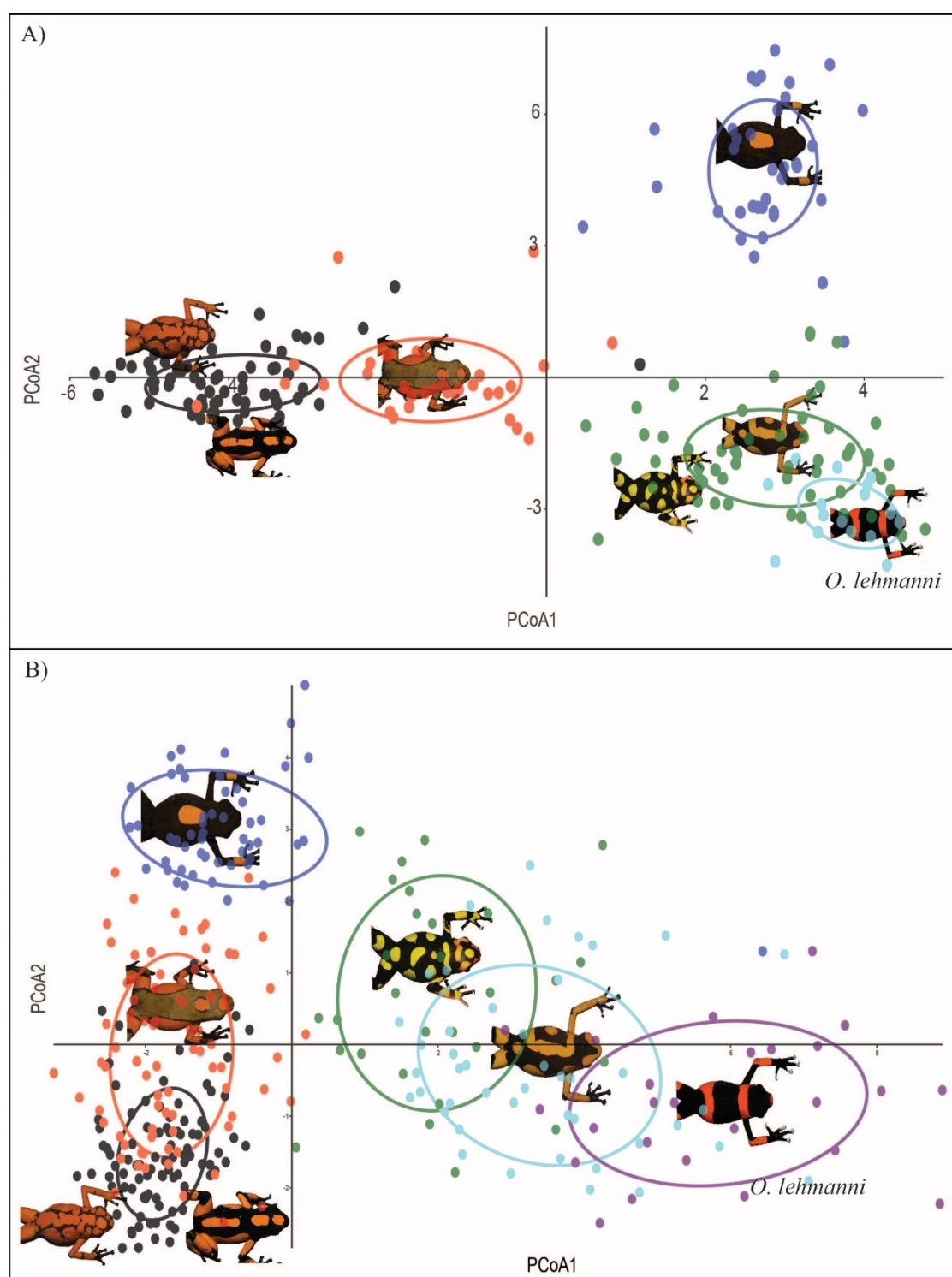


Figure S6. Gaussian clustering graph showing a plateau in BIC value at K=5. According to the *Mclust* analysis, the best model corresponded to an equal- covariance and diagonal distribution model (VEI) with five components or clusters.

