

Bases de Dados. Dados do PPBio Semiárido

**Exemplos de conjunto de dados ecológicos do Programa de Pesquisa em
Biodiversidade (PPBio) do Semiárido**

Prof. Elvio S. F. Medeiros Laboratório de Ecologia
Universidade Estadual da Paraíba Campus V, João Pessoa, PB

17/01/2026

Para entender a distribuição das espécies de peixes e seu uso de habitat, uma série de variáveis ambientais foram avaliadas como preditores da composição e riqueza da assembleia de peixes em sistemas aquáticos tropicais semiáridos. Nós pesquisamos a composição de espécies de assembleias de peixes em sistemas aquáticos semiáridos e estabelecemos seu grau de associação com a estrutura do habitat aquático. Os locais consistiam em trechos de riachos com fluxo de água superficial, poças temporárias isoladas e reservatórios artificiais (açudes). A amostragem de peixes foi realizada em quatro ocasiões durante as estações chuvosa (abril e junho de 2006) e seca (setembro e dezembro de 2006).

Índice

1 Apresentação	2
2 Sobre os dados do PPBio	3
3 Arquivos disponíveis	4
3.1 Codificação das variáveis	4
3.2 Abreviações	4
Referências	12
Apêndices	12
Sites para consulta	12
Script limpo	12

Lista de Figuras

1	Parte da planilha de dados brutos do PPBio.	3
2	<i>Astyanax bimaculatus</i> , a espécie mais comum da matriz de dados ppbio. Peru, by Eakins, R. Fonte: https://www.fishbase.se/summary/Astianax-bimaculatus.html	4
3	<i>Hoplias malabaricus</i> , espécie que cresce para se tornar um importante predador. Brazil, by Roselet, F.F.G. Fonte: https://www.fishbase.se/summary/Hoplias-malabaricus.html	5
4	<i>Apareiodon sp.</i> , importante espécie bentopelágica das bacias dos rios Jaguaribe e Paraíba. Brazil, by Ramos, T.P.A. Fonte: https://www.fishbase.se/summary/Apareiodon-davisi.html	6
5	<i>Pseudancistrus genisetiger</i> , uma espécie endêmica das bacias hidrográficas do nordeste. By Medeiros, E.S.F. Fonte: Arquivo pessoal	7
6	Associação entre a planilha de dados brutos do PPBio e o delineamento amostral do estudo.	8
7	Codificação para as variáveis, espécies de peixes, sítios de amostragem e período de amostragem.	10

Lista de Tabelas

1	Matrizes disponíveis para análises, com suas descrições e tipos de dados recomendados.	9
3	Nomenclatura das matrizes em AMD em relação aos atributos das colunas e suas abreviações nos códigos do R.	11

1 Apresentação

Para entender a distribuição das espécies de peixes e seu uso de habitat, uma série de variáveis ambientais foram avaliadas como preditores da composição e riqueza da assembleia de peixes em sistemas aquáticos tropicais semiáridos. Nós pesquisamos a composição de espécies de assembleias de peixes em sistemas aquáticos semiáridos e estabelecemos seu grau de associação com a estrutura do habitat aquático. Os locais consistiam em trechos de riachos com fluxo de água superficial, poças temporárias isoladas e reservatórios artificiais (açudes). A amostragem de peixes foi realizada em quatro ocasiões durante as estações chuvosa (abril e junho de 2006) e seca (setembro e dezembro de 2006).

Palavras-chave: rios intermitentes, reservatórios, conservação, composição de substratos.

2 Sobre os dados do PPBio

Usaremos ao longo desse livro dados que fazem parte de um estudo mais amplo sobre ecologia de rios do semiárido, coletados no Programa de Pesquisa em Biodiversidade - PPBio (Veja [Programa de Pesquisa em Biodiversidade – PPBio](#)). Parte desses dados está armazenada em planilhas de Excel `ppbio**.xlsx` (Figura 1). Essas matrizes de dados são descritas na Tabela 1. As planilhas `ppbio**.xlsx` contém vários tipos de dados arranjados em matrizes $n \times m$ que incluem dados de abundância de espécies em diferentes unidades amostrais (UA's), dados da estrutura do habitat físico, e variáveis em escala de bacia hidrográfica, dados de contagem de indivíduos ajustados para Captura Por Unidade de Esforço (CPUE), etc (Figura 1).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	sítio	ap-davis	as-bimac	as-fasci	ch-bimac	ci-oceia	ci-orien	co-macro	co-heter	cr-menez	cu-lepid	cy-gilbe	ge-brasi	he-margi	ho-malab	hy-pusar	le-melan	le-plau
2	S-A-ZA1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	S-R-CC1	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	S-R-CT1	0	194	55	0	0	5	0	1	14	0	0	3	0	1	9	0	3
5	S-R-CP1	0	19	0	0	0	0	0	0	0	0	0	0	0	5	2	0	0
6	S-A-TA1	0	23	1	13	0	0	0	0	0	0	0	0	0	0	0	0	0
7	S-R-CT2	0	142	3	3	0	69	0	0	4	0	0	0	1	17	43	0	1
8	S-R-CP2	0	5	1	0	40	9	0	0	0	0	0	0	0	10	2	0	3
9	S-A-TA2	0	46	0	178	0	0	0	0	0	0	0	0	0	2	0	0	0
10	S-R-CT3	0	206	64	0	0	25	0	0	8	0	0	1	0	31	11	0	2
11	S-R-CP3	0	16	0	0	13	24	0	0	0	0	0	0	0	4	0	0	1
12	S-A-TA3	0	234	7	238	0	0	2	0	0	0	0	0	0	20	0	0	0
13	S-R-CT4	0	0	1	0	0	5	0	0	1	0	50	3	1	4	3	0	0
14	S-R-CP4	0	0	0	0	11	6	0	0	0	0	0	0	0	2	0	0	2
15	S-A-TA4	0	394	0	273	0	0	0	0	1	0	0	1	0	9	0	0	2
16	B-A-MU1	0	12	0	0	0	0	0	0	0	0	0	190	0	0	0	0	0
17	B-R-ET1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	B-A-GU1	0	2	2	0	0	0	0	0	0	0	0	7	0	0	0	0	0
19	B-R-PC2	5	44	0	0	2	0	0	0	0	0	0	8	0	0	0	2	0
20	B-A-MU2	0	99	0	0	0	0	0	0	0	0	0	67	0	1	0	0	0
21	B-A-GU2	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0
22	B-R-PC3	22	75	7	0	4	0	0	0	0	0	21	0	16	0	2	1	0
23	B-A-MU3	0	511	0	0	0	0	0	0	0	0	0	145	0	0	0	0	0
24	B-A-GU3	0	6	0	0	0	0	0	0	0	0	0	32	0	0	0	0	0
25	B-R-PC4	0	7	17	0	0	0	0	0	0	0	0	81	5	0	1	0	1
26	B-A-MU4	0	235	0	0	0	0	0	0	0	0	0	509	0	0	0	0	0
27	B-A-GU4	0	13	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0

Figura 1: Parte da planilha de dados brutos do PPBio.

Por exemplo, essa é a **matriz bruta de dados**, porque os valores ainda não foram ajustados para os valores de Captura Por Unidade de Esforço (CPUE), nem foram relativizados ou transformados. Outros tipos de arquivos existem sobre esses dados (Tabela 1).

Várias das espécies nessa matriz tem grande importância ecológica, como é o caso de *Astyanax bimaculatus*¹ (Figura 2), que é muito comum em rios intermitentes e serve de alimento para predadores maiores (Silva et al. 2018) como a espécie *Hoplias malabaricus*² (Figura 3) (Silva et al. 2010).

Outras espécies como *Apareiodon hasemani*³ (Figura 4) tem importância trófica por estar na base da cadeia alimentar, enquanto espécies da família Loricariidae, como *Pseudancistrus*

¹A etimologia do gênero *Astyanax* vem da mitologia Grega. Heitor personagem da “Ilíada”, tinha um filho chamado Astianax.

²Do Grego, *hoplon*, arma ou armadura, em referência aos dentes caniniformes muito desenvolvidos, e forte estrutura óssea na cabeça.

³A etimologia do nome *Apareiodon* vem do Grego, *a*, sem, *pareia*, lateral ou bochecha, e *odous* dentição, em referência a ausência de dentes laterais no aparato bucal dessa espécie.”



Figura 2: *Astyanax bimaculatus*, a espécie mais comum da matriz de dados ppbio. Peru, by Eakins, R. Fonte: <https://www.fishbase.se/summary/Astianax-bimaculatus.html>

*genisetiger*⁴ (Figura 5), tem importância para taxonomia.

As planilhas ppbio**.xlsx contém o delineamento amostral de um dos estudos do Projeto PPBio (Figura 6). Nas linhas são apresentadas as abreviações dos nomes das unidades amostrais (UA's) e nas colunas são apresentados os nomes abreviados das espécies - temos portanto uma matriz comunitária (Tabela 1). No corpo da planilha temos os valores para o tipo de dados amostrado. Quantitativo, semi-quantitativo ou qualitativo.

Qual desses tipos de dados você acha que é apresentado na planilha?

3 Arquivos disponíveis

A seguir, apresento uma tabela com as informações essenciais sobre as matrizes de dados que serão utilizadas ao longo deste livro (Tabela 1). Nela, estão descritos os diferentes tipos de matrizes, suas finalidades analíticas e o tipo de dado para cada uma delas. Essas informações servirão de referência para compreender a origem, a estrutura e o tratamento recomendado para os dados empregados nas análises subsequentes.

Você pode baixar essas matrizes clicando no link para o arquivo em .xlsx na coluna Arquivo. Ao vizualizar a matriz que deseja baixar, clique em Arquivo > Baixar > Microsoft Excel (.xlsx). Atente para a pasta onde seu computador baixa o arquivo desejado, você vai precisar dessa informação depois.

3.1 Codificação das variáveis

Os arquivos da base de dados do projeto são fornecidos em formato Excel (.xlsx). Por exemplo, o arquivo ppbio06-* .xlsx, traz os dados brutos que serão usados nas análises. A matriz de dados brutos contém mais de 20 localidades ($m=$ linhas ou objetos) em estações do ano diferentes, e cerca de 35 espécies ($n=$ colunas ou atributos), antes de qualquer modificação. Portanto é uma matriz bruta. Os valores são contagens de indivíduos, e apresentam uma alta amplitude de variação, portanto, o uso de uma matriz relativizada é sugerido (Tabela 1). Nos nomes dos objetos e dos atributos são codificados de acordo com a tabela mostrada na Figura 7).

3.2 Abreviações



Figura 3: *Hoplias malabaricus*, espécie que cresce para se tornar um importante predador. Brazil, by Roselet, F.F.G. Fonte: <https://www.fishbase.se/summary/Hoplias-malabaricus.html>



Figura 4: *Apareiodon* sp., importante espécie bentopelágica das bacias dos rios Jaguaribe e Paraíba. Brazil, by Ramos, T.P.A. Fonte: [https://www.fishbase.se/summary/*Apareiodon-davisi*.html](https://www.fishbase.se/summary/Apareiodon-davisi.html)



Figura 5: *Pseudancistrus genisetiger*, uma espécie endêmica das bacias hidrográficas do nordeste. By Medeiros, E.S.F. Fonte: Arquivo pessoal

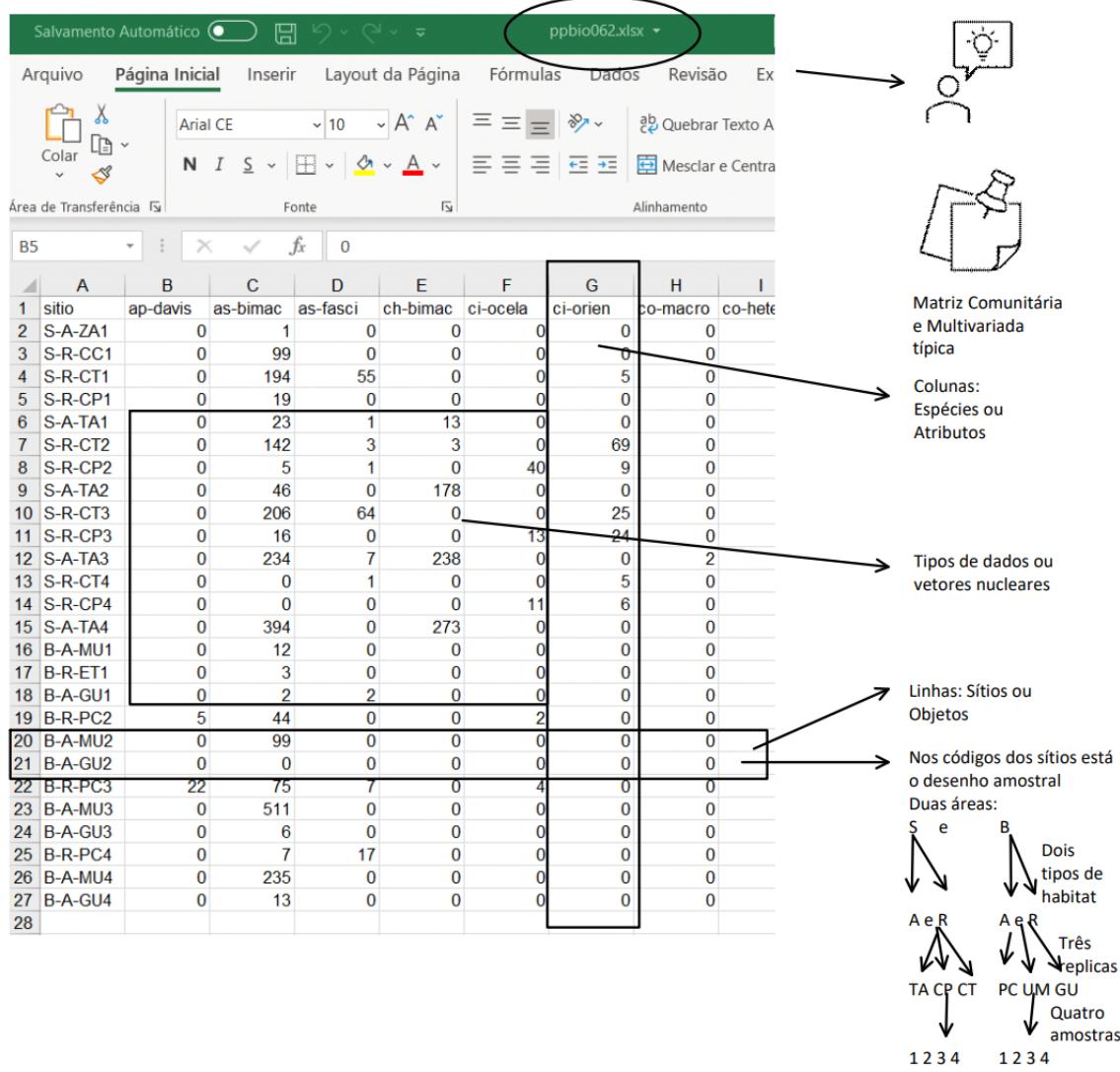


Figura 6: Associação entre a planilha de dados brutos do PPBio e o delineamento amostral do estudo.

Tabela 1: Matrizes disponíveis para análises, com suas descrições e tipos de dados recomendados.

Arquivo (.xlsx)	Tipo de matriz	Descrição	Tipo de dados
ppbio06c-peixes	Matriz comunitária	O arquivo ppbio06 traz os dados brutos que serão usados nas análises. A matriz de dados brutos contendo 26 localidades em estações do ano diferentes (objetos) x 35 espécies (atributos), antes de qualquer modificação.	Contagens de indivíduos com alta amplitude de variação, sugerido uso de matriz relativizada.
ppbio06p-amb	Matriz ambiental	O arquivo ppbio06h traz os dados brutos que serão usados nas análises. A matriz de dados brutos contendo 26 localidades em estações diferentes (objetos) x 35 variáveis ambientais (atributos) medidas em diferentes escalas espaciais, antes de qualquer modificação.	Unidades de medição diferentes (cm, m, °C, mg/L, etc.), com alta amplitude de variação, sugerido uso de matriz transformada e/ou reescalada.
ppbio06-grupos	Matriz de grupos	O arquivo ppbio06 traz os dados brutos que serão usados nessa análise. A matriz de dados brutos contendo 26 locais/ocasiões (objetos) x 35 espécies (atributos), antes de qualquer modificação.	Contagens de indivíduos com alta amplitude de variação, sugerido uso de matriz relativizada.
ppbio06-cpue	Matriz comunitária	O arquivo ppbio06cpue traz os valores após ajuste pela Captura Por Unidade de Esforço (CPUE).	Densidades de indivíduos com alta amplitude de variação, sugerido uso de matriz relativizada.

Espécies de peixes	Sítios e período de amostragem																												
ap-davis: Apareiodon hasemani as-bimac: Astyanax bimaculatus as-fasci: Astyanax fasciatus ch-bimac: Characidium bimaculatum ci-ocela: Cichla ocellaris ci-orien: Cichlasoma orientale co-macro: Colossoma macropomum co-hetero: Compsura heterura cr-menez: Crenicichla menezesi cu-lepid: Curimatella lepidura. cy-gilbe: Cyphocharax Gilbert ge-brasi: Geophagus brasiliensis he-margi: Hemigrammus marginatus ho-malab: Hoplias malabaricus hy-pusar: Hypostomus pusarum le-melan: Leporinus melanopleura le-piau: Leporinus piau le-taeni: Leporinus taeniatus mo-costa: Moenkhausia costae mo-lepid: Moenkhausia sp1 or-nilot: Oreochromis niloticus pa-manag: Parachromis managuensis pimel-sp: Pimelodella enochi po-retic: Poecilia reticulata po-vivip: Poecilia vivipara pr-brevi: Prochilodus brevis ps-losango: branquinha ps-rhomb: Psectrogaster rhomboides ps-genise: Pseudancistrus papariae se-heter: Serrapinnus heterodon se-piaba: Serrapinnus piaba se-spilo: Serrasalmus rhombeus st-noton: Steindachnerina notonota sy-marmo: Synbranchus marmoratus te-chalc: Tetragonopterus chalceus tr-signa: Triportheus signatus	<p>Nos códigos dos sítios está o desenho amostral. Duas áreas, ao norte o Seridó (S) e ao sul Buíque (B). Em cada área existem dois tipos de ambiente, açude (A) e rio (R). Ao norte estão três sítios, TA, CP e CT, e ao sul três sítios, PC, MU e GU. Em cada sítio há quatro amostragens por ano, trimestre 1, 2, 3 e 4.</p> <p>Resumo:</p> <table style="margin-left: 20px;"> <tr><td>S:</td><td>Seridó</td></tr> <tr><td>B:</td><td>Buíque</td></tr> <tr><td>A:</td><td>Açude</td></tr> <tr><td>R:</td><td>Rio</td></tr> </table> <table style="margin-left: 20px;"> <tr><td>TA:</td><td>Açude Recanto</td></tr> <tr><td>CP:</td><td>Riacho Cipó</td></tr> <tr><td>CT:</td><td>Riacho Caturé</td></tr> <tr><td>PC:</td><td>Riacho Poço da Cruz</td></tr> <tr><td>MU:</td><td>Açude Mulungu</td></tr> <tr><td>GU:</td><td>Açude Gurjão</td></tr> </table> <table style="margin-left: 20px;"> <tr><td>1:</td><td>1o. trimestre</td></tr> <tr><td>2:</td><td>2o. trimestre</td></tr> <tr><td>3:</td><td>3o. trimestre</td></tr> <tr><td>4:</td><td>4o. Trimestre</td></tr> </table> <p>Exemplo:</p> <p>A amostra S-R-CP-1 representa, a região Seridó (S), um ambiente de rio (R), o sítio riacho Cipó (CP), no primeiro trimestre do ano (1).</p>	S:	Seridó	B:	Buíque	A:	Açude	R:	Rio	TA:	Açude Recanto	CP:	Riacho Cipó	CT:	Riacho Caturé	PC:	Riacho Poço da Cruz	MU:	Açude Mulungu	GU:	Açude Gurjão	1:	1o. trimestre	2:	2o. trimestre	3:	3o. trimestre	4:	4o. Trimestre
S:	Seridó																												
B:	Buíque																												
A:	Açude																												
R:	Rio																												
TA:	Açude Recanto																												
CP:	Riacho Cipó																												
CT:	Riacho Caturé																												
PC:	Riacho Poço da Cruz																												
MU:	Açude Mulungu																												
GU:	Açude Gurjão																												
1:	1o. trimestre																												
2:	2o. trimestre																												
3:	3o. trimestre																												
4:	4o. Trimestre																												

Figura 7: Codificação para as variáveis, espécies de peixes, sítios de amostragem e período de amostragem.

Tabela 3: Nomenclatura das matrizes em AMD em relação aos atributos das colunas e suas abreviações nos códigos do R.

Nome	Atributos (colunas)	Abreviação no R
Matriz comunitaria	Os atributos são táxons ou OTU's (Unidades Taxonômicas Operacionais) (ex. espécies, gêneros, morfotipos)	m_com
Matriz ambiental	Os atributos são dados ambientais e variáveis físicas e químicas (ex. pH, condutividade, temperatura)	m_amb
Matriz de habitat	Os atributos são elementos da estrutura do habitat (ex. macróficas, algas, pedras, lama, etc)	m_hab
Matriz bruta	Os atributos ainda não receberam nenhum tipo de tratamento estatístico (valores brutos, como coletados)	m_bruta
Matriz transposta	Os atributos foram transpostos para as linhas	m_t
Matriz relativizada	Os atributos foram relativizados por um critério de tamanho ou de variação (ex. dividir os valores de cada coluna pela soma)	m_rel, m_relcol, m_rellin
Matriz transformada	Foi aplicado um operador matemático a todos os atributos (ex. raiz quadrada, log)	m_trns, m_log10, m_asrq
Matriz de distâncias	Matriz de m x m similaridades ou de distâncias (ex. Euclidiana, Manhattan, Bray-Curtis, etc)	m_dists, m_euclid, m_bray
Matriz de trabalho	Qualquer matriz que seja o foco da análise atual (ex. comunitária, relativizada, etc)	m_trab
Matriz particionada	Foram removidas linhas ou colunas (ex. linhas que são outliers e espécies zeradas)	m_part
Base de dados	Arquivo do Excel planilhado a partir de dados de campo ou de laboratório. Será manejada e particionada no R, para criar a Matriz bruta	ppbio06.xlsx, zoorebio.xlsx, bentos06.xlsx

Referências

- Silva, E. P. da, M. R. N. Duarte, e E. S. F. Medeiros. 2018. Relação peso-comprimento de duas espécies de peixes em um rio intermitente no nordeste do Brasil. *Neotropical Biology and Conservation* 13.
- Silva, M. J., B. R. S. Figueiredo, R. T. C. Ramos, e E. S. F. Medeiros. 2010. Food resources used by three species of fish in the semi-arid region of Brazil. *Neotropical Ichthyology* 8:825–833.

Apêndices

Sites para consulta

Script limpo