Formalizing Reliability in the Taxonomic Congruence Approach

Blaise Li and Guillaume Lecointre*

*Correspondence: Guillaume Lecointre, Muséum National d'Histoire Naturelle (MNHN), case postale 26, 57 rue Cuvier, 75231 Paris cedex 05. Telephone: 33 (0)1 40 79 37 51. E-mail: lecointr@mnhn.fr

Supplementary material

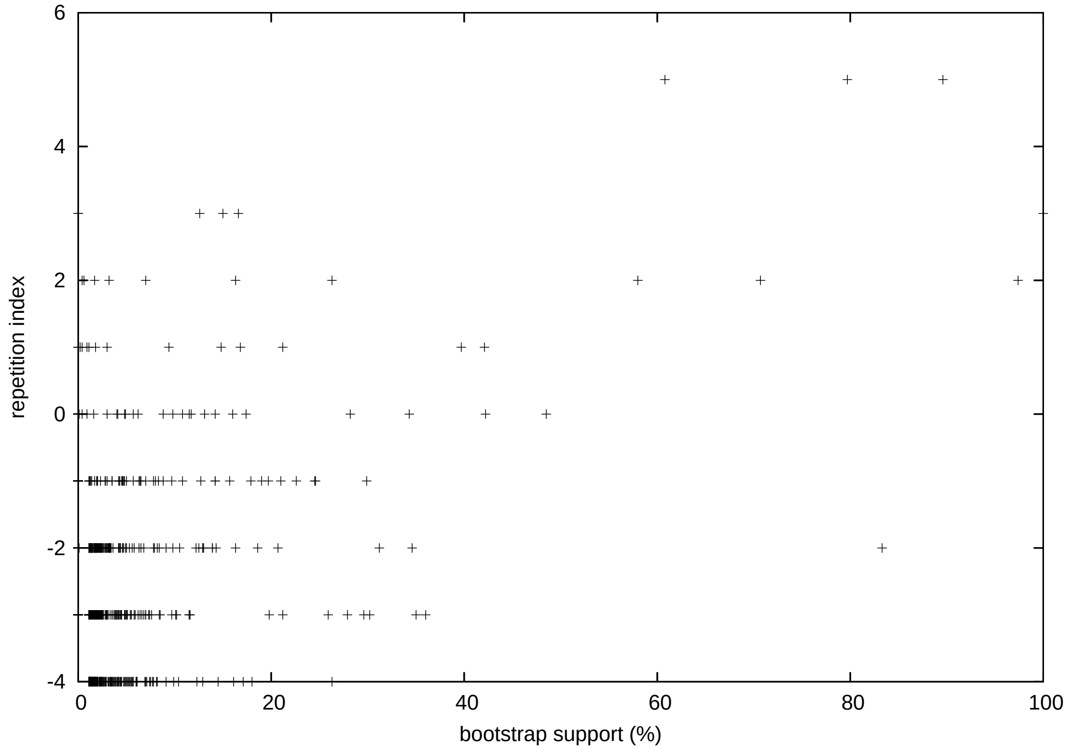

Figure 1: Repetition indices of the clades plotted against their bootstrap supports in the 'total evidence' analysis.

Table 1: Taxonomic sampling (in bold; sequences not present in Dettaï, 2004 or in Dettaï & Lecointre, 2005)

genus	28S	12S	16S	Rhodopsin	MLL	IRBP
Ammodytes	AY141689-92	AY141380	AY141450	AY141306	AY362234	**DQ309871**
Antennarius	AY372752-53	AY368287	AY368304	AY368324	AY362215	DQ168037
Apletodon	AY141557-60	AY141348	AY141418	AY141274	AY362213	DQ168039
Arnoglossus	AY141593-96	AY141358	AY141428	AY141283	AY362228	**DQ309872**
Aulostomus	AY141577-80	AY141353	AY141423	AY141279	AY362226	DQ168040

Table 1: Continued

genus	28S	12S	16S	Rhodopsin	MLL	IRBP
Barbourisia	DQ021385, 93	AY368290	AF221881	AY368333	AY362264	DQ168041
Bathypterois	AY141473-76	AY141326	AY141396	AY141257	AY362219	DQ168042
Bedotia	AY141525-28	AY141339	AY141409	AY141267	AY362271	DQ168043
Belone	AY141529-32	AY141340	AY141410	AY141268	AY362273	DQ168044
Beryx	AY141513-16	AY141336	AY141406	AY141265	AY362238	DQ168045
Bovichtus	AY141661-64	Z32702	**AY520101**	AY141299	AY362283	DQ168046
Callionymus	AY141541-44	AY141344	AY141414	AY141270	AY362225	DQ168047
Capros	AY141501-04	AY141333	AY141403	AY141262	AY362233	DQ168048
Ceratias	AY141505-08	AY141334	AY141404	AY141263	AY362270	DQ168049
Chaetodon	AY372701-04	**AJ748314**	AF055613	AY368312	AY362240	DQ168050
Channa	AY141569-72	AY141351	AY141421	AY141277	AY362241	DQ168051
Cheimarrichthys	AY141693-96	AY141381	AY141451	AY141307	AY362229	DQ168052
Chelidonichthys	AY141609-12	AY141362	AY141432	AY141287	AY362284	DQ168053
Chloroscombrus	AY141717-20	AY141387	AY141457	AY141313	AY362223	DQ168054
Citharus	AY372697-00	AF542220	AY157325	AY141323	AY362232	DQ168055
Ctenochaetus	AY141745-48	AY141394	AY141464	AY141320	AY362242	DQ168057
Ctenopoma	AY141573-76	AY141352	**AY662702**	AY141278	AY362210	DQ168058
Cyclopterus	AY372737-38	AY368284	AY368299	AY368316	AY362218	**DQ309873**
Dactylopterus	AY141589-92	AY141357	AY141427	AY141282	AY362243	DQ168059
Echeneis	AY141725-28	AY141389	AY141459	AY141315	AY362245	DQ168062
Epinephelus	AY141629-32	AY141367	AY141437	AY141291	AY362227	DQ168064
Forsterygion	AY141549-52	AY141346	AY141416	AY141272	AY362276	DQ168065
Gymnocephalus	AY141649-52	AY141373	AY141443	AY141296	AY362278	DQ168068
Halobatrachus	AY372743-44	AY368286	AY368308	AY368323	AY362246	DQ168069

Table 1: Continued

genus	28S	12S	16S	Rhodopsin	MLL	IRBP
Holacanthus	AY141753-56	AF055593	AF055614	AY141322	AY362214	DQ168072
Holanthias	AY141625-28	AY141366	AY141436	AY141290	AY362209	DQ168073
Kali	AY141697-00	AY141382	AY141452	AY141308	AY362224	DQ168074
Labrus	AY141737-40	**AF414201**	AY141462	AY141318	AY362222	DQ168075
Lagocephalus	AY141601-04	AY141360	AY141430	AY141285	AY362221	DQ168076
Lateolabrax	AY141633-36	AY141369	AY141439	AY141293	AY362253	DQ168078
Lepadogaster	AY141553-56	AY141347	AY141417	AY141273	AY362247	DQ168080
Liza	AY141521-24	AY141338	AY141408	AY141266	AY362248	DQ168082
Macroramphosus	AY141581-84	AY141354	AY141424	AY141280	AY362206	DQ168083
Mastacembelus	AY141561-64	AY141349	AY141419	AY141275	AY362249	DQ168084
Mene	AY141729-32	AY141390	AY141460	AY141316	AY362250	DQ168085
Microchirus	AY141597-00	AY141359	AY141429	AY141284	AY362275	DQ168086
Mola	AY141605-08	AY141361	AY141431	AY141286	AY362251	DQ168087
Monopterus	AY141565-68	AY141350	AY141420	AY141276	AY362252	DQ168088
Mullus	AY372719-21	AY368277	AF227680	Y18666	AY362231	DQ168090
Myripristis	AY141517-20	AY141337	AY141407	U57539	AY362265	DQ168091
Notothenia	AY141673-76	Z32712	Z32731	AY141302	AY362282	DQ168093
Ostracion	AY372722-23	AY368281	AF137213	AF137213	AY362207	DQ168095
Pampus	AY141701-04	AY141383	AY141453	AY141309	AY362220	DQ168096
Parablennius	AY141545-48	AY141345	AY141415	AY141271	AY362255	DQ168097
Pentanemus	AY141733-36	AY141391	AY141461	AY141317	AY362272	DQ168098
Perca	AY141645-48	AY141372	AY141442	AY141295	AY362279	DQ168099
Pholis	AY141657-60	AY141375	**AF420459**	AY141298	AY362285	DQ168100
Poecilia	AY141533-36	AY141342	**U80051**	AY141269	AY362203	DQ168102

Table 1: Continued

genus	28S	12S	16S	Rhodopsin	MLL	IRBP
Pogonoperca	AY372711-14	AY141368	**AF297322**	AY141292	AY362256	DQ168103
Polymixia	AY372724-26	AF049730	AF049740	AY368320	AY362208	DQ168104
Pomadasys	DQ021392	AY368293	AY368298	DQ021404	AY363643	AY362230
Psenopsis	AY141705-08	AY141384	AY141454	AY141310	AY362269	DQ168107
Psettodes	AY372717-18	AY368282	AY368302	AY368332	AY362259	DQ168108
Regalecus	AY372729-30	AY368292	AY368296	AY368328	AY362266	DQ168109
Rypticus	DQ021391	AY368295	**AF297327**	AY368329	AY362257	DQ168111
Scarus	AY141741-44	AY141393	AY141463	AY141319	AY362212	DQ168112
Scomber	AY141709-12	AY141385	**AB032521**	AY141311	AY362237	DQ168113
Scorpaena	AY141617-20	AY141364	AY141434	AY141288	AY362236	DQ168114
Serranus	AY141621-24	AY141365	AY141435	AY141289	AY362202	DQ168115
Sphyraena	AY141713-16	AY141386	AY141456	AY141312	AY362254	DQ168118
Taurulus	AH011857	AY141363	AY141433	U97275	AY362217	DQ168121
Trachinotus	AY141721-24	AY141388	AY141458	AY141314	AY362263	DQ168120
Trachinus	AY141681-84	AY141378	AY141448	AY141304	AY362277	DQ168123
Trachyrincus	AY372708-10	AY368280	AY368301	AY368318	AY362289	DQ168124
Triacanthodes	DQ021383, 96	AY368289	AY368311	AY368331	AY362258	DQ168125
Uranoscopus	AY141685-88	AY141379	AY141449	AY141305	AY362239	DQ168126
Zenopsis	AY372748-50	AY368278	AY368300	AY368314	AY362286	DQ168127
Zeus	AY141493-96	AY141331	AY141401	Y14484	AY362287	DQ168128

Why two clades with the same repetition index, same number of occurrences and contradicting one another should not have a high repetition index?

Let α and β be two clades that have the same repetition index R and the same number of occurrences M, and that contradict one another.

Let's also assume that each clade has a stable best contradictor.

If β is the best contradictor of α, then their repetition index is 0.

Is it possible that α and β have a strictly positive repetition index?

Let r be the function that to a clade γ associates its final repetition index $r(\gamma)$.

Let O be the function that to a clade γ associates its maximum number of occurrences (or sum of supports) over independant analyses $O(\gamma)$.

Let C be the function that to a clade γ associates its best contradictor $C(\gamma)$.

$r(\gamma) = O(\gamma) - O(C(\gamma))$

Let C_i be the sequence of the clades such that $C_0 = \alpha$ and for $i \geq 0$: $C_{i+1} = C(C_i)$.

Let O_i be the sequence of the number of ocurrences of the clades C_i. For $i \geq 0$: $O_i = O(C_i)$.

Let r_i be the sequence of the repetition indices of the clades C_i. For $i \geq 0$: $r_i = r(C_i)$.

$C_0 = \alpha$

$O_0 = O(\alpha) = M$

$r_0 = r(\alpha) = R$

Let us suppose (1): $R > 0$.

β is one of the contradictors of α, therefore $r_1 \geq r(\beta) = R$.

If $r_1 = R$, then $O_1 \geq O(\beta) = M$.

Then $R = r_0 = O_0 - O_1 \leq M - M = 0$: impossible because (1).

Therefore (2): $r_1 > R$.

$r_0 = O_0 - O_1$, therefore $O_1 = O_0 - r_0 = M - R$.

α is one of the contradictors of C_1, therefore $r_2 \geq r(\alpha) = R$.

If $r_2 = R$, then $O_2 \geq O(\alpha) = M$.

Now $O_1 = M - R$ and $r_1 = O_1 - O_2$.

Therefore $r_1 = O_1 - O_2 \leq M - R - M = -R$: impossible because (1) and (2).

Therefore $r_2 > R$.

$r_1 = O_1 - O_2$, therefore $O_2 = O_1 - r_1 = M - R - r_1 < M - 2R$.

Property P_i: $r_i > r_{i-2} \geq R$ and $O_i < M - iR$.

Let us try to demonstrate P_i, for $i \geq 2$.

P_2 is true: $r_2 > R = r_0$ and $O_2 < M - 2R$.

Let j be an interger such that $j > 2$.

Let us suppose that P_i is true for each integer i such that $2 \leq i < j$.

C_{j-2} is a contradictor of C_{j-1}, therefore $r_j \geq r_{j-2}$.

If $r_j = r_{j-2}$, then $O_j \geq O_{j-2}$.

Then $r_{j-1} = O_{j-1} - O_j \leq O_{j-1} - O_{j-2}$.

Now $O_{j-1} - O_{j-2} = -r_{j-2}$ and $r_{j-2} > R$ (from (2) if $j = 3$, or from P_{j-2}).

Therefore $r_{j-1} < -r_{j-2} < -R$: impossible because $r_{j-1} > R$ (from P_{j-1}).

Therefore $r_j > r_{j-2} \geq R$.

$r_{j-1} = O_{j-1} - O_j$, therefore $O_j = O_{j-1} - r_{j-1}$.

Now $O_{j-1} < M - (j-1)R$ and $r_{j-1} > R$ (from P_{j-1}).

Therefore $O_j < M - (j-1)R - R = M - jR$: P_j is true.

P_2 is true, therefore P_i is true for every integer i such that $i \geq 2$.

For i an integer such that $i > M/R$, $O_i < 0$ (from P_i).

However, by definition, $O_i \geq 0$ for every value of i.

Therefore (1) cannot be true:

Two clades with the same repetition index and the same number of occurrences that contradict one another cannot have a strictly positive repetition index (in the case where all clades have a stable best contradictor).

If the repetition index is an average, it's more difficult to see if α and β have a low repetition index.

References

Dettaï, A. (2004), *La phylogénie des Acanthomorpha (Teleostei) inférée par l'étude de la congruence taxinomique*. Ph.D. thesis, Université Paris VI Pierre et Marie Curie.

Dettaï, A. & Lecointre, G. (2005), Further support for the clades obtained by multiple molecular phylogenies in the acanthomorph bush. *Comptes Rendus Biologies*, *328*, 674–689.