

**Supplementary material**

**A data-driven method for selecting candidate reference sites for stream bioassessment programs using generalised dissimilarity models**

*P. M. Rose<sup>A,D</sup>, M. J. Kennard<sup>A</sup>, F. Sheldon<sup>A</sup>, D. B. Moffatt<sup>B</sup> and G. L. Butler<sup>C</sup>*

<sup>A</sup>Australian Rivers Institute, Griffith University, Kessels Road, Nathan, Qld 4111, Australia.

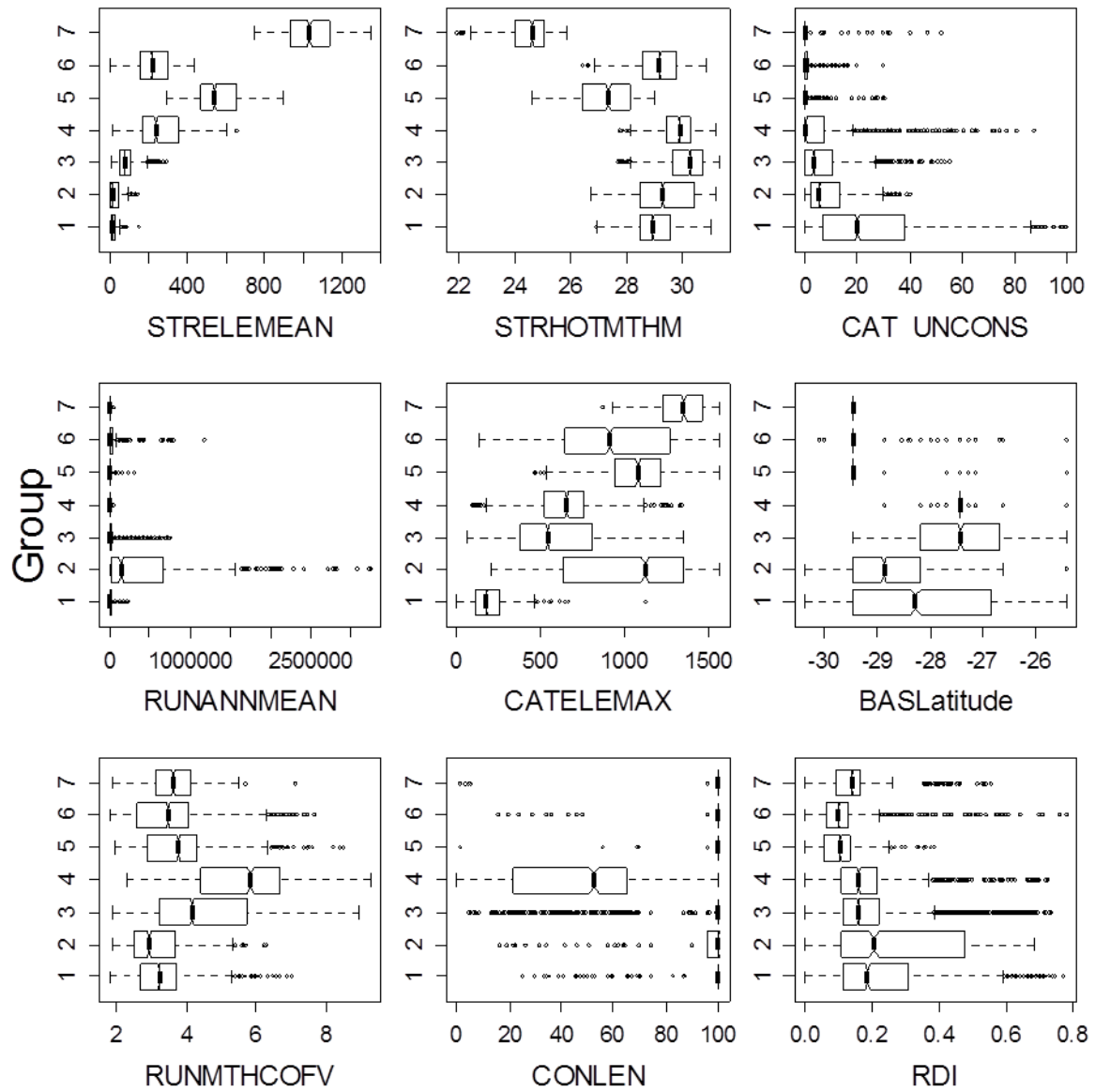
<sup>B</sup>Department of Science, Information Technology, Innovation and the Arts, EcoSciences Precinct, GPO Box 5078, Brisbane, Qld 4001, Australia.

<sup>C</sup>NSW Department of Primary Industries, Grafton Fisheries Centre, Private Mail Bag 2, Grafton, NSW 2460, Australia.

<sup>D</sup>Corresponding author. Email: [peter.rose@griffithuni.edu.au](mailto:peter.rose@griffithuni.edu.au)

**Table S1. List of fish species included in the analysis, and their prevalence in the generalised dissimilarity modelling (GDM) training dataset ( $n = 396$ )**

Species	Prevalence (%)
<i>Ambassis agassizii</i>	33.8
<i>Ambassis marianus</i>	2.3
<i>Anguilla australis</i>	28.5
<i>Anguilla reinhardtii</i>	83.3
<i>Arrhamphus sclerolepis</i>	1.5
<i>Craterocephalus marjoriae</i>	21.7
<i>Craterocephalus stercusmuscarum</i>	18.2
<i>Galaxias olidus</i>	3.0
<i>Glossamia aprion</i>	3.5
<i>Gobiomorphus australis</i>	49.2
<i>Gobiomorphus coxii</i>	21.0
<i>Hypseleotris compressa</i>	38.1
<i>Hypseleotris galii</i>	71.5
<i>Hypseleotris klunzingeri</i>	38.9
<i>Kuhlia rupestris</i>	0.5
<i>Maccullochella ikei</i>	6.6
<i>Macquaria novemaculeata</i>	30.3
<i>Maccullochella peelii mariensis</i>	1.0
<i>Melanotaenia duboulayi</i>	67.9
<i>Mogurnda adspersa</i>	21.5
<i>Mordacia</i> sp. (undescribed)	0.3
<i>Mugil cephalus</i>	25.8
<i>Nannoperca oxleyana</i>	2.0
<i>Nematalosa erebi</i>	8.3
<i>Neoceratodus forsteri</i>	0.3
<i>Neosilurus hyrtlii</i>	0.5
<i>Notesthes robusta</i>	13.9
<i>Ophisternon</i> sp.	2.8
<i>Philypnodon grandiceps</i>	40.7
<i>Philypnodon macrostomus</i>	25.5
<i>Porochilus rendahli</i>	0.3
<i>Potamalosa richmondii</i>	10.6
<i>Pseudomugil mellis</i>	0.8
<i>Pseudomugil signifer</i>	27.0
<i>Retropinna semoni</i>	58.1
<i>Rhadinocentrus ornatus</i>	14.6
<i>Trachystoma petardi</i>	18.4
<i>Tandanus tandanus</i>	73.0



**Fig. S1.** Notched boxplots of retained untransformed predictor variables for each ecotype. Non-overlapping notches roughly indicate significant differences among median values. See Table 1 in the main paper for predictor-variable descriptions.

**Table S2. Differences in assemblage composition among generalised dissimilarity modelling (GDM)-defined ecotypes using the evaluation dataset ( $n = 128$ ), assessed using PERMANOVA**

Probabilities are significant at: \*,  $P < 0.05$ ; \*\*,  $P < 0.01$

Pairwise ecotype comparison	Pseudo- $F$	$P$ -value
1 v. 2	16.1	0.001***
1 v. 3	25.6	0.001***
1 v. 4	23.1	0.001***
1 v. 5	25.5	0.001***
1 v. 6	45.3	0.001***
1 v. 7	10.5	0.002***
2 v. 3	8.0	0.001***
2 v. 4	12.0	0.001***
2 v. 5	22.6	0.001***
2 v. 6	20.9	0.001***
2 v. 7	10.8	0.001***
3 v. 4	2.8	0.023*
3 v. 5	14.1	0.001***
3 v. 6	3.3	0.012*
3 v. 7	7.1	0.001***
4 v. 5	10.6	0.001***
4 v. 6	6.3	0.001***
4 v. 7	5.9	0.001***
5 v. 6	13.4	0.001***
5 v. 7	0.9	0.47
6 v. 7	9.5	0.001***
Global analysis	27.3	0.001***