The Data, the Matrix, and the Message: Comments on Begle’s “Relationships of the Osmeroid Fishes”

COLIN PATTERTSON and G. DAVID JOHNSON

1Natural History Museum, London SW7 5BD, London, UK
2National Museum of Natural History, Washington, D.C. 20560, USA

In 1991, this journal (then Systematic Zoology) included a paper on osmeroid fishes (smelts and their relatives) and the use of reductive characters in phylogenetic analysis (Begle, 1991). That study was based on a matrix of 84 characters coded in 19 taxa and a hypothetical ancestor. Thirteen of the taxa were genera, and the other six were collective at higher levels. In a revision of the systematics of osmeroids and other lower euteleostean fishes (Johnson and Patterson, 1996), we had to pay close attention to Begle’s paper. It was immediately obvious that his matrix contained many mistakes, and we eventually found it necessary to check each entry against specimens. Begle’s matrix is reproduced in Appendix 1 with our corrected coding and a summary of differences between the two versions. A detailed commentary was published elsewhere (Johnson and Patterson, 1996), and an abbreviated commentary on some characters is provided in Appendix 2.

RESULTS

When we analyzed Begle’s matrix, as published, with Hennig86 (Farris, 1988), we found the same result as Begle did: two shortest trees with a length of 128 steps and a consistency index of 0.69 when the 14 autapomorphic characters are excluded. The two trees differ only in how they relate salmonoids and neoteleosts, two outgroups to the osmeroids; one tree is shown in Figure 1a (cf. Begle, 1991: fig. 3).

The corrected matrix (Appendix 1) gives very different results. The results depend on the treatment of polymorphism within collective terminals, entered in the matrix as B for polymorphism in states 0 and 1 and C for polymorphism in states 1 and 2. Begle treated polymorphism in collective terminals as missing data (e.g., ? in the data matrix for character 55 in salmonoids and characters 63 and 82 in argentinoids). Using that method, with B’s and C’s coded as ?, the corrected matrix gives 225 equally parsimonious trees; the strict consensus is shown in Figure 1b. If B’s are coded as 0 and C’s as 1, which in every case is the most parsimonious interpretation of their distribution in the collective terminals (for details, see Johnson and Patterson, 1996), the corrected matrix gives 64 shortest trees; the strict consensus is shown in Figure 1c. Of the 15 nodes in Figure 1a, only 4 are found in Figures 1b or 1c. Major differences between Figures 1b or 1c and Figure 1a are (1) the position of Argentinioidei, placed in the basal polytomy rather than as the sister of osmeroids; (2) the position of esocoids, placed within osmeroids (as the sister of Lepidogalaxias, Lovettia, Aplochiton, and galaxiids in Fig. 1c) rather than in the basal polytomy; (3) the position of salangids, placed no nearer to Lepidogalaxias, Lovettia, Aplochiton, and galaxiids than are esocoids; and (4) the lack of resolution within osmerids (node 1 in Figs. 1b, 1c). The differences between Figure 1a and Figures 1b and 1c demonstrate the effects of Begle’s errors in coding, but the results from the corrected matrix reflect weakness in Begle’s choice of characters rather than the relationships of the taxa concerned (Fig. 1d).

COMMENTS ON OUR RECODING

Errors in Begle’s matrix (Appendix 1) range from elementary to arcane. At the elementary end of the scale are entries of undescribed states (e.g., state 2 in character 9, states 2 and 3 in character 16, state 2 in character 26, state 2 in character 53). These errors are sufficiently obvious to ring warning bells about the quality of other aspects of the work. Almost as elementary, in that spotting them requires no knowl-
edge of fishes, are mistakes in which a state, rather than a ?, is entered for features of a structure that is absent. For example, under character 43, the uncinate process of the second pharyngobranchial was (correctly) coded by Bégle as absent in galaxiids, Aplochiton, Retropinna, Stokellia, and Prototroctes (it is absent also in Lovettia). Character 45 concerns the orientation of that uncinate process; galaxiids, Aplochiton, Retropinna, Stokellia, and Prototroctes were coded by Bégle as having the missing process directed laterally or caudally. The cluster of ? and asterisks entered in the data matrix for Salangidae under characters 39–45 concern a more technical aspect of uncinate processes on gill-arch elements. Salangids are neotenic, and their gill arches are entirely cartilaginous (Roberts, 1984). Uncinate processes can be discriminated only in ossified gill-arch elements through separation by bone from
the cartilaginous head of the structure; salangids should therefore be coded? for all characters concerning uncinate processes. At the arcane end of the spectrum is character 71, concerning a process on the pterosphenoid bone. On investigation, we had to delete that character and replace it with four two-state characters (71A–D in Appendix 1). The first character concerns presence or absence of the pterosphenoid: it is absent in Lovettia and salangids, which must therefore be coded? for any feature of the bone. The other three characters (71B–D) concern three different, nonhomologous processes on the pterosphenoid, one dorsomedial, one ventral, and one posteroverentral, directed toward the prootic.

In addition to Begle's original character 71, we deleted or discarded 13 other characters. Reasons for discarding them range again from elementary to arcane. Examples at the elementary end of the spectrum include character 25, caudally directed "pegs" on the basioccipital, coded as present in galaxiids. The "pegs" occur only in one species of a group comprising six genera and almost 40 species, Galaxias fasciatus, the species illustrated in McDowall's (1969) survey of galaxiid osteology. Another deleted character also concerns McDowall's (1969) publication. Begle (character 67) used curvature of the maxilla to discriminate galaxiids and Aplochiton from other taxa in the matrix. We believe this character is based on a misreading of McDowall's (1969) figure 7, showing the maxilla and premaxilla in eight galaxiids, and results from taking 7H, 7C, and 7F to be 7F, 7G, and 7H. Similarly, Begle's character 81 discriminates Argentinoidae from other taxa as having expanded or elongate gill rakers on the fourth and fifth gill arches. We believe this character is based on a misreading of Greenwood and Rosen's (1971) account of the cruminal organ in Argentinoidae (argentinoids and alepocephaloids), not on attributes of specimens. Several characters were deleted because we could not confirm supposed differences between taxa (e.g., nos. 44, 45, orientation of uncinate processes on second and third pharyngobranchials; no. 64, form of first basibranchial; no. 74, form of postorbital process). Character 64, form of first basibranchial, is significant as the only credible character of Begle's Osmerae (Argentinoidae + Osmeroidae) (Fig. 1). We found (Johnson and Patterson, 1996: fig. 8) that Begle's (1991: fig. 5) "derived" state occurs in such varied teleosts as Elops, herrings (e.g., Chirocentrus), and stomiiforms and does not occur in argentinoids. Some characters were deleted for technical reasons, such as no. 19, which represents a state of character 6 that was unrecognized there, and no. 23, which concerns the position of a misidentified bone (dermopalatine taken to be ectopterygoid). Some characters were deleted because the supposed primitive state does not exist (e.g., no. 63, pterosphenoids meeting in midline; no. 77, first preural centrum with more than one neural arch).

In Appendix 1, an asterisk is entered beneath every instance where we detected an error in Begle's coding. Those asterisks do not always coincide with differences between our coding and his because in some characters we added extra states.

Statistics

As Appendix 1 shows, according to our reading of the specimens and the literature, Begle's matrix contains errors in 68 (81%) of his 84 characters. Among the 16 characters that are error free, 7 (almost half) are autapomorphic, leaving only 9 that might be used to group taxa. Of those nine, five (nos. 5, 7, 18, 21, 54) were taken from Fink (1984) and three (nos. 29, 30, 51) were from Howes and Sanford (1987a, 1987b).

A more accurate way of assessing Begle's error rate is to compare the number of asterisks in Appendix 1 (each an error) with the total number of entries. Excluding seven deleted characters (nos. 62–64, 67, 71, 74, 77) in which errors cannot be assessed for one reason or another, the total entries in the matrix are 19 × (84 − 7) = 1,463. There are 165 asterisks in Appendix 1, giving an error rate of 11%.

Discussion

With general acceptance of the cladistic approach, it is now commonplace for systematists to publish a matrix, with only a brief list of characters, and to concentrate on
the results of manipulating that matrix with parsimony or other programs. In other words, the emphasis has shifted from observation, the source of the matrix, to whatever message may be extracted from the matrix. So it was with Begle's (1991) paper, where the emphasis was on the use of reductive characters in phylogenetic analysis (all but two [nos. 7, 18] of his 31 "reductive" characters included errors in coding). This change of emphasis replaces our pernicious old black box, evolutionary systematics, with a new one, the matrix. In general, systematists regard matrices published in respected journals such as Systematic Biology as reliable because they have been carefully vetted. In consequence, systematists often take data sets from the literature and use them in comparative studies; Begle's (1991) data have already been used in that way (Colless, 1995), and in a discussion of Colless's (1995) work, Heard and Mooers (1996:117) assessed Begle's data as one of two sets (out of 11) in which the quality was good. On the contrary, as we have shown, Begle's data are riddled with error, as will be analyses based on them. Nevertheless, Begle's paper has influenced other workers. For example, Stearley and Smith (1993) followed Begle in selecting Thaleichthys as the most primitive osmerid (Fig. 1a), whereas it and Spirinchus are actually the most derived (Fig. 1d); and the systematic treatment of osmeroids in Nelson's (1994) textbook followed Begle's work in several groups. Few will have the material, the specialized knowledge, and the incentive necessary to check published matrices as thoroughly as we checked Begle's, but the warning should be clear: If the primary work, studying specimens, is not done with as much care and in as much detail as possible, what follows will be weakened and can be almost worthless.

There is also a message here for editors and reviewers. Standards in our professional journals are ultimately the responsibility of editors and of boards or committees, but in practice much of the responsibility passes to expert reviewers. The early warnings in Begle's matrix were obvious; no particular expertise is necessary to spot entries for undescribed states of characters. Signs of negligence or incompetence will not always be so clear, and the responsibilities of reviewers may be harder to meet. All systematists can play a part by finding time to review manuscripts in our fields and by using our knowledge of those fields to check the data provided with as much attention as we would give to conclusions drawn from those data.

ACKNOWLEDGMENTS

For comments on a draft, we are grateful to C. C. Baldwin, J. A. Coddington, S. L. Jewett, L. R. Parenti, R. H. Rosenblatt, V. G. Springer, and S. H. Weihtman.

REFERENCES


Received 9 May 1996; accepted 20 December 1996

Associate Editor: David Cannatella
## APPENDIX 1

### DATA MATRIX

Begle's (1991) matrix of 84 characters in 19 taxa of teleost fishes and a hypothetical ancestor was recoded after checking each entry against specimens (or against the literature where specimens were inadequate, e.g., characters 4, 24). Characters are described in Appendix 2. There are three rows for each taxon. The upper row is Begle's coding. The second row is our coding, where X = deleted characters, B = polymorphism in states 0 and 1, C = polymorphism in states 1 and 2, and underlining = inferred errors in Begle's coding. The third row summarizes those errors with an asterisk beneath each underlined entry and a total at the end. These asterisks do not coincide with all differences between Begle's coding and ours because in some characters our coding differs through deleting nonexistent states.

<table>
<thead>
<tr>
<th>Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ancestor</td>
</tr>
<tr>
<td>Galaxiinae</td>
</tr>
<tr>
<td>Lovettia</td>
</tr>
<tr>
<td>Aplochiton</td>
</tr>
<tr>
<td>Salangidae</td>
</tr>
<tr>
<td>Retropinna</td>
</tr>
<tr>
<td>Stokella</td>
</tr>
<tr>
<td>Prototroctes</td>
</tr>
<tr>
<td>PlecoGLOSSUS</td>
</tr>
<tr>
<td>Allosmerus</td>
</tr>
<tr>
<td>Hypomusus</td>
</tr>
<tr>
<td>Omerus</td>
</tr>
<tr>
<td>Spirinchus</td>
</tr>
<tr>
<td>Mallothus</td>
</tr>
<tr>
<td>Thaleichthys</td>
</tr>
<tr>
<td>Lepidogalaxias</td>
</tr>
<tr>
<td>Argentinobeli</td>
</tr>
<tr>
<td>Bocoida</td>
</tr>
<tr>
<td>Salmonoid</td>
</tr>
<tr>
<td>Neoteleostei</td>
</tr>
</tbody>
</table>
APPENDIX 2

CHARACTER LIST

Characters are from Bogle (1991: appendix 1). Comments in brackets detail changes we made or add explanation. Full commentary was provided by Johnson and Patterson (1996).

1. **Posterior shaft of vomer.**—0 = long; 1 = short. [State 2 added for absence of vomer in *Aplachiton* and salangids.]
2. **Articular.**—0 = fused; 1 = reduced. [Fusion with the angular is general in Clupeocephala; reduced or delayed ossification occurs in all osmeroids and in esocoids.]
3. **Mesopterygoid teeth.**—0 = broadly distributed over surface of bone; 1 = narrow band of larger teeth along mesial border; 2 = absent.
4. **Anchor membrane of egg.**—0 = egg not adhesive; 1 = egg surrounded by adhesive “anchor” membrane.
5. **Caudal fusion pattern.**—0 = uroneurals and neural arches unfused or, if fused, arches fuse first with compound centrum, then with uroneurals; 1 = uroneural 1 fuses to compound centrum, followed in some forms by fusion with neural arches.
6. **Orientation of infraorbital sensory canals.**—0 = anterior portion of infraorbital canal posterodorsally deflected; 1 = anterior portion of canal posteroventrally deflected. [Two conditions are conflated under state 1: the galaxiid condition, in which the postorbital infraorbital bones and sensory canal are absent, here coded 2, and the retropinnid condition, in which the postorbital bones and canal are present but the canal is interrupted behind the eye and the anterior portion runs posteroventrally to cross the preopercular canal, here coded 1. *Lepidogalaxias* lacks all infraorbital bones and the sensory canal and is coded 3. See also no. 19.]
7. **Mesocoracoid.**—0 = present; 1 = absent.
8. **Dorsal fin position.**—0 = anterior; 1 = posterior.
9. **Principal caudal rays.**—0 = 10–19; 1 = 9–18 or fewer.
10. **Palatine teeth.**—0 = present; 1 = absent.
11. **Ectopterygoid.**—0 = present; 1 = absent. [In retropinnids, a small ectopterygoid is normally fused with the dermal palatine; Bogle identified the entire bone as the ectopterygoid. See also no. 23.]
12. **Extrascapular.**—0 = present; 1 = absent.
13. **Cleithrum process.**—0 = cleithrum with ventral process meeting coracoid; 1 = cleithrum without ventral process. [Two conditions are conflated under state 0: in the primitive state, here coded 0, there is a long-based triangular process and the coracoid lies against it; in the osmerid/retropinnid state, here coded 1, there is a narrow-based columnar process that interdigitates with the coracoid. Absence is here coded 2.]
14. **Posterior pubic symphysis.**—0 = present; 1 = absent.
15. **Scales.**—0 = present; 1 = absent.
16. **Vomerine teeth.**—0 = present; 1 = absent. [Vomer absent in *Aplachiton* and salangids, here coded ?.]
17. **Posterior border of suspensorium.**—0 = smooth; 1 = emarginate. [Character refers to the margin of the opercular bones, also serrated or fimbriate in many argentinoids and alepocephaloids.]
18. **Number of hypurals.**—0 = six; 1 = five.
19. **Extent of infraorbital canal.**—0 = not extending to preopercle; 1 = extending to preopercle. [State 1, in retropinnids, here added to no. 6; this character deleted.]
20. **Ceratohyal border.**—0 = more or less straight, with branchiostegals along most of its length; 1 = deeply concave anteriorly, with branchiostegals restricted to area posterior to concavity.
21. **Horny midventral abdominal keel.**—0 = absent; 1 = present.
22. **Ovaries.**—0 = left and right present; 1 = right only. [It is the left ovary only that is present in retropinnids (and *Plecoglossus*).]
23. **Location of ectopterygoid.**—0 = posterior to autopalatine; 1 = ventral to autopalatine. [The bone is the dermal palatine (no. 11); character deleted.]
24. **Cucumber odor.**—0 = fresh specimens lacking cucumber odor; 1 = fresh specimens with distinctive cucumber odor.
25. **Basiooccipital lateral pegs.**—0 = basiooccipital without lateral extensions; 1 = with pegs projecting caudally along each side of first vertebra. [Character deleted; see comment in text.]
26. **Lateral hyomandibular spur.**—0 = hyomandibular without lateral projection; 1 = with lateral spur at level of its articulation with opercle.
27. **Posterior border of caudal fin.**—0 = incised; 1 = rounded. [Among esocoids, this fin is forked in esocids, rounded in umbriids.]
28. **Adipose fin.**—0 = present; 1 = absent.
29. **Articulation of premaxilla with maxilla.**—0 = premaxillary articulation not tightly adhering to maxillary head; 1 = syndesmosis between the two bones.
30. **Palatine-maxilla articulation.**—0 = palatine without articulatory head meeting similar head on maxilla; 1 = maxilla and palatine with head-to-head articulation.
31. **Sympophysial cartilages.**—0 = paired cartilages not present at dentary symphysis; 1 = paired cartilages present.
32. **Ectopterygoid flange.**—0 = dorsal rim of ectopterygoid not directed laterally; 1 = dorsal rim directed laterally in a horizontal flange.
33. **Nuptial tubercles.**—0 = present; 1 = absent. [? entered for hypothetical ancestor because the only outgroup taxa with tubercles are ostariophysans (some gonorynchiforms, some otophysans). In neoteleosts, tubercles are recorded only in Percidae and Gadidae; the group must be coded 1.]
34. **Petaloid girdle.**—0 = without ventral condyle; 1 = with ventral condyle articulating with the first three or four hemitr�cs.
35. **Autopalatine.**—0 = not in the shape of a dumbbell; 1 = in distinctive dumbbell shape.
36. **Adipose cartilage.**—0 = absent; 1 = present; 2 = present, pear shaped.
37. **Dentary symphysis.**—0 = without medial process; 1 = with medial tusklike process.
38. **Metapterygoid.**—0 = without lateral shelf; 1 = with
short lateral shelf; 2 = with prominent diagonal shelf.
39. Uncinate process on first epibranchial.—0 = present; 1 = absent.
40. Uncinate process on second epibranchial.—0 = present; 1 = absent.
41. Uncinate process on third epibranchial.—0 = present; 1 = absent.
42. Uncinate process on third pharyngobranchial.—0 = present; 1 = absent.
43. Uncinate process on second pharyngobranchial.—0 = present; 1 = absent.
44. Uncinate process of third pharyngobranchial.—0 = not extending over second epibranchial; 1 = extending over second epibranchial. [We could not confirm the difference between Hypomesus and other osmerids implied by Begg's coding, found the osmerid condition in various arginodontids, and found that distinguishing the states is in part subjective. Character problematic and so deleted.]
45. Uncinate process of second pharyngobranchial.—0 = directed laterally or caudally; 1 = directed anteriorly. [We found state 1 more widely distributed (e.g., in salmoninoids and arginodontids), and distinguishing the states is, in part subjective. Character problematic and so deleted.]
46. Hypomandibula.—0 = not fused to palatine [pala-toquadrate intended]; 1 = fused with palatine [palaquadrate].
47. Third pharyngobranchial.—0 = with narrow anterior extension, reaching first pharyngobranchial; 1 = without anterior extension. [We divided Begg's state 1, with state 1 reserved for Pb3 of intermediate length, extending anteriorly alongside about two-thirds of Pb2, and state 2 for Pb3 without anterior extension.]
48. Occiput.—0 = not greatly depressed; 1 = greatly depressed.
49. Separate fourth hypobranchial.—0 = absent; 1 = present.
50. Levator process on fourth epibranchial.—0 = wide; 1 = narrow. [We interpreted state 0 as absence of a levator process and added state 2 for esocoids, in which Eb4 has an uncinate process but lacks a levator process or a posterior elevation (except in Umbra).]
51. Nasal lamellae.—0 = arranged in a rosette; 1 = parallel, longitudinal.
52. Preopercular and supraorbital sensory canal pores.—0 = five or more; 1 = three.
53. Anal rays/scales.—0 = not modified in males; 1 = scales anterior to dorsal margin of anal fin greatly enlarged, anal-fin rays and supports also more or less modified.
54. Retractor dorsalis.—0 = absent; 1 = present.
55. Occipital condyle.—0 = formed by basioccipital; 1 = formed by basioccipital and exoccipitals.
56. Temporal sensory canal.—0 = present; 1 = absent.
57. Basiphenoid.—0 = present; 1 = absent.
58. Orbitosphenoid.—0 = present; 1 = absent.
59. Fusion of fifth epibranchial to form circular foramen for efferent artery.—0 = no; 1 = yes. [Two different patterns are conflated under state 1: in the osmerid/salangid pattern, here coded 2, there is a levator process on Eb4 and a vascular notch is enclosed between it and Eb5, which fuses with Eb4 at its lower end; in the galaxiid/aplochitonid pattern, here coded 1, there is no levator process on Eb4 and fusion between Eb5 and Eb4 is initiated at its upper end.]
60. Sculpture on quadrate.—0 = absent; 1 = present. [The "sculpture" is struts of membrane bone, commonly developed in other teleosts.]
61. Sculpture on metapterygoid.—0 = absent; 1 = present. [See no. 60.]
62. Ethmoid.—0 = without foramen; 1 = with midlateral foramen. [The "foramen" is the termination of the groove for the hyoidian artery, primitive for teleosts. Character deleted.]
63. Pterosphenoid.—0 = not reduced, meeting on midline; 1 = reduced, separated. [State 0 does not exist; character deleted.]
64. First basibranchial.—0 = unmodified; 1 = with ventral cartilaginous vane. [See comments in text; character deleted.]
65. Anterodorsal border of opercle.—0 = without spine, horizontal; 1 = with notch and spine; 2 = with deep narrow notch.
66. Ascending process of premaxilla.—0 = short, knob-like; 1 = sharply triangular.
67. Maxilla.—0 = more or less straight in lateral profile; 1 = curved dorsally in lateral profile. [See comments in text; character deleted.]
68. Epipharyngeal.—0 = usually greater than half length of ceratohyal; 1 = short, very much less than half length of ceratohyal.
69. Fifth ceratobranchial.—0 = without laminar extensions; 1 = with anterior laminal extension.
70. Palatine.—0 = contacting maxilla by small knob, if at all; 1 = lateral knob overlying maxilla.
71. Pterosphenoid.—0 = without ventral process; 1 = with ventral flange midway along its length. [Character deleted, replaced by four separate characters (71A-D); see comments in text.]
72. Anterior extent of prootic.—0 = smooth; 1 = notched with a dorsal projection. [See no. 73.]
73. Prootic/pterosphenoid contact.—0 = at dorso lateral part of prootic; 1 = more medial, contact by interdigitations. [Character deleted, derived condition added as state 2 of no. 72.]
74. Lateral spine on sphenotic.—0 = blunt; 1 = rodlike. [Character deleted; see comment in text.]
75. Pyloric caeca.—0 = present; 1 = absent.
76. Urobranchials.—0 = more than one; 1 = one.
77. First preural centrum (PU1).—0 = more than one neural arch over PU1; 1 = a single rudimentary neural arch over PU1. [Character deleted; state 1 is primitive, state 0 occurs only in some arginodontids and galaxiids.]
78. Accessory cartilage at tip of fifth ceratobranchial.—0 = absent; 1 = present.
79. Fourth ceratobranchial.—0 = unmodified; 1 = distal end expanded.
80. Fleshy membrane on fourth arch.—0 = absent; 1 = present. [The membrane is a component of the
81. Gill rakers on fourth and fifth arches.—0 = unmodified; 1 = expanded/elongate. [Character deleted; see comment in text.]

82. Opercle.—0 = extending dorsally above articulation with hyomandibular; 1 = not extending above articulation.

83. Alveolar process of premaxilla.—0 = short; 1 = extended, more than half length of maxilla.

84. Midlateral band of silver pigment.—0 = absent; 1 = present.