

Assessing phylogenetic dependence of morphological traits using co-inertia prior to investigate character evolution in Loricariinae catfishes

Raphaël Covain^{a,b,*}, Stéphane Dray^c, Sonia Fisch-Muller^a, Juan I. Montoya-Burgos^b

^a Département d'herpétologie et d'ichtyologie, Muséum d'histoire naturelle, 1 route de Malagnou, C.P. 6434, CH-1211 Genève 6, Switzerland

^b Département de zoologie et de biologie animale, Université de Genève, Sciences III, 30 quai E. Ansermet, CH-1211 Genève 4, Switzerland

^c Université de Lyon, université Lyon 1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, 43 boulevard du 11 novembre 1918, Villeurbanne F-69622, France

Received 15 June 2007; revised 4 December 2007; accepted 5 December 2007

Available online 25 December 2007

Abstract

With the increase of laboratory facilities, molecular phylogenies are playing a predominant role in evolutionary analyses. However, understanding the evolution of morphological traits remains essential for a comprehensive view of the evolution of a group. Here we present a new approach based on co-inertia analysis for identifying characters which variations are dependent to the phylogeny, a prerequisite for analyzing the evolution of characters. Our approach has the advantage of treating the full data set at once, including qualitative and quantitative variables. It provides a graphical output giving the contribution of each variable to the co-structure, allowing a direct discrimination among phylogenetically dependent and independent variables. We have implemented this approach in deciphering the evolution of morphological traits in a highly specialized group of Neotropical catfishes: the Loricariinae. We have first inferred a molecular phylogeny of this group based on the 12S and 16S mitochondrial genes. The resulting phylogeny indicated that the subtribe Harttiini was restricted to the single genus *Harttia*, and within the subtribe Loricariini, two sister subtribes were distinguished, Sturisomina (new subtribe), and Loricariina. Among Loricariina, the morphological groups *Loricariichthys* and *Loricaria* + *Pseudohemiodon* were confirmed. The co-inertia analysis highlighted a strong relationship between the morphological and the genetic data sets, and identified three quantitative and eight qualitative variables linked to the phylogeny. The evolution of quantitative variables was assessed using the orthogram method and showed a major punctual event in the evolution of the number of caudal-fin rays, and a more gradual pattern of evolution of the number of teeth along the phylogeny. The evolution of qualitative variables was inferred using ancestral states reconstructions and highlighted parallel patterns of evolution in characters linked to the mouth, suggesting co-evolution of the traits for adapting to divergent substrates.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Siluriformes; Loricariidae; Molecular phylogeny; Ribosomal genes; Morphology; Co-inertia analysis; Orthogram; Character mapping

1. Introduction

The increasing amount of robust molecular phylogenies, often based on multiple genes, is gradually setting aside the

concern of reconciling the phylogenies based on molecules and morphology. As a consequence, studies based on morphological traits incur the risk of a significant decline. The evolution of morphology, visible traits or phenotypes, remain however essential for the understanding of the evolutionary history of a group. A meaningful approach for a comprehensive understanding of the evolution of a given taxon is to first generate a well supported molecular phylogenetic tree and thereafter interpret the evolution of morphological traits in the light of this phylogeny. This

* Corresponding author. Address: Département d'herpétologie et d'ichtyologie, Muséum d'histoire naturelle, 1 route de Malagnou, C.P. 6434, CH-1211 Genève 6, Switzerland. Fax: +41 22 418 63 01.

E-mail addresses: Raphael.Covain@ville-ge.ch (R. Covain), dray@biomserv.univ-lyon1.fr (S. Dray), Sonia.Fisch-Muller@ville-ge.ch (S. Fisch-Muller), juan.montoya@zoo.unige.ch (J.I. Montoya-Burgos).

principle is followed in the character mapping methods available, either via parsimony or using stochastic models.

The evolution of traits (morphology, ecology, behavior...) may be plastic and stochastic or, to the contrary, traits may evolve according to a trend tightly linked to the phylogeny of the group. Only those characters displaying variations correlated to a given phylogeny may have their evolution interpreted in the light of that phylogeny. Therefore, testing for phylogenetic dependence is a first and unavoidable step to study the evolutionary relationship between a life trait and the phylogeny (Ollier et al., 2006). Several methods have been developed to detect phylogenetic dependence in comparative data (e.g. Felsenstein, 1985a; Cheverud et al., 1985; Gittleman and Kot, 1990; Harvey and Pagel, 1991; Lynch, 1991; Diniz-Filho et al., 1998; Pagel, 1999a; Abouheif, 1999; Blomberg et al., 2003; Ollier et al., 2006; for reviews see Rhoif, 2001; Blomberg et al., 2003). Probably the most popular tests were developed by Abouheif (1999) who modified two previously existing tests, the Test For Serial Independence (TFSI) (von Neumann et al., 1941), and the RUNS test (Sokal and Rohlf, 1995), which can detect phylogenetic autocorrelation for quantitative and qualitative variables, respectively. These tests have the advantage of needing only the topological structure of the tree, which allows the use of a wide range of tree sources (cladograms, phylograms, consensus trees, supertrees...). Each character under study must be however individually tested according to its quantitative or qualitative nature. Therefore, this procedure becomes fastidious when the tree topology is complex, and when the number of traits under study is important.

After testing for the phylogenetic dependence of the character, their evolution can be reconstructed along the given phylogenetic tree. Several methods have been proposed for reconstructing ancestral states or for mapping characters on the tree in order to test hypotheses about the evolution of the selected characters (Schluter et al., 1997; Pagel, 1999b; Huelsenbeck et al., 2003; Pagel et al., 2004). They provide a graphical view of the best possible reconstruction of the evolution of the trait assuming an implicit process of evolution (Maximum Parsimony mapping) or an explicit model of evolution (Maximum likelihood, Stochastic, and Bayesian ancestral state reconstructions). Alternatively, the orthogram method developed by Ollier et al. (2006) represents a relevant approach that detects and characterizes phylogenetic dependence, and at the same time highlights different pattern of evolution along a phylogenetic tree. However, this method can treat only quantitative data.

Here we present a new approach to detect phylogenetic dependence of characters of different type in a fast, simultaneous and reliable way. Our approach is based on the co-inertia analysis (Dolédéc and Chessel, 1994) to assess the common information present within a genetic and a morphological data set. It allows identifying the morphological variables that possess a strong covariation with the phylog-

eny in a pool of many different morphological variables, either quantitative or qualitative. The strength of our approach is to provide a direct graphical interpretation of the explored data sets. The phylogenetically informative morphological variables can be easily detected as well as the variables unlinked to the phylogeny. This last class of variables is discarded from further analyses as they represent evolutionary “noise”. The co-structure can be represented in a phylogram summarizing the total amount of convergent information present in both molecular and morphological data sets.

We have implemented our new approach for identifying morphological characters varying dependently from the phylogeny, and have reconstructed their evolution in a group of highly derived catfishes, the Loricariinae. Our work has therefore started by the reconstruction of a robust molecular phylogeny of this group based on partial 12S and 16S mitochondrial genes. The Loricariinae represents a diversified subfamily among the large Neotropical catfish family Loricariidae, or armored catfish. Loricariids are characterized by a modification of the mouth structure into a sucker disk, by a body covered with bony plates, and by a unique pair of maxillary barbels. Loricariids have undergone an evolutionary radiation at a subcontinental scale, from Costa Rica to Argentina, which has been compared to that of the Cichlidae of the Great Lakes of the Rift Valley in Africa (Schaefer and Stewart, 1993). Among Loricariids, members of the Loricariinae subfamily are characterized by a long and depressed caudal peduncle and by the absence of an adipose fin. They live stuck to the substrate and show marked variations in body shape due to the various habitats colonized, from lotic to lentic systems, on inorganic or organic substrates. Some groups have numerous teeth, pedunculated, and organized in comb, while other groups have few teeth or even no teeth on premaxillae. These later are often strongly differentiated, and can be bicuspid straight and thick, spoon-shaped, reduced in size or very long. An important diversity in lips structure, which can be strongly papillose, filamentous or smooth, also characterizes this subfamily (Covain and Fisch-Muller, 2007).

Modern classification of Loricariinae started with Isbrücker (1979) who proposed a subdivision into four tribes and eight subtribes on the basis of morphology. These included the Loricariini (comprising six subtribes), the Harttiini (including two subtribes), the Farlowellini, and the Acestridiini. Schaefer (1987) established the monophyly of the Loricariinae on the basis of morphological data, and placed the Acestridiini into another subfamily, the Hypoptopomatinae (Schaefer, 1991). Rapp Py-Daniel (1997) confirmed the monophyly of the subfamily and of the Loricariini *sensu* Isbrücker (1979), and redefined the Harttiini comprising former Farlowellini. Further on, Montoya-Burgos et al. (1998) proposed the first molecular phylogeny of the family Loricariidae based on mitochondrial markers. They confirmed the position of the Farlowellini nested within Harttiini and provided the first

evidence for a splitting of the subfamily into two lineages, *Harttia*, on one side and all other Loricariinae on the other side. They also found that *Farlowella* and *Sturisoma* form the sister group to the Loricariini. In a recent work, Covain and Fisch-Muller (2007) recognized 203 valid species distributed in 30 genera. Based on external morphological analyses, they partly confirmed the splitting of the subfamily into two tribes, the Harttiini and the Loricariini, and proposed four morphological groups within the Loricariini: (1) the *Pseudohemiodon* group, (2) the *Loricaria* group, (3) the *Rineloricaria* group, and (4) the *Loricariichthys* group. The morphological data set of Covain and Fisch-Muller (2007) was used here to test our new approach for detecting morphological characters linked to the phylogeny. Then, the evolution of the retained characters has been inferred.

2. Materials and methods

2.1. Taxonomic sampling

The molecular phylogeny was established for 14 genera totalizing 20 species of Loricariinae. Taxonomic sampling was chosen in a way to include at least one representative of the different morphological groups defined in Covain and Fisch-Muller (2007). The outgroup was chosen in another subfamily of Loricariidae. The list of material used for this study is given in Table 1. The analyzed samples came from the tissue collection of MHNG, Geneva, and the sequences were deposited in GenBank. The morpholog-

ical characters analyzed in this study are presented in Covain and Fisch-Muller (2007) and summarized at the end of Table 3.

2.2. DNA extraction, amplification and sequencing

Tissue samples were preserved in 80% ethanol and stored at -20°C . Total genomic DNA was extracted with the DNeasy Tissue Kit (Qiagen) following the instructions of the manufacturer. The PCR amplification of partial 12S and 16S were carried out using the *Taq* PCR Core Kit (Qiagen). The primers used were: An12S-2D 5'-GCC AGC TTA CCC TGT GAA GG-3' and H3059 5'-CCG GTC TGA ACT CAG ATC ACG T-3'. The amplifications were performed in a total volume of 50 μl , containing 5 μl of 10 \times reaction buffer, 1 μl of dNTP mix at 10 mM each, 1 μl of each primer at 10 μM , 0.2 μl of *Taq* DNA Polymerase equivalent to 1 U of Polymerase per tube, and 1–4 μl of DNA. Cycles of amplification were programmed with the following the profile: (1) 3 min at 94°C (initial denaturing), (2) 35 s at 94°C , (3) 30 s at $52\text{--}54^{\circ}\text{C}$, (4) 2 min at 72°C , and (5) 5 min at 72°C (final elongation). Steps 2–4 were repeated 27–39 times according to the quality and concentration of DNA. PCR products were purified with the High Pure PCR Product Purification Kit (Roche). Sequencing reactions were performed with the Big Dye Terminator Cycle Sequencing Ready Reaction 3.1 Kit (Applied Biosystems) following instructions of the manufacturer, and were loaded on an automatic sequencer 3100-Avant Genetic Analyzer (Applied Biosystems, Perkin-Elmer). To obtain

Table 1
Taxa list, specimen and sequence data for the 20 species of Loricariinae, and outgroup analyzed in this study

Species	Catalog number	Field number	Locality	mt 12S+16S bases + GenBank No.
<i>Crossoloricaria venezuelae</i>	INHS 35467	VZ 049	Venezuela, Rio Santa Rosa	1687 EU310444
<i>Dasylicaria tuyrensis</i>	MHNG 2674.052	PA00-012	Panama, Rio Ipeti	1687 EU310445
<i>Farlowella platoryncha</i>	MHNG 2588.093	PE96-071	Peru, Rio Ucayali	1700 EU310446
<i>Farlowella oxyrryncha</i>	MHNG 2588.064	PE96-022	Peru, Rio Tambopata	1700 EU310443
<i>Harttia guianensis</i>	MHNG 2543.016	GF00-351	French Guiana, Marouini River	1704 EU310447
<i>Hemiodontichthys acipenserinus</i>	MHNG 2651.012	GY04-15	Guyana, Rupununi River	1688 EU310448
<i>Lamontichthys stibaros</i>	MHNG 2677.039	MUS 208	Peru, aquarium trade, Rio Itaya according to the exporter	1701 EU310449
<i>Limatulichthys griseus</i>	MHNG 2651.013	GY04-18	Guyana, Rupununi River	1689 EU310450
<i>Loricaria clavipinna</i>	MHNG 2640.044	PE98-002	Peru, Rio Putumayo	1693 EU310451
<i>Loricaria parnahybae</i>	MHNG 2602.067	BR98-274	Brazil, Rio Parnahyba	1689 EU310452
<i>Loricariichthys maculatus</i>	MHNG 2621.042	SU01-56	Surinam, Sarramacca River	1694 EU310453
<i>Loricariichthys microdon</i>	MHNG 2650.054	GY04-12	Guyana, Rupununi River	1694 EU310454
<i>Metaloricaria paucidens</i>	MHNG 2677.086	GF00-083	French Guiana, Marouini River	1703 EU310455
<i>Planiloricaria cryptodon</i>	MHNG 2677.038	MUS 211	Peru, aquarium trade, Rio Itaya according to the exporter	1690 EU310456
<i>Rineloricaria platyura</i>	MHNG 2651.009	GY04-83	Guyana, Rupununi River	1692 EU310458
<i>Rineloricaria lanceolata</i>	MHNG 2588.059	PE96-011	Peru, Rio Tambopata	1689 EU310457
<i>Rineloricaria</i> sp. Tocantins	UFRJ batch 6-EF4	BR 1114	Brazil, Rio Maranhão	1691 EU310459
<i>Sturisoma nigrirrostrum</i>	MHNG 2588.055	PE96-001	Peru, Rio de las Piedras	1706 EU310460
<i>Sturisoma monopelte</i>	MHNG 2651.033	GY04-187	Guyana, Sawarab River	1707 EU310461
<i>Sturisomatichthys citirensis</i>	MHNG 2676.004	PA97-032	Panama, Rio Tuyra	1703 EU310462
<i>Ancistrus cirrhosus</i> ^a	MHNG 2645.037	MUS 202	Argentina, Rio Uruguay	1698 EU310442

The abbreviations of institutions follow Leviton et al. (1985).

^a Outgroup.

the complete sequence of the amplified region, an internal primer was designed: Lor12S-3D 5'-CCT CGT ACC TTT TGC ATC ATG-3'.

2.3. Sequence alignment and phylogenetic reconstruction

The DNA sequences were edited and assembled using BioEdit 7.0.1 (Hall, 1999). Alignment was realized using ClustalW (Thompson et al., 1994) and optimized by eye. Regions with ambiguous alignments were excluded from the analyses. Phylogenetic reconstructions were performed with PAUP* 4.0b10 (Swofford, 2003) following three methods: Neighbor-joining (NJ) (Saitou and Nei, 1987), Maximum parsimony (MP), and Maximum likelihood (ML) (Felsenstein, 1981). The model of substitution that best fitted the data was determined by Modeltest 3.06 (Posada and Crandall, 1998). The best fit model was used for the ML tree reconstructions and to correct the distance matrix for NJ analysis. Robustness of the results was estimated by resampling the data set with the nonparametric bootstrap (Efron, 1979) following Felsenstein's (1985b) methodology with 1000 replicates for NJ and MP methods, and with 200 replicates for ML method. Alternative topologies were investigated using the Shimodaira and Hasegawa (SH) test (1999) that allows comparison between the best ML tree and an alternative topology (Goldman et al., 2000). SH tests were performed using PAUP* 4.0b10 with 2000 RELL replicates.

2.4. Co-structure analysis between morphology and genetics (CIA)

To highlight a possible relationship between the information of morphological data provided in Covain and Fisch-Muller (2007), and the one provided by our molecular data, both data sets were analyzed by co-inertia analysis (CIA) (Dolédéc and Chessel, 1994). Taxonomical sampling was modified in order to keep the same 14 genera present in both data sets. For this, when more than one representative of a given genus was present in the molecular data set, all except one were pruned, that is to say: *Loricaria parnahybae*, *Loricariichthys microdon*, *Rineloricaria* sp. Tocantins, *Rineloricaria platyura*, *Farlowella oxyrryncha*, and *Sturisma nigrirostrum*. Because morphological characters analyzed here are homogenous within genera, only generic names are given in the CIA and subsequent analyses. Molecular data were converted into a distance matrix corrected according to the model of substitution re-estimated by Modeltest 3.06 (Posada and Crandall, 1998). This matrix was then rendered Euclidian using Lingoes's (1971) method. Principal Coordinate Analysis (PCoA) (Gower, 1966) was performed on this corrected distance matrix to reveal the structuring of the genetic data set. This analysis provide a tree free representation of the phylogenetic data set onto axes, where the pairwise distances between genera are exactly the genetic pairwise distances of the matrix. Morphological data were analyzed by Hill

and Smith Analysis (HSA) (1976) to reveal their structuring. The HSA consists in a Principal Component Analysis (PCA) of a table mixing quantitative and qualitative variables. These two simple analyses (PCoA and HSA) were then coupled by a CIA to study a possible co-structure of each type of information. This analysis describes the common structure of both tables measured on the same statistical units (herein the genera). The mathematical model of CIA is given in Dolédéc and Chessel (1994) and in Dray et al. (2003). Results of the CIA consist in two sets of scores (morphological in table $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_p]$, and genetic in table $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_p]$) of maximum covariance (i.e., maximization of product: $\text{cov}(\mathbf{a}, \mathbf{b}) = \text{var}(\mathbf{a})^{1/2} \cdot \text{var}(\mathbf{b})^{1/2} \cdot \text{cor}(\mathbf{a}, \mathbf{b})$). Thus, the CIA maximizes a compromise between the structure of morphological information ($\text{var}(\mathbf{a})$), the structure of phylogenetic information ($\text{var}(\mathbf{b})$), and their link ($\text{cor}(\mathbf{a}, \mathbf{b})$). To assess the significance of the CIA results, a Monte-Carlo permutation test was computed on the RV coefficient (Robert and Escoufier, 1976). This procedure tests the link between two tables by permuting simultaneously the rows of both tables. Then, the common structure of the tables has been extracted by computing Euclidian distances between genera using the CIA scores of both tables. This new distance matrix was submitted to a hierarchical analysis using Fitch and Margoliash (FM) (1967) algorithm, resulting in a phylogram showing the relationships that are strictly congruent between both data sets. The FM phylogram was calculated using the global optimization criterion with negative branch lengths allowed, and 999 random permutations for the input order of taxa. The most external group within Loricariinae, according to the results of the phylogeny, was used to root the tree. Multivariate analyses were conducted using ADE-4 software (Thioulouse et al., 1997), and the phylogram was calculated with the Fitch module in PHYLIP 3.66 package (Felsenstein, 2004).

2.5. Identification of phylogenetically dependent variables

Phylogenetically dependent variables are given by the CIA as those that show the strongest covariation with the phylogeny (i.e., variables with the most important absolute contributions, and the longest vectors when projected onto axes). In order to have an independent confirmation of the results obtained by our approach (CIA results), quantitative variables were submitted to the Test For Serial Independence (TFSI) (von Neumann et al., 1941), and qualitative variables to the RUNS test (Sokal and Rohlf, 1995) following Abouheif's (1999) procedures, as implemented in Phylogenetic Independence version 2.0 (Reeve and Abouheif, 2003). These tests against phylogenetic autocorrelation allow to detect self-similarities among adjacent (ordered) observations. The computation of the statistics requires a topology and the value of a trait for the tips. An average statistics (C—mean for TFSI test, and Runs—mean for RUNS test) is calculated for a random representative sampling of all possible branch swap-

ping. This average statistics (observed) was then compared to a null hypothesis sampling distribution of randomized average statistics obtained by calculating an average statistics on a representative sampling of all possible branch swapping on topologies obtained after randomly shuffling the tips of the original topology. The tree topology used here corresponds to the ML tree calculated from the molecular data set used for the CIA analysis. Average statistics were estimated after 10,000 random permutations of tips around nodes and compared to the randomized average statistics obtained after 10,000 random shuffling of tips.

2.6. Analysis of character evolution

We first analyzed quantitative phylogenetically dependent variables by using a canonical procedure that allowed decomposition of their variance along the phylogenetic tree (Ollier et al., 2006). Prior to the variance analysis, an orthonormal vectorial basis was constructed to represent the topology of the phylogenetic tree. This tree (root, nodes, and tips) was described by a set of ordered dummy variables corresponding to a tip or a node (and its descendant tips). These dummy variables were then orthonormalized to obtain the orthonormal basis. Vectors of this basis were linear combinations of the dummy variables ranked according to the initial ranking of the dummy variables. This allowed the interpretation of the successive vectors in terms of decreasing phylogenetic dissimilarities. Then, a linear regression was performed with the centered and standardized trait variable as response variable, and the orthonormal basis as explanatory variables. Regression coefficients allowed reconstructing the trait variable, and squared coefficients provided variance decomposition of the trait onto the orthonormal basis. The plotting of the squared coefficients and of the cumulative squared coefficients provides two graphical tools called orthogram and cumulative orthogram (Ollier et al., 2006). Four permutation procedures associated to orthograms are used to test the null hypothesis of phylogenetic independence. These procedures are based on different statistics and consider different alternative hypotheses. The R2Max statistics was used to test against the alternative hypothesis that one vector explained a significant part of the trait variance (punctual effect). SkR2k was used to test against the alternative hypothesis that vectors near the tips (or the root) explained a significant part of the trait variance. SkR2k is high when the trait variance was rather explained by last vectors (towards tips) and low when explained by first vectors (towards root). D_{\max} is a Kolmogorov–Smirnov-like statistic and was used to test if the vector of squared coefficients may be an ordered random sample of the uniform distribution on (0,1). D_{\max} was used to test against the alternative hypothesis that some successive vectors explained a significant part of the trait variance. Finally, SCE is a measure of the average local variation of the orthogram and tests against the alternative hypothesis that there are significant differences in variance explained by

vectors and their neighbors (precedent or subsequent). Distribution of the statistics under the null hypothesis and confidence limits of (cumulative) orthograms were built using 9999 random permutations of the trait values. Orthograms and associated tests (Ollier et al., 2006) were conducted using ade4 package (Chessel et al., 2004) in R 2.4.0 (Ihaka and Gentleman, 1996).

We then analyzed qualitative phylogenetically dependent variables using Maximum likelihood ancestral state reconstruction as implemented in the Stochchar 1.1 package (Maddison and Maddison, 2006a), in Mesquite 1.12 (Maddison and Maddison, 2006b). This method estimates for each node the ancestral states that maximize the probability of observing the different character states in the terminal taxa, given a stochastic model of evolution. We used the Lewis's (2001) Mk model which assumes equal rates of change from one state to another, for forward as well as backward rates. This global rate was directly estimated from the data, and scaled using our ML tree branch lengths.

3. Results

3.1. Phylogenetic analysis of the subfamily Loricariinae

We sequenced the partial 12S and 16S mitochondrial genes of 20 Loricariinae species representing 14 genera. The sequence alignment included 1739 positions from which 238 corresponded to the 12S rRNA gene, 73 corresponded to the tRNA Val gene, and 1428 belonged to the 16S rRNA gene. The model GTR+G+I (Tavaré, 1986) fitted our data the best as indicated by Modeltest.

MP, ML and NJ analyses lead to comparable tree topologies. The MP tree (not shown) included 1572 steps (CI = 0.513; RI = 0.528). In the ML tree ($-\ln L = 9414.26784$), shown in Fig. 1, the Loricariinae was split into two lineages: the Harttiini (Fig. 1, clade 1), including the single genus *Harttia*, and the Loricariini (Fig. 1, clade 2). The Loricariini was divided into two clades, the Sturisosmina (new subtribe) (clade A), and the Loricariina (clade B). The genus *Lamontichthys* was the first diverging genus within the clade Sturisosmina, a position strongly supported by bootstrap values (100/100/99). The remaining Sturisosmina representatives were then split into two lineages, one comprising a species of *Farlowella* and the two representatives of *Sturisoma*, and a second comprising another species of *Farlowella* and *Sturisomatichthys*. However, the first group was not found in NJ and MP tree topologies, and the node giving *Farlowella platoryncha* as sister genus of *Sturisomatichthys* was weakly supported by the same two methods. The polyphyly of *Farlowella* was assessed by a one-tailed SH test with, as alternative topology, the enforced monophyly of *Farlowella* as sister group of *Sturisosmina* and *Sturisomatichthys* (without hypotheses concerning their interrelationships). The result indicated that the monophyly of *Farlowella* was significantly rejected ($p = 0.0495$). Our phylogenetic reconstructions all showed the monophyly of the subtribe Loricariina yet only the

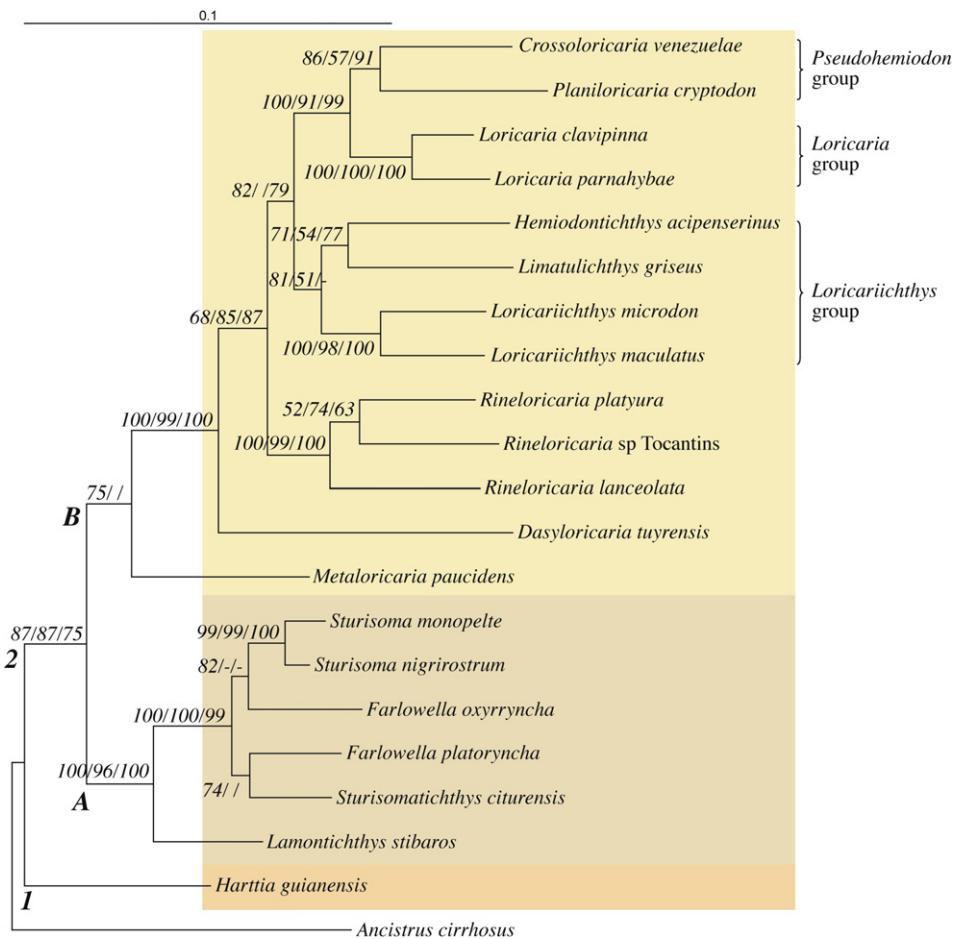


Fig. 1. Maximum likelihood tree of the Loricariinae including 14 genera and 20 species inferred from the analysis of partial 12S and 16S gene sequences ($-\ln L = 9414.26784$). The best fit substitution model was GTR+G+I with the following parameter values: base frequencies: $f_A = 0.3686$, $f_C = 0.2361$, $f_G = 0.1856$, $f_T = 0.2096$; substitutions rates $[A \leftrightarrow G] = 11.5316$, $[C \leftrightarrow T] = 37.5424$, $[A \leftrightarrow C] = 4.5376$, $[A \leftrightarrow T] = 4.8089$, $[C \leftrightarrow G] = 0.0089$, $[G \leftrightarrow T] = 1$; proportion of invariable sites $I = 0.4527$; gamma shape parameter: $\alpha = 0.6563$. Numbers above branches indicate bootstrap supports above 50 for ML, MP, and NJ trees, respectively. Sign (–) indicates that the node was not found in some topologies. 1, Harttiini; 2, Loricariini; A, Sturisomina; B, Loricariina. Scale indicates the number of substitution per site as expected by the model.

ML analysis gave good bootstrap support. Within the Loricariina, *Metaloricaria* branched at the base of the clade. The sister group of *Metaloricaria* was strongly supported (100/99/100) with *Dasylicaria* as sister genus of all remaining representatives of the subtribe. The sister group of *Dasylicaria* was then split into two clades: the first corresponding to *Rineloricaria* representatives, and the second comprising the remaining genera studied herein. This last group contained two clades with on one hand representatives of the *Loricariichthys* group, and on the other hand representatives of the *Loricaria* group plus *Crossoloricaria* and *Planiloricaria*, these last two genera belonging to the so called *Pseudohemiodon* group (*sensu* Covain and Fisch-Muller, 2007). Within the *Loricariichthys* group, the nominal genus occupied a sister position to *Hemiodontichthys* and *Limatulichthys*. The NJ tree showed however an unresolved polytomy among these three genera. Within the *Loricaria-Pseudohemiodon* clade, all methods placed *Loricaria* representatives as the sister lineage to the *Pseudohemiodon* group.

3.2. Co-structure analysis of molecular and morphological data

In order to highlight the co-structure of the morphological data as compared to the molecular data, the CIA was performed on the restricted data sets comprising the same taxonomic sampling. The new genetic distance matrix was calculated using a re-estimated model of substitutions which characteristics were: GTR+G+I: base frequencies: $f_A = 0.3563$, $f_C = 0.2376$, $f_G = 0.1970$, $f_T = 0.2091$; substitutions rates $[A \leftrightarrow G] = 11.7966$, $[C \leftrightarrow T] = 42.5481$, $[A \leftrightarrow C] = 5.7480$, $[A \leftrightarrow T] = 4.6241$, $[C \leftrightarrow G] = 0.0714$, $[G \leftrightarrow T] = 1$; proportion of invariable sites $I = 0.4521$; gamma shape parameter: $\alpha = 0.5508$. A first assessment of the relationships between morphology and genetics was performed using the RV coefficient, and showed a strong and significant correlation between both data sets ($p < 0.0001$; $RV = 0.832$). The projection of inertia axes of the PCoA of the genetic data and of the HSA of the morphological data onto co-inertia axes (Fig. 2c) placed

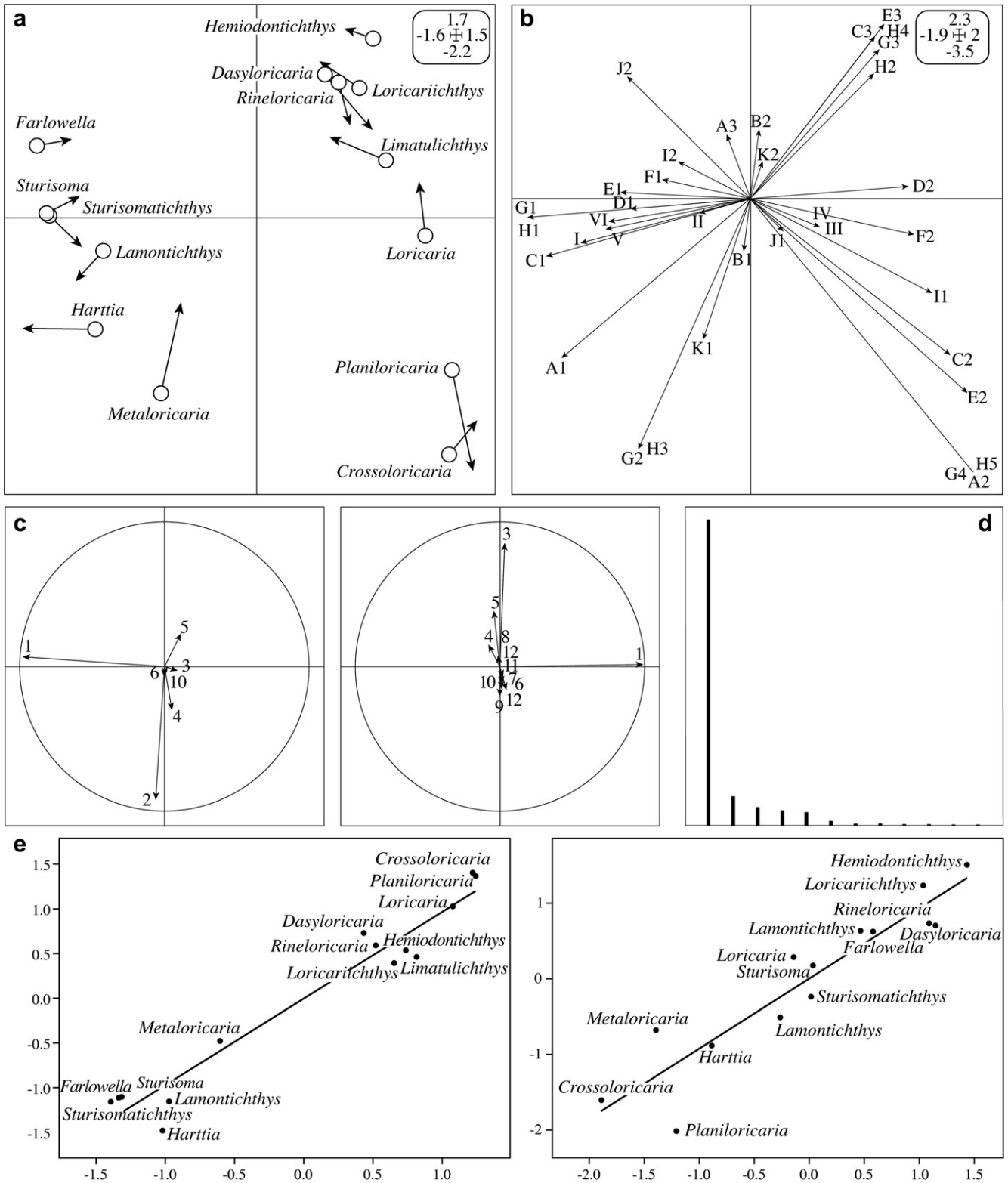


Fig. 2. Co-inertia analysis. Projection of data coordinates of preliminary analyses (PCoA of genetic data and HSA of morphological data) onto axes 1–2 of the co-inertia analysis. (a) Normalized individuals' scores in the co-inertia plan: genetic data (origin of arrows) and morphological data (extremity of arrows). (b) Coordinates of morphological variables in the co-inertia plan (numbered as in Table 3). (c) Projection of inertia axes of simple analyses onto co-inertia axes: inertia axes of PCoA of genetic data (left); inertia axes of HSA of morphological data (right). (d) Eigenvalues of co-inertia analysis. (e) Bivariate plots of correlations of normalized individuals' scores (genetic data in abscise and morphological data in ordinate) for the first (left) and second (right) co-inertia axes.

Table 2
Main characteristics of co-inertia analysis

Co-inertia axes	Covariance	Variance 1	Variance 2	Correlation	Inertia 1	Inertia 2
1	0.04601	0.004599	0.487	0.9722	0.004643	0.4971
2	0.01419	0.0009029	0.2549	0.9295	0.001139	0.2752

Covariance: covariance between both systems of coordinates of co-inertia analysis (maximized by the analysis).

Variance 1: inertia of the genetic data projected onto co-inertia axes.

Variance 2: inertia of the morphological data projected onto co-inertia axes.

Correlation: correlation between both systems of coordinates of co-inertia analysis.

Inertia 1: maximum inertia projected onto axes of the simple analysis of genetic data (eigenvalues of PCoA).

Inertia 2: maximum inertia projected onto axes of the simple analysis of morphological data (eigenvalues of HSA).

plan 1–3 of the genetic data analysis in relation to plan 1–2 of the morphological data analysis. Thus, CIA found that both axes 1 were associated, and that axis 3 of the genetic data was associated to axis 2 of the morphological data. The first plan of CIA accounted for 85.84% of the total co-structure (78.47% for axis 1 and 7.37% for axis 2) (Fig. 2d). CIA characteristics are given in Table 2. Covariance associated to the first axis was almost four times greater than the one associated to other axes. Co-inertia plan 1–2 was of the same quality than plans 1–3 and 1–2 of the initial analyses. The inertia projected onto co-inertia axes was equivalent to the one projected onto inertia axes of the initial analyses: 99.05% (0.004643/0.004599) of the genetic data structure and 97.96% (0.487/0.4971) of the morphological data structure was recovered by axis 1 of the co-structure analysis. Correlations between both data sets were also very high (more than 0.97 on the first co-inertia axis and 0.92 on the second one). Axis 1 of co-inertia analysis defined the tribal rank of the subfamily and split Harttiini, Sturisomina, and *Metaloricaria* on one hand and Loricariina on the other hand. Axis 2 defined the generic rank and ordered the genera according to their morphological and genetic proximity. The projection of morphological and genetic data coordinates onto co-inertia axes is given in Fig. 2. Superimposition of both sets of coordinates, after normalization for scaling (Fig. 2a), allowed to display the most important differences between genetic (origin of arrows) and morphological data (extremity of arrows). These differences mainly concerned the generic rank (axis 2) and particularly genera *Planiloricaria*, *Dasylicaria*, and *Metaloricaria* among Loricariini, and *Harttia* concerning Harttiini. The co-structure highlighted concerned thereby the tribal rank and the grouping of genera in some groups (morphological and genetic) which were Sturisomina and the *Loricaria*–*Pseudohemiodon* group. The position of *Hemiodontichthys* was also consistent between both representations, whereas *Metaloricaria* was placed together with Harttiini and Sturisomina. In the same manner, Sturisomina was grouped with Harttiini. The morphological variables involved the most in the co-structure were identified by the projection of the variables onto the first co-inertia plan (Fig. 2b) and by the inertia analysis. Absolute contributions of the variables to the axes are given in Table 3. Concerning axis 1 (tribal rank), these variables corresponded, in decreasing order, to: mouth and tooth shapes (variables G and H which con-

tributed to 12.38% of the explained inertia by this axis), the absence or presence of deep or weak postorbital notches (variable C, 12.04% of the explained inertia), the number of caudal-fin rays (variable I, 10.72% of the explained inertia), the lip structure (variable E, 8.9% of the explained inertia), the number of premaxillary and dentary teeth (variables V and VI, respectively, 7.84% and 7.49% of the explained inertia), the presence or absence of predorsal keels (variable D, 7.16% of the explained inertia), the presence or absence of fringed barbels (variable F, 5.63% of the explained inertia), and the characteristics of the maxillary barbels (variable I, 4.95% of the explained inertia). Concerning axis 2 (generic rank), the strongest contributions were registered for: the tooth and mouth shape (variables H and G which contributed, respectively, to 21.16% and 20.94% of the explained inertia by this axis), the absence or presence of deep or weak postorbital notches (variable C, 13.47% of the explained inertia), the absence or presence of a complete or incomplete abdominal cover (variable A, 13.39% of the explained inertia), and the lip's structure (variable E, 12.13% of the explained inertia). Bivariate plots of the individuals' normalized scores concerning co-inertia axes 1 and 2 (Fig. 2e) showed a better ordination of the genera along first axis, knowing the phylogenetic tree topology. Along axis 2, representatives of the *Pseudohemiodon* group were indeed grouped with Harttiini and *Metaloricaria*, whereas *Loricaria* was placed among Sturisomina. A part of the incongruent information (background noise) was thus integrated on axis 2 and following ones, and these axes were consequently discarded for the calculation of the phylogram depicting the amount of phylogenetic information strictly congruent between the morphological and the genetic data sets. This strict congruence phylogram was thus reconstructed by taking, for each genus, the scores of the morphological and genetic data only on the first CIA axis to compute dissimilarities between individuals. *Harttia* was used as the rooting group according to previous results. The tree that best fit the distance matrix (Fig. 3) showed a topology comparable to the one of the ML tree. The first difference was that Sturisomina was partly retrieved by grouping *Sturisoma*, *Sturisomatichthys*, and *Farlowella* but not *Lamontichthys*. The relationships within Sturisomina stayed unresolved because of contradictions between genetics and morphology. The second difference lied within the Loricariina where *Rineloricaria*,

Table 3
Main characteristics of variables tested for phylogenetic dependence

	I	V	VI	III	IV	II	G	H	C	E	D	F	I	A	J	K	B
Absolute contribution to co-inertia axis 1 (in %)	10.72	7.84	7.49	1.90	1.90	0.93	12.38	12.38	12.04	8.90	7.16	5.63	4.95	3.83	1.54	0.21	0.01
Absolute contribution to co-inertia axis 2 (in %)	1.60	0.74	0.40	0.68	0.68	0.14	20.94	21.16	13.47	12.13	0.09	0.59	2.92	13.39	3.37	4.50	3.02
TFSI test (C-mean)	0.7103	0.5606	0.6086	-0.0411	-0.043	0.0554	—	—	—	—	—	—	—	—	—	—	—
P value ($X \leq X$ obs.)	0.9998	0.9995	0.9991	0.6205	0.5708	0.8028	—	—	—	—	—	—	—	—	—	—	—
P value ($X \geq X$ obs.)	0.0003	0.0006	0.0010	0.3796	0.4293	0.1973	—	—	—	—	—	—	—	—	—	—	—
RUNS test (Runs-mean)	—	—	—	—	—	—	5.3745	6.627	6.3082	3.9297	5.4107	4.3323	4.3323	3.9671	5.7954	5.2008	9.5661
P value ($X \leq X$ obs.)	—	—	—	—	—	—	0.0002	0.0002	0.0017	0.0001	0.0318	0.0243	0.0243	0.0063	0.4931	0.2447	0.9529
P value ($X \geq X$ obs.)	—	—	—	—	—	—	0.9999	0.9999	0.9984	1	0.9683	0.9758	0.9758	0.9938	0.5070	0.7554	0.0472
R2Max test	0.6851	0.5166	0.5082	—	—	—	—	—	—	—	—	—	—	—	—	—	—
P value ($X \leq X$ obs.)	0.9995	0.8209	0.8525	—	—	—	—	—	—	—	—	—	—	—	—	—	—
P value ($X \geq X$ obs.)	0.0016	0.2029	0.2016	—	—	—	—	—	—	—	—	—	—	—	—	—	—
SkR2k test	2.5857	2.2934	2.6229	—	—	—	—	—	—	—	—	—	—	—	—	—	—
P value ($X \leq X$ obs.)	0.0007	0.0001	0.0005	—	—	—	—	—	—	—	—	—	—	—	—	—	—
P value ($X \geq X$ obs.)	0.9996	1	0.9997	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Dmax test	0.7002	0.7582	0.7225	—	—	—	—	—	—	—	—	—	—	—	—	—	—
P value ($X \leq X$ obs.)	0.9995	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—
P value ($X \geq X$ obs.)	0.001	0.0001	0.0002	—	—	—	—	—	—	—	—	—	—	—	—	—	—
SCE test	2.2602	2.4662	2.1879	—	—	—	—	—	—	—	—	—	—	—	—	—	—
P value ($X \leq X$ obs.)	0.9996	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—
P value ($X \geq X$ obs.)	0.0007	0.0001	0.0001	—	—	—	—	—	—	—	—	—	—	—	—	—	—

Variables are titled as in Covain and Fisch-Muller (2007), and are ordered according to their absolute contributions to the first co-inertia axis. Quantitative variables I–VI. I, number of caudal-fin rays (including spines); II, number of pectoral-fin rays (including spine); III, number of pelvic-fin rays (including spine); IV, number of dorsal-fin rays (including spine); V, number of premaxillary teeth; VI, number of dentary teeth. Qualitative variables A–K. A, abdominal cover with three modalities: 1 = absent, 2 = present incomplete, 3 = present complete; B, secondary organization in the abdominal cover with two modalities: 1 = absent, 2 = present; C, postorbital notches with three modalities: 1 = absent, 2 = present weak, 3 = present deep; D, predorsal keels with two modalities: 1 = absent, 2 = present; E, lips structure with three modalities: 1 = papillose, 2 = filamentous, 3 = rather smooth; F, fringed barbels with two modalities: 1 = absent, 2 = present; G, mouth shape with four modalities: 1 = elliptical, 2 = horse shoe like, 3 = bilobate, 4 = bilobate with trapezoidal opening; H: tooth shape with five modalities: 1 = pedunculated, 2 = straight bicuspid, 3 = pedunculated size reduced, 4 = straight bicuspid size reduced, 5 = spoon shaped size reduced; I, maxillary barbels with two modalities: 1 = conspicuous, 2 = inconspicuous; J, rostrum with two modalities: 1 = absent, 2 = present; K, snout shape with two modalities: 1 = rounded, 2 = pointed. Absolute contribution to co-inertia axis: contribution of each variable to the total inertia explained by the axis. TFSI and RUNS tests: tests against phylogenetic autocorrelation, respectively, for quantitative and qualitative variables as defined by Abouheif (1999). R2Max, SkR2k, Dmax, and SCE tests: tests against phylogenetic dependence as defined by Ollier et al. (2006). Bold types indicate significant tests for $\alpha = 5\%$.

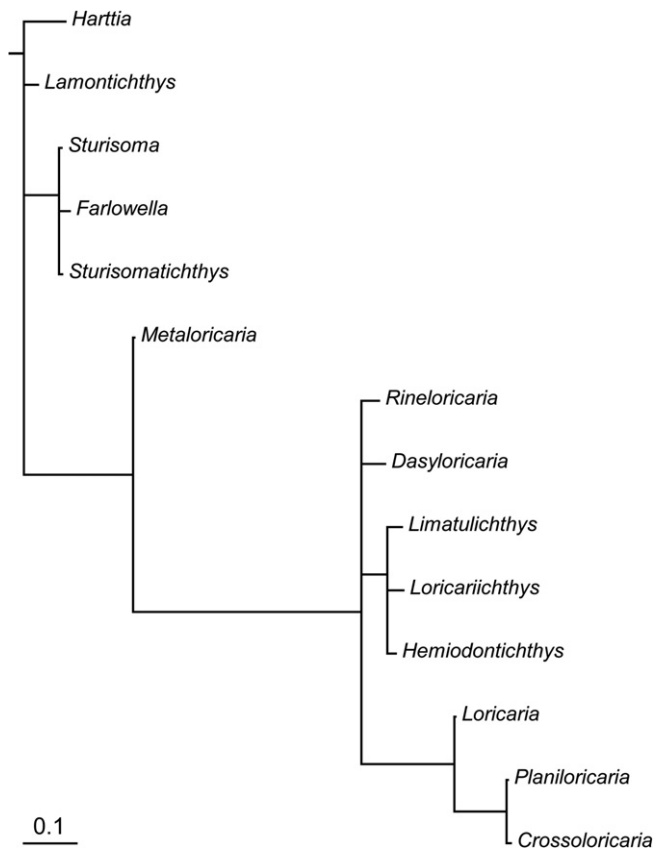


Fig. 3. Strict congruence phylogram computed from individuals' scores on the first co-inertia axis of the morphological and genetic data using Fitch and Margoliash algorithm. Sum of squares = 0.36173, average percent standard deviation (APSD) = 4.48288. Scale indicates the quantity of information computed from the morphological and genetic data sets.

Dasyloritaria, the *Loricariichthys* group and the *Loricaria* + *Pseudoheiodon* groups were all retrieved but with unresolved interrelationships. The last difference was the polytomy within the *Loricariichthys* group due to conflicting information between morphological and genetic data.

3.3. Identification of morphological phylogenetically dependent variables

The quality of the obtained strict congruence phylogram allowed the recognition of several morphological groups that were congruent with the molecular phylogeny, and highlighted the level of resolution reached by the morphological variables to describe these groups from a phylogenetic point of view. The variables involved in the characterization of these groups were then tested for phylogenetic dependence following our new approach. The CIA results are summarized in the first two lines of Table 3. The contributions of quantitative variables to the first axis ranged from 10.72% to 0.93%, while quantitative variables ranged from 12.38% to 0.01%. On axis 2, the absolute contributions of quantitative variables were small (1.6–0.14%), while qualitative variables showed generally high contributions (21.16–0.09%). These results were compared to the

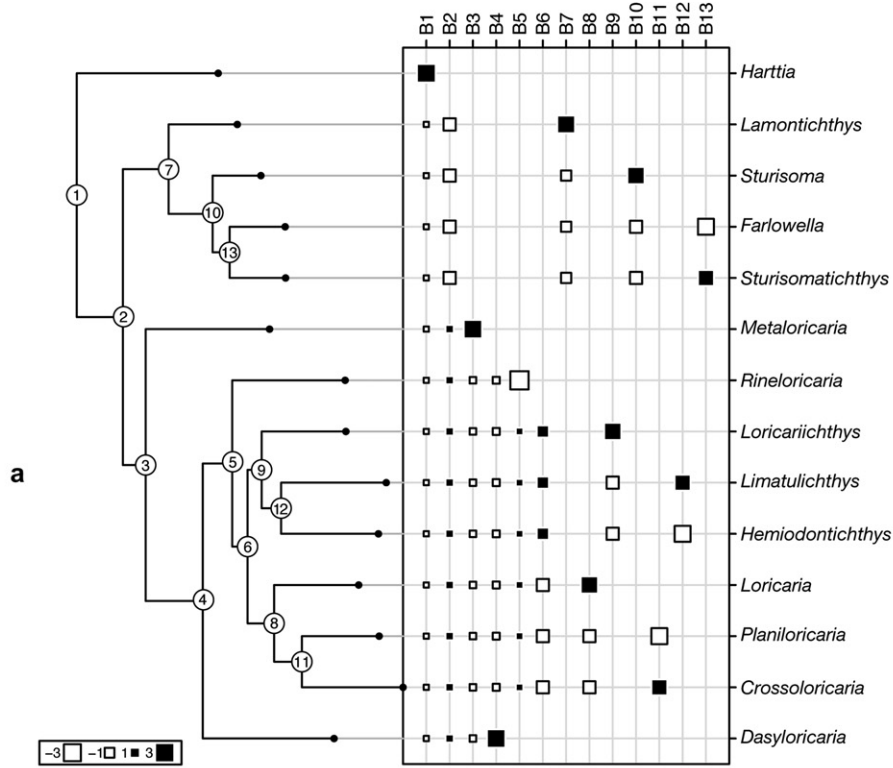
outputs of the TFSI tests (Table 3) which identified three quantitative variables to be strongly positively autocorrelated with the phylogeny: (1) the number of caudal-fin rays (I), (2) the numbers of premaxillary (V) and (3) dentary teeth (VI). These three variables also showed the strongest contributions to co-inertia axis 1, ranging from 10.72% (I) to 7.49% (VI). On axis 2, all quantitative variables were weakly informative. The CIA results were then compared to the outputs of the RUNS tests conducted on qualitative variables (Table 3) which showed a significant autocorrelation to the phylogeny for the following characters: abdominal cover present or absent (A), postorbital notches shape (C), predorsal keels present or absent (D), lip structure (E), fringed barbels present or absent (F), mouth shape (G), the tooth shape (H), and maxillary barbel length (I). The null hypothesis of absence of phylogenetic autocorrelation was consequently rejected for all these variables. To the contrary, absence of phylogenetic autocorrelation was not significantly rejected for the following variables: secondary organization in the abdominal plating (B), rostrum present or absent (J), and snout shape (K). All phylogenetically autocorrelated variables possessed the strongest contributions to axis 1, ranging from 12.38% (G, H) to 3.83% (A). On axis 2, phylogenetically autocorrelated variables such as predorsal keels present or absent (D), and fringed barbels present or absent (F) appeared weakly informative (0.09% and 0.59%, respectively), whereas uninformative variables such as rostrum present or absent (J) and the snout shape (K) played a more important role on the axis, contributing, respectively, to 3.37% and 4.5%. This means that one part of the background noise was integrated on axis 2, and provided an *a posteriori* justification for the rejection of axis 2 and next ones in the calculation of the strict congruence phylogram. In summary, the variables that contributed more than 3.83% to the co-inertia axis 1 were significantly correlated to the phylogeny according to TFSI and RUNS results.

3.4. Evolutionary analysis of phylogenetically dependent morphological variables

Quantitative variables I, V, and VI (Table 3) were analyzed using the orthogram approach (Fig. 4). The tree topology together with the vectorial basis (Fig. 4a) allowed the identification of the ranking of the nodes, and consequently to see which vector accounted for which node. The orthogram of the first quantitative variable analyzed, the number of caudal-fin rays (I) (Fig. 4b, top), indicated that vector 2 explained the greatest part of the variance. This vector showed a strong departure from the expected value under the hypothesis of absence of phylogenetic dependence (given by the solid line in Fig. 4b, top), and peaked outside of the confidence limit (given by the dashes). The cumulative orthogram (Fig. 4b, down) confirmed predominance of vector 2 in the variance distribution. A significant departure from H0 was registered for this vector, and this pattern was preserved for several suc-

cessive vectors. The maximum deviation from the expected value was given for the sum of the three first vectors (vertical arrow in Fig. 4b, down) meaning that maximum var-

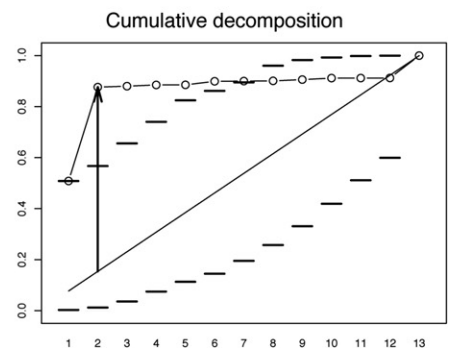
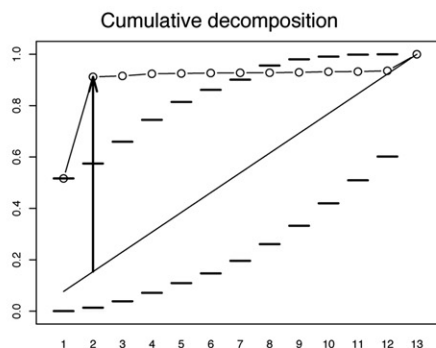
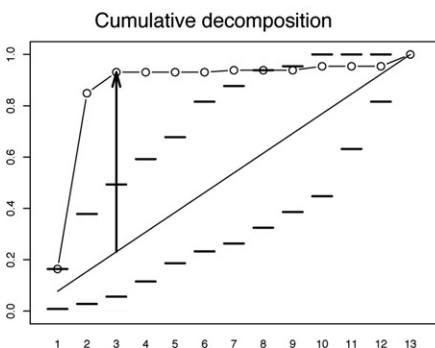
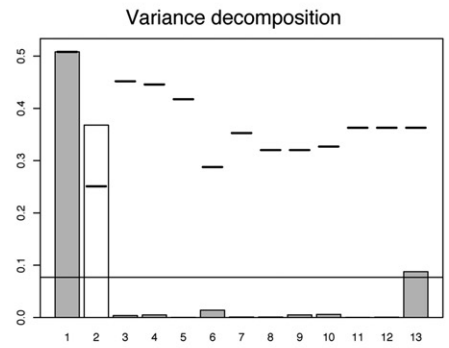
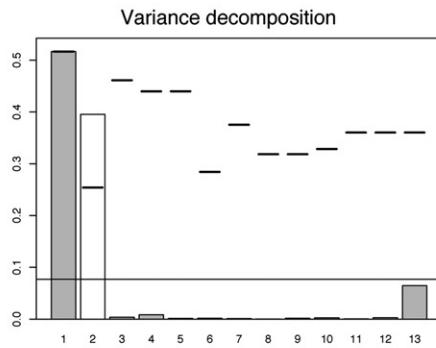
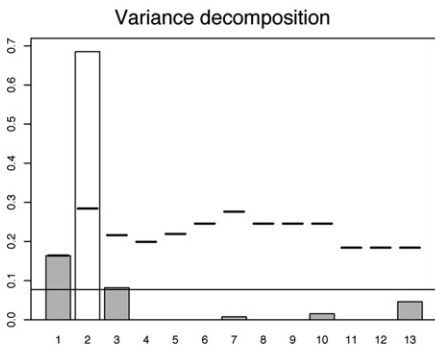
iation was registered on these three vectors. All four statistical tests were also significant, particularly R2Max (Table 3; $p(X \geq X_{obs}) = 0.0016$), indicating that a single



b NUMBER OF CAUDAL-FIN RAYS (I)

c NUMBER OF PREMAXILLARY TEETH (V)

d NUMBER OF DENTARY TEETH (VI)



punctual modification of the trait (number of caudal-fin rays) occurred at a particular node and that it stayed unchanged afterward. Moreover, the variance distribution was rather skewed towards the root (Table 3; SkR2k: $p(X \leq X_{\text{obs}}) = 0.0007$), indicating that the deepest nodes of the phylogeny explained the variance distribution. These results suggested that this trait has been shaped deep in the phylogeny. In summary, a single major punctual event occurred at node 2, between Sturisomina and Loricariina lineages, with a reduction of the number of caudal-fin rays in Loricariina.

In the second and third quantitative traits analyzed, the numbers of premaxillary (V) and dentary (VI) teeth, variance decomposition showed similar patterns. The orthogram plot (Fig. 4c and d, up) pointed vectors 1 and 2 as explaining the major part of the variance distribution. Cumulative orthograms (Fig. 4c and d, down) confirmed this fact with a maximum departure from the expected value under absence of phylogenetic dependence registered for the sum of two first vectors (arrow on vector 2). Out of the four statistics tested (Table 3), only R2Max was not significant meaning that a rather gradual effect was responsible of the variance distribution. Moreover, this distribution was skewed towards the root (Table 3, SkR2k: $p(X \leq X_{\text{obs}}) = 0.0001$ and $p(X \leq X_{\text{obs}}) = 0.0002$ for numbers of premaxillary and dentary teeth, respectively). Consequently, these two traits have been also shaped rather deep in the phylogeny. Two major successive events can be reconstructed in the overall gradual trend: a first decrease in the number of premaxillary and dentary teeth between Harttiini and Loricariini lineages (Fig. 4a, node 1), and a second decrease between Sturisomina and Loricariina lineages (Fig. 4a, node 2).

Qualitative variables A, C, D, E, F, G, H, and I were analyzed using Maximum likelihood ancestral state reconstruction (Fig. 5). The mouth shape (Fig. 5a) evolved from circular in Harttiini and Sturisomina, to bilobate in all Loricariina except *Metaloricaria* which displays a horse shoe like mouth. Therefore, the ancestral state reconstruction showed an unclear state at the root of Loricariina, with a slight preference for the elliptical state ($p_{G1} = 0.6186$). A second step in the specialization of the mouth in Loricariina occurred in the *Pseudohemiodon* group which displays a bilobate mouth but with a trapezoi-

dal opening. The tooth shape (Fig. 5b) showed a similar pattern of evolution than the mouth shape. Tooth evolved from pedunculated in Harttiini and Sturisomina to more specialized in Loricariina. A first step occurred at the basal diversification of the Loricariina where the teeth evolved from an ancestor possessing more probably pedunculated teeth ($p_{H1} = 0.6120$), to teeth pedunculated yet reduced in size in *Metaloricaria*, and straight and bicuspid in the sister lineage. In this last lineage, two other modifications occurred later on: a reduction in size in the *Loricariichthys* group and a change towards spoon shaped teeth reduced in size in the *Pseudohemiodon* group. The postorbital notches (Fig. 5c) appeared in the ancestor of the Loricariina. This feature regressed two times toward weak postorbital notches: a first time in *Limatulichthys*, and a second time in the *Pseudohemiodon* group. The lip structure (Fig. 5d) evolved from papillose in Harttiini, Sturisomina, and basal Loricariina, to rather smooth in the *Loricariichthys* group, and filamentous in the *Loricaria* and *Pseudohemiodon* groups. The sudden diversification of the lip structure made it difficult to reconstruct the ancestral state at the origin of this diversification (Fig. 5d, $p_{E1} = 0.4118$, $p_{E2} = 0.2265$, $p_{E3} = 0.3617$). Predorsal keels (Fig. 5e) appeared most probably in the ancestor of the Loricariina lineage not comprising *Metaloricaria* ($p_{D1} = 0.8401$). Thereafter, this feature regressed in several representatives of the *Loricariichthys* group such as *Loricariichthys* and *Limatulichthys*. Fringed barbels (Fig. 5f) are present only in some members of the Loricariina, yet the first appearance of this feature was difficult to assess and consequently none of the deeper ancestral nodes within this tribe displayed a clear state assignment. It seemed however clear that this feature regressed in representatives of the *Loricariichthys* group while it has never been present in *Metaloricaria*. The maxillary barbels (Fig. 5g) evolved from inconspicuous to conspicuous in two Loricariina lineages: the *Loricaria* and *Pseudohemiodon* groups. The abdominal cover (Fig. 5h) is absent in the species representing Harttiini and present in extant Loricariini, making it difficult to assess the state of the ancestor, yet the Maximum likelihood ancestral state reconstruction method slightly favors the presence of an abdominal cover ($p_{A3} = 0.7171$). Later on, this character evolved from a complete abdominal cover to an incomplete cover in the *Pseudohemiodon* group.

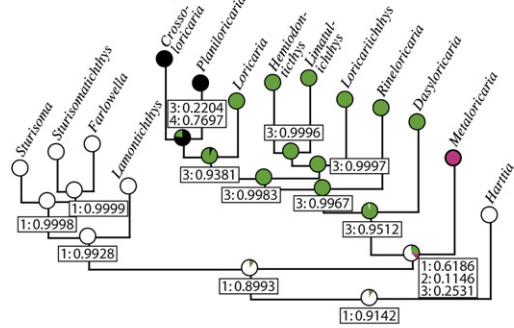
Fig. 4. Variance decomposition of three quantitative morphological traits across the orthonormal basis defined by the phylogenetic tree topology. (a) Phylogenetic tree (left) and description of the topology of the tree by the orthonormal vectors B1 to B13 which represent nodes and descendent tips (right). Node numbering in the phylogenetic tree (1–13) indicates the number of the vector (B1–B13) that accounts for the variance associated to the node. The indicative scale show squares with sizes proportional to the values of the orthonormal vectors (white and black for negative and positive values, respectively). (b) Variance decomposition of the number of caudal-fin rays (I) using the orthogram plot (upper panel) and the cumulative orthogram plot (lower panel). (c) Variance decomposition of the number of premaxillary teeth (V) using the orthogram plot (upper panel) and the cumulative orthogram plot (lower panel). (d) Variance decomposition of the number of dentary teeth (VI) using the orthogram plot (upper panel) and the cumulative orthogram plot (lower panel). In the orthogram plots, the abscise gives the number of the vectors associated to nodes while the ordinate shows the contribution of the vector to the variance of the trait given by the squared regression coefficient (white and gray for positive and negative coefficients, respectively); dashes correspond to the upper confidence limit at 5% deduced from 9999 Monte Carlo permutations; solid line represents the mean value. In the cumulative orthogram plots the ordinate shows the cumulated contribution of successive vectors to the variance; circles represent the observed value of cumulated squared regression coefficients; solid diagonal line represents expected value under absence of phylogenetic dependence; dashes correspond to the bilateral 95% confidence interval. Vertical arrow indicates the position of maximum deviation from the expected value (diagonal line).

4. Discussion

In this work, we were interested in reconstructing the evolutionary history of the Loricariinae, a highly special-

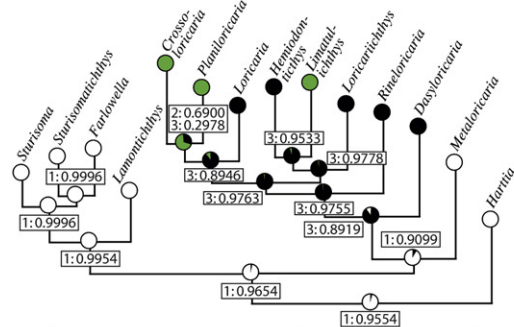
ized group of neotropical catfishes, and in deciphering the evolution of their morphological traits. For this purpose, we used a new approach to detect phylogenetic dependence of character variations to the phylogeny, which

a MOUTH SHAPE (G)



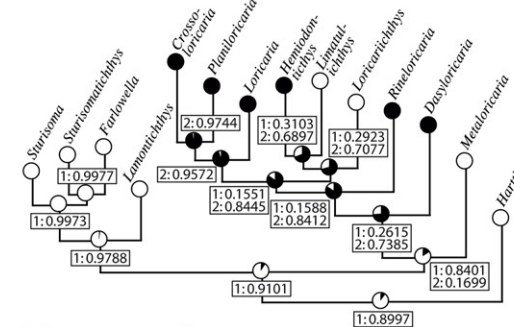
- ① elliptical
- ② horse shoe like
- ③ bilobate
- ④ bilobate with trapezoidal opening

c POSTORBITAL NOTCHES (C)



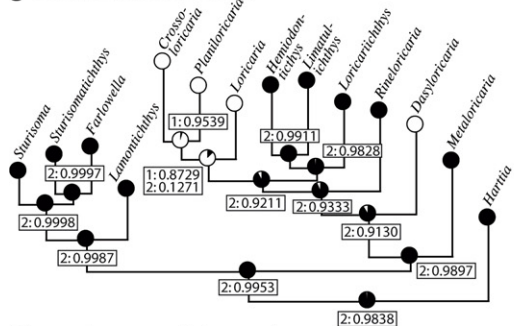
- ① absent
- ② present weak
- ③ present deep

e PREDORSAL KEELS (D)



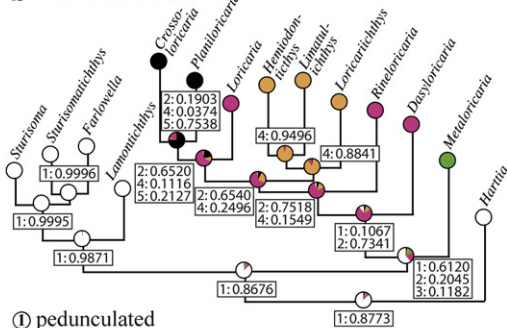
- ① absent
- ② present

g MAXILLARY BARBELS (I)



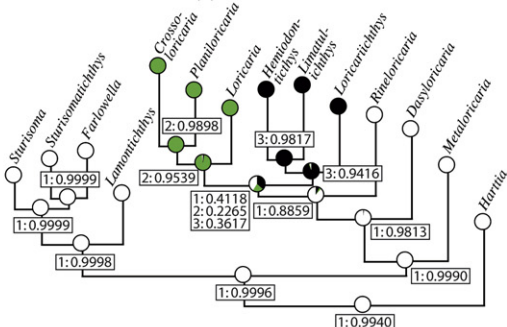
- ① conspicuous
- ② inconspicuous

b TOOTH SHAPE (H)



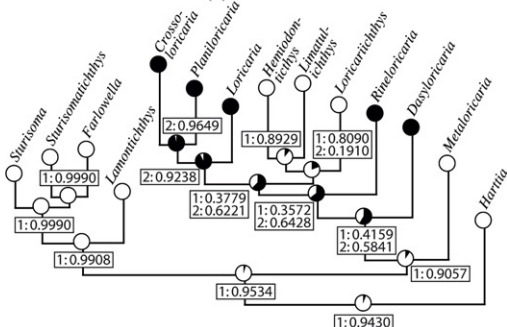
- ① pedunculated
- ② straight bicuspid
- ③ pedunculated size reduced
- ④ straight bicuspid size reduced
- ⑤ spoon shape size reduced

d LIPS STRUCTURE (E)



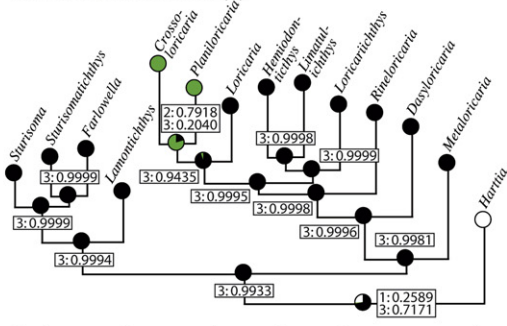
- ① papillose
- ② filamentous
- ③ rather smooth

f FRINGED BARBELS (F)



- ① absent
- ② present

h ABDOMINAL COVER (A)



- ① absent
- ② present incomplete
- ③ present complete

is a prerequisite for a sensible evolutionary analysis of characters. Our approach using the CIA has the advantage over existing methods to treat the full morphological data set at once, including qualitative and quantitative variables. The CIA offers a graphical output which allows a detailed analysis of the contribution of individual variables to the overall trend, within the frame of the phylogenetic tree. Contradictions and congruencies between both data sets are highlighted on the factorial map of individuals (Fig. 2a) by the relative position of both systems of coordinates (genetic and morphological) onto co-inertia axes. Incongruence between both data sets is given by the size of arrows representing the differences between genetics and morphology. Longer arrows, or origin of arrows in positive values and extremity in negative values imply strong contradictions between both data sets. In our case, no strong contradiction was highlighted by the CIA. The factorial map of variables (Fig. 2b), reveals the contribution of each variable to the co-structure, and identifies the groups defined by these variables. The graph of eigenvalues (Fig. 2d) identifies the axis explaining the major part of the congruent information between both data sets. Thus, the CIA provides an ordination of the variables according to their contribution to the co-inertia axes and by this mean allows the identification of phylogenetically dependent variables as well as the identification of the axes containing phylogenetic “noise” which are then discarded from the calculation of the strict congruence phylogram. The CIA approach has also the advantage of having no theoretical limitations and can be generalized to K tables displaying the same taxonomical sampling. These data can be of many different types (genetic, morphological, ethological, geographical, ecological...). The robustness of the CIA approach was assessed by comparing the level of correlation to the phylogeny as obtained by this method and the *p*-values obtained by classical tests, namely the TFSI test for quantitative variables, and the RUNS test for qualitative variables (Abouheif, 1999). The results of the comparisons (Table 3) show a strict correspondence between our approach and Abouheif’s tests for assessing phylogenetic dependence and in this way we have shown that variables contributing more than 3.83% to the co-inertia axis 1 were significantly correlated to the phylogeny.

In order to study the morphological evolution of the Loricariinae catfishes, we first inferred the phylogeny of

the subfamily using 12S and 16S mitochondrial genes. The results show that Harttiini *sensu* Rapp Py-Daniel (1997) is not a monophyletic assemblage due to the scattered positions of its representatives in the phylogenetic tree, with a basal position of *Harttia* (type genus) as the sister group to all other Loricariinae analyzed. This corroborates the findings of Montoya-Burgos et al. (1998) who recovered this topology with a more restricted Loricariinae sampling. According to our results, we propose that the Harttiini should be restricted to the single genus *Harttia*. In the phylogenetic tree, the Loricariini *sensu* Isbrücker (1979) was not retrieved. We thus redefine the Loricariini as the clade comprising two sister subtribes, (1) the Loricariina including all former Loricariini *sensu* Isbrücker (1979), and (2) a new monophyletic subtribe named Sturisomina, from the name of the first described genus of this group. Sturisomina includes at its base *Lamontichthys* as sister genus of *Farlowella*, *Sturisoma*, and *Sturisomatichthys* whose relationships still deserve further investigations. The paraphyly of *Farlowella*, even though surprising, is supported by the significant rejection of the constrained monophyly of the genus as assessed using the SH-test. A larger taxonomic sampling remains however necessary to definitely answer this question.

The relative position of *Metaloricaria* at the base of the Loricariina clade, is not consistent with the classification of Isbrücker (1979) who assigned it to the Harttiini tribe, and Metaloricariina subtribe. The position of *Metaloricaria* in our trees is poorly supported by bootstrap values for MP and NJ analyses, and should therefore be considered cautiously. However, the topology agrees with the hypothesis of Rapp Py-Daniel (1997) who suggested a placement within Loricariini (*sensu* Isbrücker, 1979). Herein, the Loricariina constitutes the sister group of Sturisomina. Within the Loricariina, *Dasylicaria* occupies a basal position, just after *Metaloricaria*, while *Rineloricaria* has a derived position relative to *Dasylicaria* and constitutes the sister group to all other Loricariina. This topological situation renders the Rineloricariina subtribe proposed by Isbrücker, 1979 paraphyletic. Indeed, this subtribe comprised *Dasylicaria*, *Rineloricaria*, *Ixinandria*, and *Spatuloricaria*, a grouping which is incompatible with our results. In addition, this subtribe was already questioned by Rapp Py-Daniel (1997) who found a paraphyly between *Spatuloricaria* and *Rineloricaria*. Here, the *Loricariichthys* group

Fig. 5. Maximum likelihood ancestral state reconstructions of eight qualitative life-history traits along the phylogenetic tree using a single forward-backward rate (mK) model. Traits are ordered according to their absolute contribution to co-inertia axis 1. (a) Ancestral state reconstruction of the mouth shape with four modalities (G): estimated rate of change 1.981274554, $-\log L = 13.32290011$; (b) ancestral state reconstruction of the tooth shape with five modalities (H): estimated rate of change = 2.431420956, $-\log L = 18.364202556$; (c) ancestral state reconstruction of the postorbital notches with three modalities (C): estimated rate of change = 3.33717473, $-\log L = 12.268601550$; (d) ancestral state reconstruction of the lips structure with three modalities (E): estimated rate of change = 1.84519947, $-\log L = 10.47907210$; (e) ancestral state reconstruction of the predorsal keels with two modalities (D): estimated rate of change = 7.174725755, $-\log L = 8.39048636$; (f) ancestral state reconstruction of the fringed barbels with two modalities (F): estimated rate of change = 5.604381355, $-\log L = 8.07064024$; (g) ancestral state reconstruction of the maxillary barbels with two modalities (I): estimated rate of change = 4.041408096, $-\log L = 7.524279656$; (h) ancestral state reconstruction of the abdominal cover with three modalities (A): estimated rate of change = 1.88539421, $-\log L = 9.22821077$. Boxes indicate the marginal probabilities of the most probable states. Likelihoods are reported as proportional likelihoods.

constitutes the sister clade of *Loricaria* plus the *Pseudohemiodon* groups. On the basis of the present taxonomic sampling, *Loricaria* is the sister clade of the *Pseudohemiodon* group represented here by *Crossoloricaria* and *Planiloricaria*. This agrees with Rapp Py-Daniel's (1997) results who found *Loricaria* branching at the base of the Planiloricariina (comprising *Planiloricaria* and *Crossoloricaria* among others). Nevertheless, these relationships deserve a wider taxonomic sampling for being confidently supported.

An overview of the morphological groups recently proposed by Covain and Fisch-Muller (2007) and the molecular phylogenetic results obtained herein, suggested that common information was shared between both data types. A strong correlation was indeed observed ($RV = 0.832$). This analysis suggested that several morphological groups were not obtained by chance or by character convergence, but followed a phylogenetic classification. The amount of congruent information between both data sets is in fact significant as summarized in the strict congruence phylogram (Fig. 3). This phylogram based on the co-structure analysis confirmed the natural status of several morphological groups like the Harttiini and, among Loricariini: Loricariina (including *Loricariichthys*, and *Loricaria-Pseudohemiodon* groups), and one part of Sturisomina. The *Rineloricaria* group did not constitute a natural group as defined by incompatible molecular and morphological hypotheses. The co-structure showed that the variables used by Covain and Fisch-Muller (2007) were relevant to characterize tribal and subtribal ranks, as well as several morphological groups, but were insufficient to define the generic rank. Therefore, the lack of resolution at the generic level in the phylogram came mainly from the restricted morphological data set rather than from incompatibilities (6 discrete quantitative and 11 qualitative variables). However, the quality of the strict congruence phylogram obtained validates the co-inertia approach in a phylogenetic context by identifying morphological variables correlated to the phylogeny in a pool of different types of variables.

Maximum likelihood ancestral state reconstructions of qualitative variables underlined similar patterns of evolution of traits linked to the mouth. Moreover, the mouth characteristics appeared as the most important features for discriminating the different groups of this subfamily, as traits linked to this organ show the strongest variations correlated to the phylogeny. Therefore, we believe that the mouth shape, the tooth shape, the lips structure and the barbels shape may have co-evolved due to identical selective pressure acting on this organ. The co-variation of these traits may reflect adaptations to the large number of ecological niches conquered by the Loricariinae. For instance, species occurring over sandy substrates, such as the representatives of the *Pseudohemiodon* and *Loricaria* groups, possess a bilobate mouth with filamentous lips, whereas more rheophilic species like representatives of *Harttia* or *Lamontichthys* (which live on stones) possess an elliptical mouth with papillose lips. Our conclusions also highlight

the difficulties in defining evolutionary independent morphological characters for phylogenetic purposes.

Some qualitative variables retained as phylogenetically dependent were homoplastic as referring to the molecular phylogenetic tree such as the predorsal keels, the fringed and maxillary barbels. The two first characters show local losses while the third displays two independent gains, which is a case of evolutionary convergence. This indicates that the CIA approach is not too restrictive and allows the retention of characters with some degree of homoplasy which can be of different nature (losses or independent gains). However, although retaining them as interesting characters, the CIA ordered them as the less informative among the retained ones (see absolute contributions on axis 1 in Table 3).

The analysis of the quantitative variables with the orthogram method (Ollier et al., 2006) not only showed that these variables were shaped by the evolutionary history of this group but also described how these variables evolved along the phylogeny. The analysis of the number of caudal-fin rays indicated a significant drop at the base of the Loricariina lineage, with a reduction of rays from 14 (13 in *Farlowella*) in Sturisomina, to 12 in Loricariina (13 in *Metaloricaria*). We have noticed that in Loricariina, the loss of caudal-fin rays was accompanied by the appearance of a thicker caudal-spine bearing a whip used as a defensive weapon. These concomitant morphological changes may therefore be linked and the formation of the thicker caudal-spine with its whip may be the outcome of fin ray fusions. Contrasting with the instance of caudal-fin rays number variation linked to the phylogeny presented above, the punctual reduction of caudal-fin rays in *Farlowella* and *Metaloricaria* were not dependent to the phylogeny but rather randomly distributed events and were thus discarded from an evolutionary interpretation. The analysis of the caudal-fin rays exemplifies the possibility that a given morphological trait may display changes that are linked to the phylogeny and others that arise in a stochastic manner. Yet, we have the tools to discern between these two situations. The study of the number of premaxillary and dentary teeth revealed a more gradual evolution of these features, as indicated by the non significance of the R2Max test. The decrease in the number of teeth extended gradually along the phylogeny, from Harttiini (bearing 80 premaxillary and 70 dentary teeth) to Loricariini (bearing less than 60 premaxillary and 50 dentary teeth), and then between Sturisomina (bearing 20 to 60 premaxillary and 15 to 50 dentary teeth) and Loricariina (bearing 0 to 15 premaxillary and 3 to 15 dentary teeth).

As shown in our study, the orthogram method of Ollier et al. (2006) proved to be a powerful tool to detect phylogenetic dependence and to analyze the patterns of evolution of quantitative life-history traits. However, this method suffers from the fact that it can not treat qualitative variables; a weakness that can be partly overcome by using the CIA approach. The convincing results given by the orthograms encourage nevertheless the development of

the method for analyzing qualitative data or even a complete table mixing different types of data. The theoretical background for generalizing the orthogram method is in progress and its implementation will be performed soon.

Acknowledgments

We are grateful to Volker Mahnert, MHNG, Geneva, for being the supervisor of first author during the early stage of this work, to Claude Weber, and Andreas Schmitz, MHNG, Geneva, and Lawrence M. Page, University of Florida, Gainesville for their helpful advices; Jan Pawlowski, University of Geneva, for the laboratory facilities, and José Fahrni, University of Geneva, for the sequencing process. We would also like to thank Philippe Debey, Poisson Ange Company, Lure, for providing specimens and samples from the trade; the G. and A. Claraz Foundation for their financial support for the missions in Surinam in 2001 and 2005; the Guyana Environmental Protection Agency for collecting permit; and the NRDDDB and Iwokrama organization for their field logistic. This project was partly supported by the Fond National Suisse de la Recherche Scientifique (JIMB 3100A0–104005), and by a field grant of the Académie Suisse des Sciences Naturelles (ASSN). The figures were finalized by Florence Marteau, MHNG.

References

- Abouheif, E., 1999. A method to test the assumption of phylogenetic independence in comparative data. *Evol. Ecol. Res.* 1, 895–909.
- Blomberg, S.P., Garland Jr., T., Ives, A.R., 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57, 717–745.
- Chessel, D., Dufour, A.B., Thioulouse, J., 2004. The ade4 package—One-table methods. *R News* 4, 5–10.
- Cheverud, J.M., Dow, M.M., Leutenegger, W., 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. *Evolution* 39, 1335–1351.
- Covain, R., Fisch-Muller, S., 2007. The genera of the Neotropical armored catfish subfamily Loricariinae (Siluriformes: Loricariidae): a practical key and synopsis. *Zootaxa* 1462, 1–40.
- Diniz-Filho, J.A.F., de Sant'Ana, C.E.R., Bini, L.M., 1998. An eigenvector method for estimating phylogenetic inertia. *Evolution* 52, 1247–1262.
- Dolédéc, S., Chessel, D., 1994. Co-inertia analysis: an alternative method for studying species–environment relationships. *Freshwater Biol.* 31, 277–294.
- Dray, S., Chessel, D., Thioulouse, J., 2003. Co-inertia analysis and the linking of ecological data tables. *Ecology* 84, 3078–3089.
- Efron, B., 1979. Bootstrap methods: another look at the jackknife. *Ann. Stat.* 7, 1–26.
- Felsenstein, J., 1981. Evolutionary trees from DNA-sequences—a maximum-likelihood approach. *J. Mol. Evol.* 17, 368–376.
- Felsenstein, J., 1985a. Phylogenies and the comparative method. *Am. Nat.* 125, 1–15.
- Felsenstein, J., 1985b. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Felsenstein, J., 2004. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
- Fitch, W.M., Margoliash, E., 1967. Construction of phylogenetic trees. *Science* 155, 279–284.
- Gittleman, J.L., Kot, M., 1990. Adaptation: statistics and null model for estimating phylogenetic effects. *Syst. Zool.* 39, 227–241.
- Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* 49, 652–670.
- Gower, J.C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53, 325–338.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Harvey, P.H., Pagel, M.D., 1991. *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford, 248 pp.
- Hill, M.O., Smith, A.J.E., 1976. Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon* 25, 249–255.
- Huelsenbeck, J.P., Nielsen, R., Bollback, J., 2003. Stochastic mapping of morphological characters. *Syst. Biol.* 52, 131–158.
- Ihaka, R., Gentleman, R., 1996. R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* 5, 299–314.
- Isbrücker, I.J.H., 1979. Description préliminaire de nouveaux taxa de la famille des Loricariidae, poissons-chats cuirassés néotropicaux, avec un catalogue critique de la sous-famille nominale (Pisces, Siluriformes). *Rev. Fr. Aquariol.* 5, 86–116.
- Leviton, A.E., Gibbs Jr., R.H., Heal, E., Dawson, C.E., 1985. Standards in herpetology and ichthyology: part I. Standard symbolic codes for institutional resource collections in herpetology and ichthyology. *Copeia* 1985, 802–832.
- Lewis, P.O., 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50, 913–925.
- Lingoes, J.C., 1971. Some boundary conditions for a monotone analysis of symmetric matrices. *Psychometrika* 36, 195–203.
- Lynch, M., 1991. Methods for the analysis of comparative data in evolutionary biology. *Evolution* 45, 1065–1080.
- Maddison, W.P., Maddison, D.R., 2006a. StochChar: a package of mesquite modules for stochastic models of character evolution. Version 1.1 <<http://mesquiteproject.org>>.
- Maddison, W.P., Maddison, D.R., 2006b. Mesquite: a modular system for evolutionary analysis. Version 1.12 <<http://mesquiteproject.org>>.
- Montoya-Burgos, J.I., Muller, S., Weber, C., Pawlowski, J., 1998. Phylogenetic relationships of the Loricariidae (Siluriformes) based on mitochondrial rRNA gene sequences. In: Malabarba, L.R., Reis, R.E., Vari, R.P., Lucena, Z.M.S., Lucena, C.A.S. (Eds.), *Phylogeny and Classification of Neotropical Fishes*. Edipucrs Porto Alegre, Porto Alegre, Brazil, pp. 363–374.
- Ollier, S., Couteron, P., Chessel, D., 2006. Orthonormal transform to decompose the variance of a life-history trait across a phylogenetic tree. *Biometrics* 62, 471–477.
- Pagel, M., 1999a. Inferring the historical patterns of biological evolution. *Nature* 401, 877–884.
- Pagel, M., 1999b. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.* 48, 612–622.
- Pagel, M., Meade, A., Barker, D., 2004. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* 53, 673–684.
- Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Rapp Py-Daniel, L.H., 1997. Phylogeny of the Neotropical armored catfishes of the subfamily Loricariinae (Siluriformes: Loricariidae). Unpublished Ph.D. Thesis, University of Arizona, Tucson, 280 pp.
- Reeve, J., Abouheif, E., 2003. Phylogenetic Independence Version 2.0. Distributed freely by the authors on request. Department of Biology, McGill University.
- Robert, P., Escoufier, Y., 1976. A unifying tool for linear multivariate statistical methods: the RV coefficient. *Appl. Stat.-J. Roy. St. C* 25, 257–265.
- Rholf, F.J., 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution* 55, 2143–2160.

- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Schaefer, S.A., 1987. Osteology of *Hypostomus plecostomus* (Linnaeus) with a phylogenetic analysis of the loricariid subfamilies (Pisces: Siluroidei). *Contrib. Sci., Los Angeles Co. Mus. Nat. Hist.* 394, 1–31.
- Schaefer, S.A., 1991. Phylogenetic analysis of the loricariid subfamily Hypoptopomatinae (Pisces: Siluroidei: Loricariidae), with comments on generic diagnoses and geographic distribution. *Zool. J. Linn. Soc.-Lond.* 102, 1–41.
- Schaefer, S.A., Stewart, D.J., 1993. Systematics of the *Panaque dentex* species group (Siluriformes: Loricariidae), wood-eating armored catfishes from tropical South America. *Ichthyol. Explor. Freshwater* 4, 309–342.
- Schluter, D., Price, T., Mooers, A.O., Ludwig, D., 1997. Likelihood of ancestor states in adaptive radiation. *Evolution* 51, 1699–1711.
- Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114–1116.
- Sokal, R.R., Rohlf, F.J., 1995. *Biometry*, third ed. W.H. Freeman & Co., New York, 887 pp.
- Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.
- Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. In: Miura, R.M. (Ed.), *Some Mathematical Questions in Biology-DNA Sequence Analysis*. American Mathematics Society, Providence, RI, pp. 57–86.
- Thioulouse, J., Chessel, D., Dolédec, S., Olivier, J.M., 1997. ADE-4: a multivariate analysis and graphical display software. *Stat. Comput.* 7, 75–83.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- von Neumann, J., Kent, R.H., Bellinson, H.R., Hart, B.I., 1941. The mean square successive difference. *Ann. Math. Stat.* 12, 153–162.