

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
DEPARTAMENTO DE ZOOLOGIA

BEATRIZ BAPTISTA KELLERMANN

Análise da eficiência do uso de códigos de barras moleculares na identificação de
helmintos do filo Acanthocephala

Orientadora

Prof.^a Dr.^a Cláudia Calegari-Marques

Coorientador

Me. Leonardo Tresoldi Gonçalves

Periódico de referência

Mitochondrial DNA Part A

Porto Alegre

2021

BEATRIZ BAPTISTA KELLERMANN

**Análise da eficiência do uso de códigos de barras moleculares na identificação
de
helminhos do filo Acanthocephala**

Trabalho de Conclusão de Curso apresentado à
Universidade Federal do Rio Grande do Sul, como
requisito parcial para obtenção do grau de
Bacharel em Ciências Biológicas.

Orientadora: Prof.^a Dr.^a Cláudia Calegari-Marques

Coorientador: Me. Leonardo Tresoldi Gonçalves

Porto Alegre

2021

**Análise da eficiência do uso de códigos de barras moleculares na identificação
de
helminhos do filo Acanthocephala**

Beatriz Baptista Kellermann

Prof.^a Dr.^a Maríndia Deprá

Me. Stephanie Lopes de Jesus

Me. Leonardo Tresoldi Gonçalves

Prof.^a Dr.^a Cláudia Calegari-Marques

SUMÁRIO

Agradecimentos	5
Introdução geral	6
Resumo	9
Introdução	10
Material e métodos	12
Resultados	14
Discussão	16
Referências bibliográficas	19
Apêndice 1	26
Apêndice 2	28

Agradecimentos

Agradeço primeiramente a Deus, que tem me dado forças e ânimo para seguir em frente com as pesquisas mesmo em tempos conturbados. A Ele seja toda a glória.

Aos meus orientadores, Cláudia e Leonardo, que me acompanharam de perto desde o início deste projeto e mesmo antes disso, compartilhando conhecimento e chamando minha atenção para detalhes que de outra forma talvez eu não percebesse;

Aos professores da cadeira de TCC I, especialmente à professora Karen Haag, por me ajudarem na difícil tarefa de escrever um projeto pela primeira vez e não perder a cabeça no processo. Agradeço pelos conselhos e pelo clima mais descontraído que trouxeram ao tema.

A todos os colegas do laboratório de helmintologia da Ufrgs, que fizeram parte da minha trajetória nas pesquisas com helmintos parasitos;

À professora Suzana Bencke Amato, que colocou em minhas mãos os primeiros acantocéfalos que investiguei, pela confiança e ensinamentos;

Agradeço também aos amigos que fiz nestes cinco anos e meio da graduação, cuja presença e apoio me ensinaram muitas coisas e me inspiraram de diversas maneiras;

Aos meus pais, Alex e Sheila, e ao meu irmão Thiago, que sempre estiveram ao meu lado quando precisei, fosse presencialmente ou digitalmente.

E a todos os parentes, amigos, professores e colegas que, de uma forma ou de outra, fizeram parte da minha trajetória acadêmica até aqui.

Introdução geral

Nas pesquisas científicas que envolvem seres vivos é essencial, em primeiro lugar, que tais seres possam ser corretamente classificados e identificados, de forma a evitar enganos que venham a comprometer a confiabilidade dos seus resultados (Kirton, 2005). A forma mais tradicional de proceder a identificação de espécimes macroscópicos e, em certos casos, também dos microscópicos, é a partir de seus caracteres morfológicos, tanto externos quanto internos. Porém, em muitas situações, tal abordagem consagrada não é suficiente. Casos em que isso ocorre incluem os complexos de espécies com morfologia críptica, espécimes em mau estado de conservação, larvas ou outros estágios de desenvolvimento que não apresentam todos os caracteres diagnósticos da espécie, ou espécies cuja descrição original seja atualmente insuficiente. Em tais casos, pode-se recorrer a outros métodos, como a identificação por meio de um “código de barras” molecular, ou *DNA barcode* (Hebert, Cywinska, et al., 2003), que servem como ferramenta alternativa ou, mais acertadamente, auxiliar (Schlick-Steiner et al., 2010).

O conceito de código de barras genético (*DNA barcode*) foi sugerido formalmente pela primeira vez em trabalhos publicados em 2003 (Hebert, Cywinska, et al., 2003; Hebert, Ratnasingham, et al., 2003). A técnica de *DNA barcoding* baseia-se na utilização de genes ou outras sequências nucleotídicas que apresentam grande variação entre espécies, porém pouca variação dentro de uma mesma espécie. Tais marcadores serviriam como um tipo de código de barras que pode ser comparado para determinar, através do grau de semelhança com sequências de referência, se um espécime pertence a um dado táxon. O marcador de *DNA barcoding* mais utilizado em estudos com espécies animais corresponde a um fragmento de cerca de 650 pares de bases, situado na região 5' do gene mitocondrial citocromo c oxidase subunidade I (*cox1*) (Hebert, Cywinska, et al., 2003).

Em 2004 foi lançado o CBOL (*Consortium for the Barcode of Life*), iniciativa internacional com o objetivo de tornar o *DNA barcoding* um padrão global para a identificação de espécies, que em 2007 já incluía mais de 120 organizações envolvidas com

as pesquisas relacionadas ao DNA *barcode* (Ratnasingham & Hebert, 2007). No ano seguinte (2005), começa a ser desenvolvido o BOLD (*Barcode of Life Data System*, <http://www.boldsystems.org/>), repositório digital para sequências de *barcode* (Hajibabaei et al., 2005) que atualmente inclui mais de 9 milhões de sequências. Um repositório deste tipo é fundamental para a disseminação da técnica de *DNA barcoding*, visto que ela depende de uma biblioteca de sequências previamente identificadas que sirvam de referência. Infelizmente, apesar dos esforços recentes de vários pesquisadores para preencher tais lacunas no conhecimento, muitos táxons ainda carecem de sequências disponíveis nas bibliotecas que as organizam (Alcántar-Escalera et al., 2013), especialmente certos grupos que não apresentam grande importância econômica ou cujos indivíduos são costumeiramente fixados com substâncias que afetam o seu material genético (Campos & Gilbert, 2012).

Os acantocéfalos são um grupo de helmintos parasitos cujo estágio adulto habita o intestino de seu hospedeiro. Apresentam ciclo de vida heteroxeno, ou seja, que utiliza mais de um hospedeiro para completar seu desenvolvimento. Em geral, o hospedeiro intermediário (HI), que ingere os ovos, é um artrópode. Nele os acantocéfalos passam pelas fases de acântor, acantela e, finalmente, cistacanto, que é o estágio infectante. O hospedeiro definitivo (HD), onde se desenvolve o adulto, é um vertebrado (Crompton & Nickol, 1985), podendo haver um ou mais hospedeiros paratênicos, que facilitam a transmissão entre HIs e HDs (Santos & Amato, 2010). A existência de estágios larvais com sistema reprodutivo ausente e relativamente poucas características diferenciáveis, além da existência de espécies morfologicamente muito semelhantes, acarretam dificuldades consideráveis para a correta identificação dos animais deste grupo (Song & Proctor, 2020). A revisão geral mais recente do filo *Acanthocephala* Rudolphi, 1802, relaciona 26 famílias, englobando 157 gêneros com 1298 espécies válidas (Amin, 2013), tornando o grupo um campo vasto para a pesquisa.

Na década de 90, vários estudos propuseram uma realocação do clado *Acanthocephala* dentro do filo *Rotifera* Cuvier, 1817, com base em evidências morfológicas

e moleculares (Garey et al., 1996). A discussão, continuada por vários estudos nos anos seguintes (Fontaneto & Jondelius, 2011; Gazi et al., 2012; Sielaff et al., 2016) tem levado à adoção do termo Syndermata para referir-se ao clado composto pelos dois grupos, muito embora Acanthocephala continue sendo comumente tratado como um filo.

Grande parte dos estudos disponíveis sobre *barcodes* de acantocéfalos foram publicados a partir de 2019 (Lewisch et al., 2020; Reier et al., 2019). Antes disso, é mais comum encontrar trabalhos que utilizam a região ribossomal 18S como marcador para a identificação molecular dos acantocéfalos (Amin et al., 2015; Near, 2002), região esta que evolui em taxas menos aceleradas que as do gene *cox1* (Hebert, Cywinska, et al., 2003). Há, porém, alguns artigos mais antigos em que o *cox1* é o gene escolhido, como um estudo de 2007 que se presta a avaliar a existência de espécies crípticas de *Leptorhynchoides thecatus* (Linton, 1891) (Steinauer et al., 2007) e um de 2013, que utilizava códigos de barras genéticos para relacionar cistacantos a espécimes adultos de *Polymorphus brevis* (Van Cleave, 1916) (Alcántar-Escalera et al., 2013). Os trabalhos mais recentes com biologia molecular de acantocéfalos incluem principalmente sequenciamento e análise de genomas mitocondriais completos de algumas espécies de *Centrorhynchus* Lühe, 1911 (Muhammad et al., 2019, 2020) e análises de *barcode* de espécies parasitas de peixes, especialmente do gênero *Pomphorhynchus* Monticelli, 1905 (Lewisch et al., 2020; Reier et al., 2019).

O desenvolvimento e aperfeiçoamento das técnicas de *barcoding* e dos bancos de dados de sequências de acantocéfalos pode representar um avanço desejável para os estudos taxonômicos do grupo e para o campo da taxonomia integrativa (utilização de diferentes métodos taxonômicos em conjunto; Schlick-Steiner et al., 2010) como um todo. Este trabalho busca analisar a eficácia do uso do *cox1* como marcador de *barcoding* no filo Acanthocephala, e discorrer brevemente sobre a situação taxonômica de alguns gêneros do grupo, e foi escrito na forma de artigo, com formatação e estruturação baseadas no periódico *Mitochondrial DNA Part A*.

Resumo: O filo Acanthocephala Rudolphi, 1802 compreende mais de 1300 espécies de endoparasitos. Uma busca geral por sequências do gene *cox1* de acantocéfalos foi realizada no *Barcode of Life Data System*, repositório especializado em sequências de *barcode*. Dados referentes a 13 gêneros, contendo 998 sequências do gene *cox1*, foram analisados a fim de testar a eficiência deste marcador na identificação do filo. As análises realizadas incluíram a visualização das distâncias intra e interespecíficas dos gêneros e a comparação de duas métricas para avaliação da utilização do marcador. Os resultados indicam que o marcador escolhido tem grande potencial no auxílio de pesquisas taxonômicas com acantocéfalos. O gênero *Pomphorynchus* Monticelli, 1905 apresentou um desempenho mais baixo nas análises, e uma busca na literatura indicou dificuldades de consenso taxonômico relativas a algumas de suas espécies. Outros casos taxonômicos relacionados aos resultados dos testes e lacunas no conhecimento dos táxons envolvidos são brevemente discutidos.

Palavras-chave: acantocéfalos, *cox1*, *DNA barcoding*, parasitos, taxonomia integrativa

Introdução

Qualquer estudo científico envolvendo seres vivos necessita, em primeiro lugar, que tais seres sejam corretamente identificados e classificados, de forma a evitar enganos que poderiam comprometer ou confundir os resultados (Kirton, 2005). A forma mais comum de proceder a identificação de espécimes de seres vivos em geral é a partir de seus caracteres morfológicos, tanto externos quanto internos. Porém, em casos como: complexos de espécies com morfologia críptica, espécimes em mau estado de conservação, larvas ou outros estágios de desenvolvimento que não apresentam todos os caracteres diagnósticos da espécie, ou espécies cuja descrição original esteja defasada, tal abordagem consagrada muitas vezes não é suficiente. Em tais situações, a identificação por meio de um “código de barras” molecular, ou *DNA barcode* (Hebert, Cywinska, et al., 2003), pode servir como uma ferramenta auxiliar, pois permite relacionar os seres vivos através do seu material genético.

A técnica de *DNA barcoding* baseia-se na utilização de genes ou outras sequências nucleotídicas que apresentam grande variação entre espécies, porém pouca variação dentro de uma mesma espécie. Estas sequências são utilizadas como um tipo de código de barras que pode ser comparado para determinar, através do grau de semelhança com sequências de referência, se um espécime pertence a um dado táxon. Em animais, o marcador mais utilizado em estudos de *DNA barcoding* corresponde a um fragmento de cerca de 650 pares de bases, situado na região 5' do gene mitocondrial citocromo c subunidade I (*cox1*) (Hebert, Cywinska, et al., 2003).

O conceito de código de barras genético foi sugerido formalmente pela primeira vez em trabalhos de 2003 (Hebert, Cywinska, et al., 2003; Hebert, Ratnasingham, et al., 2003). Em 2004 foi lançado o CBOL (*Consortium for the Barcode of Life*), que em 2007 já incluía mais de 120 organizações envolvidas com as pesquisas relacionadas ao *DNA barcode* (Ratnasingham & Hebert, 2007). Trata-se de uma iniciativa internacional com o objetivo de tornar o *DNA barcoding* um padrão global para a identificação de espécies. No ano seguinte, começa a ser desenvolvido o BOLD (*Barcode of Life Data System*,

<http://www.boldsystems.org/>), repositório digital para sequências de *barcode* (Hajibabaei et al., 2005) que atualmente inclui mais de 9 milhões de sequências. Um repositório deste tipo é fundamental para a disseminação da técnica de *DNA barcoding*, visto que ela depende de uma biblioteca de sequências previamente identificadas que sirvam de referência. Infelizmente, apesar dos esforços recentes de vários pesquisadores para preencher tais lacunas no conhecimento, muitos táxons ainda carecem de sequências disponíveis nas bibliotecas que as organizam (Alcántar-Escalera et al., 2013), especialmente certos grupos que não apresentam grande importância econômica.

Os acantocéfalos são helmintos endoparasitos cujo estágio adulto habita o intestino de seu hospedeiro, e têm um ciclo de vida heteroxeno, ou seja, que utiliza mais de um hospedeiro para completar seu desenvolvimento. Em geral, o hospedeiro intermediário (HI), que ingere os ovos, é um artrópode, e o hospedeiro definitivo (HD), onde se desenvolve o adulto, é um vertebrado, (Crompton & Nickol, 1985) podendo haver um ou mais hospedeiros paratênicos, que facilitam a transmissão entre HIs e HDs (Santos & Amato, 2010). A existência de estágios larvais com sistema reprodutivo ausente e relativamente poucas características diferenciáveis, além da existência de espécies morfologicamente muito semelhantes, acarretam dificuldades consideráveis para a correta identificação dos animais deste grupo (Song & Proctor, 2020). As estimativas mais recentes publicadas relacionam para o filo Acanthocephala 26 famílias, englobando 157 gêneros com 1298 espécies válidas (Amin, 2013), tornando o grupo um campo vasto para a pesquisa.

Entre os trabalhos publicados sobre acantocéfalos, é comum observar a utilização da região ribossomal 18S como marcador para a sua identificação molecular (Amin et al., 2015; Near, 2002), região esta que evolui em taxas menos aceleradas que as do gene *cox1* do *barcoding* (Hebert, Cywinska, et al., 2003). Estudos disponíveis em que o *cox1* é o gene escolhido incluem um de 2007, que busca avaliar a existência de espécies crípticas de *Leptorhynchoides thecatus* (Linton, 1891) (Steinauer et al., 2007) e um de 2013, que utiliza o *DNA barcoding* para relacionar larvas encistadas (cistacantos) a espécimes adultos de *Polymorphus brevis* (Van Cleave, 1916) (Alcántar-Escalera et al., 2013). Grande parte dos

trabalhos disponíveis sobre acantocéfalos que utilizam o *cox1* foram publicados a partir de 2019 (Lewisch et al., 2020; Reier et al., 2019). Os trabalhos mais recentes com biologia molecular de acantocéfalos incluem sequenciamento e análise de genomas mitocondriais completos de algumas espécies de *Centrorhynchus* Lühe, 1911 (Muhammad et al., 2019, 2020) e análises de *barcode* de espécies parasitas de peixes, especialmente do gênero *Pomphorhynchus* (Lewisch et al., 2020; Reier et al., 2019).

O desenvolvimento e aperfeiçoamento das técnicas de *barcoding* e dos bancos de dados de sequências de acantocéfalos pode representar um avanço desejável para os estudos taxonômicos do grupo. Neste trabalho, busca-se analisar a eficácia do uso do *cox1* como marcador de *barcoding* no filo Acanthocephala, e discorrer brevemente sobre a situação taxonômica de alguns gêneros do grupo.

Material e métodos

As sequências de genes *cox1* relativas a acantocéfalos foram obtidas em 06/10/2020, através de uma busca no BOLD (*Barcode of Life Data System*, <http://www.boldsystems.org/>), um repositório de sequências de *barcode*. Para evitar confusão com sequências pertencentes a um gênero de insetos com mesmo nome, foram usados os termos "Acanthocephala -Insecta". As sequências obtidas foram visualizadas através do programa AliView (Larsson, 2014) e analisadas seguindo o fluxo de trabalho descrito por Bianchi & Gonçalves (2021). Resumidamente, foram selecionadas apenas as sequências anotadas até nível de espécie, que correspondiam à região canônica de *barcode* (região COI-5P) e que, após um alinhamento e corte preliminar, possuíam 400 ou mais bases (fig. 1). Após isso, ainda foram excluídas da análise 5 sequências com inserções de bases que foram consideradas possíveis artefatos do sequenciamento. O site MAFFT (Kato et al., 2019) foi utilizado para proceder os alinhamentos das sequências, com configuração de parâmetros padrão.

O pacote em R Spider (Brown et al., 2012) foi utilizado para calcular as distâncias-p intra e interespecíficas e para gerar diagramas de caixas, de forma a facilitar a visualização das distâncias encontradas (Badotti et al., 2017; Bianchi & Gonçalves, 2021). Tais diagramas foram classificados em três categorias de desempenho: “bom”, “mediano” e “fraco”, sendo considerado “bom” o diagrama onde não há sobreposição entre as distâncias intra e interespecífica, “mediano” quando apenas os limites extremos dos diagramas se sobrepõe e “fraco” aquele em que as caixas se sobrepõe (Badotti et al., 2017). Também foi calculada a probabilidade de correta identificação ou PCI (Hollingsworth et al., 2009; Schoch et al., 2012) usando funções do mesmo pacote (Apêndice 1) e os resultados foram comparados com os de um script adaptado (Apêndice 2) do pacote *BarcodingR* (Zhang et al., 2017), a fim de testar a efetividade do marcador na identificação das sequências analisadas. O cálculo de PCI leva em consideração a existência de um intervalo entre as distâncias intra e interespecíficas em cada espécie, retornando uma porcentagem de “sucessos” para cada gênero analisado (Suwannasai et al., 2013). O pacote *BarcodingR* foi utilizado com um script modificado a partir da função *barcodes.eval*, originalmente pensada para comparar a efetividade de dois marcadores. Esta se baseia na comparação de cada sequência com todas as demais do banco de dados analisado, retornando a porcentagem de sequências cujo par mais semelhante esteja anotado para a mesma espécie.

O único gênero que apresentou desempenho abaixo de “bom” na classificação dos diagramas de caixas obtidos foi selecionado para uma busca de literatura adicional. Artigos relativos ao táxon foram buscados no Google Acadêmico, utilizando-se como palavra-chave o nome do gênero em questão. As publicações consideradas mais relevantes foram selecionadas para comparar e analisar o histórico taxonômico em busca de possíveis dificuldades na identificação morfológica das espécies envolvidas. Tal análise da literatura buscou identificar problemas como descrições muito antigas e não revisadas, estudos com identificação incompleta de espécimes, falta de consenso entre diferentes pesquisadores, espécies com apenas um sexo descrito na literatura ou espécies diferenciadas apenas por ligeiras variações morfológicas. Por fim, material previamente estudado, relativo ao gênero

de acantocéfalos *Centrorhynchus*, foi brevemente revisitado e comentado a título de comparação.

Resultados

O número inicial de sequências obtidas foi de 1792. Após a primeira seleção, retirando as sequências sem anotação de espécie ou que não correspondiam à região desejada, restaram 1240 sequências. Depois de um alinhamento preliminar, as sequências foram comparadas com uma sequência de referência oriunda do trabalho de Alcántar-Escalera et al. (2013) (número de acesso do GenBank KC549447), e cortadas para corresponderem apenas à região tradicional das análises de *barcode*. Em seguida ao corte, foram eliminadas as sequências com menos de 400 bases e algumas que continham inserções de bases possivelmente resultantes de falha no sequenciamento, restando ao fim do processo 1224 sequências.

No conjunto final, constavam sequências anotadas para 74 espécies, pertencentes a 37 gêneros. Entre as espécies contempladas, há representantes de todas as 4 classes de Acanthocephala admitidas por Amin (2013). Uma grande parte das espécies constitui-se de espécies-tipo dos respectivos gêneros. Porém, algumas análises apenas são possíveis com a presença de no mínimo duas espécies do mesmo gênero, em que ao menos uma delas apresente duas ou mais sequências disponíveis, de forma que seja possível fazer comparações inter e intra-específicas. Apenas 13 dos 37 gêneros cumpriram tais requisitos.

Conforme a classificação utilizada por Badotti et al. (2017), os diagramas correspondentes a 12 gêneros apresentaram desempenho “bom” e apenas 1 apresentou desempenho “mediano”, desconsiderando os *outliers* (pontos discrepantes). O gênero *Pomphorhynchus*, com 375 sequências disponíveis depois dos processos de filtragem, as quais pertenciam a 4 espécies, foi o único cujo diagrama teve desempenho “mediano”, apresentando também um grande número de *outliers* (Figura 1).

Os testes de PCI e de *barcodes.eval* adaptado geraram resultados diferentes para 11 dos 13 gêneros analisados. A maioria dos gêneros exibiu valores de 100% no cálculo de PCI, e os restantes apresentaram valores de PCI mais baixos que os do *barcodes.eval* (Tabela 1).

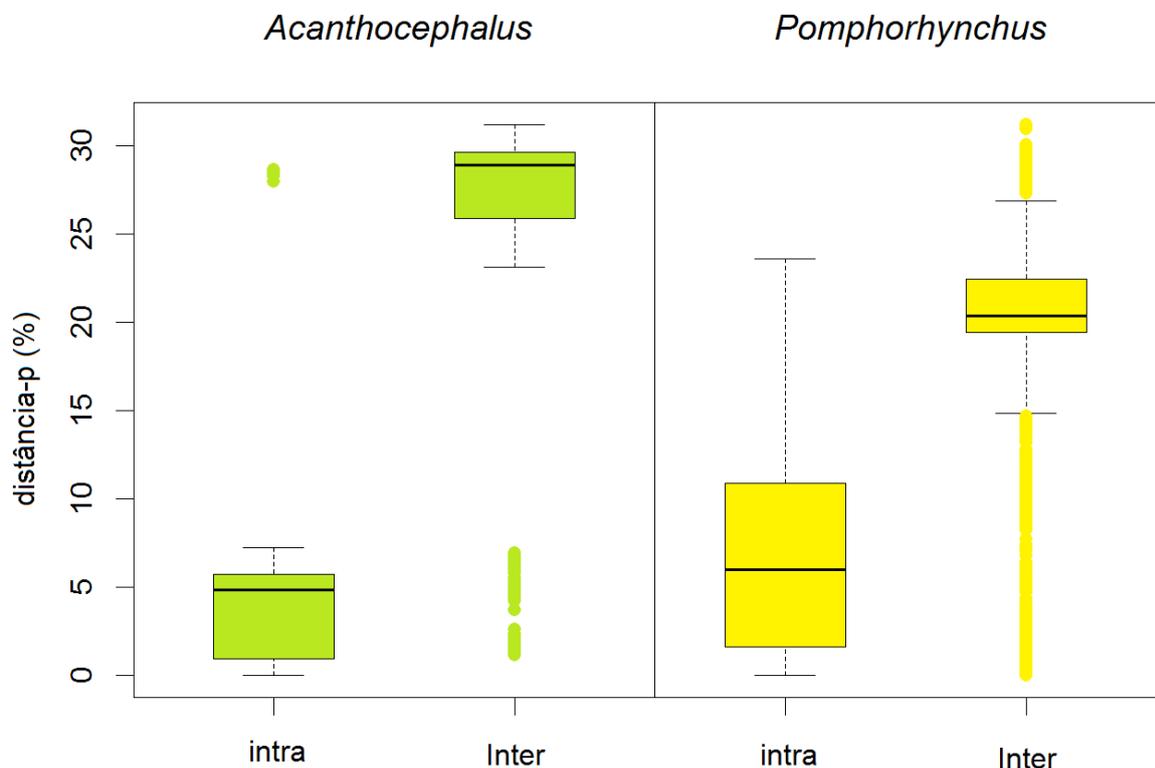


Figura 1: Exemplos de diagramas de caixas gerados e suas classificações. O gênero *Acanthocephalus* obteve desempenho “bom” por exibir um intervalo claro entre o limite superior das distâncias intraespecíficas (intra) e o limite inferior das distâncias interespecíficas (inter). *Pomphorhynchus* obteve desempenho “mediano” pois, apesar de exibir as caixas separadas, apresentou uma superposição dos limites extremos dos diagramas. Nenhum gênero analisado obteve desempenho “fraco”.

Tabela 1: resultados dos testes rodados em R: Probabilidade de Identificação Correta (PCI) e o sucesso calculado por meio da função *barcodes.eval*. Valores em porcentagem (%).

Gênero	PCI	Sucesso no <i>barcodes.eval</i>
<i>Acanthocephalus</i>	33,33	93,67
<i>Centrorhynchus</i>	100,00	66,67
<i>Corynosoma</i>	80,00	94,96
<i>Dollfusentis</i>	100,00	100,00
<i>Echinorhynchus</i>	100,00	77,27
<i>Macracanthorhynchus</i>	100,00	88,89
<i>Neoechinorhynchus</i>	100,00	55,56
<i>Polymorphus</i>	100,00	95,45
<i>Pomphorhynchus</i>	25,00	98,67
<i>Profilicollis</i>	60,00	97,98
<i>Pseudoacanthocephalus</i>	100,00	92,31
<i>Pseudocorynosoma</i>	100,00	100,00
<i>Sphaerirostris</i>	100,00	83,33

Discussão

O filo Acanthocephala compreendia, em listagem de 2013, 1296 espécies válidas (Amin, 2013). Muitas outras foram descritas após essa última revisão disponível sobre o filo como um todo, como demonstra o site ITIS (*Integrated Taxonomic Information System*, <https://www.itis.gov/>) por exemplo, onde é indicada a existência de 1330 espécies descritas (consultado em 16/04/2021). O ITIS apresenta uma base de dados taxonômicos ampla e bem suportada, e foi utilizado neste estudo para consultas relativas a números de espécies conhecidas e outras informações semelhantes. Apesar da amostra aqui utilizada cobrir menos de 6% das espécies de acantocéfalos conhecidos, os resultados dos testes de distância-p indicam que o método pode ser eficiente na distinção de espécies do filo na maioria dos gêneros analisados. O fato de ocorrer sobreposição das distâncias inter e intraespecíficas unicamente no gênero *Pomphorhynchus* pode indicar falhas no sequenciamento, falhas na identificação dos indivíduos sequenciados e/ou a existência de espécies crípticas ainda não descritas. Uma busca rápida no Google Acadêmico, em 16/03/2021, pelo nome *Pomphorhynchus* retornou 3.320 resultados. Vários artigos indicam

problemas na distinção entre as espécies do gênero, em especial *Pomphorhynchus laevis* (Zoega in Müller, 1776) e *Pomphorhynchus tereticollis* (Rudolphi, 1809) (A. F. Brown, 1987; Reier et al., 2019), o que pode estar se refletindo na anotação das sequências disponíveis no BOLD e, conseqüentemente, influenciando em nossos resultados.

Quanto aos testes de PCI e *barcodes.eval*, cada um baseia-se em cálculos distintos, portanto era esperado que os resultados variassem para cada gênero. Como o PCI leva em conta a máxima distância intraespecífica e a mínima distância interespecífica para cada gênero, pode-se dizer que este teste apresenta uma visão mais geral da efetividade do *barcoding* para os gêneros. Por outro lado, a função *barcodes.eval* baseia-se na comparação de pares individuais de sequências, e o cálculo de sucesso tem sensibilidade menor a sequências discrepantes. A comparação dos dois grupos de resultados gerados pode ajudar a inferir o tipo de problema taxonômico que pode estar envolvido na porcentagem de “insucessos” de cada gênero: um PCI mais baixo do que o valor retornado pelo *barcodes.eval* pode indicar a existência de problemas taxonômicos em várias espécies ou mesmo em todo o gênero, enquanto a situação oposta pode ser devida a erros de anotação ou outros problemas em grande quantidade, porém em poucas espécies, por exemplo. Os gêneros *Dollfusentis* Golvan, 1969 e *Pseudocorynosoma* Aznar, Pérez-Ponce de León and Raga, 2006 foram os únicos cuja taxa de “sucesso”, em ambos os testes, foi 100%. Os dois gêneros dispunham de sequências anotadas para apenas duas espécies cada, portanto se o conjunto de dados analisado fosse mais amplo, é possível que surgisse alguma variação.

Havia uma sequência, no conjunto analisado, anotada como *Profilicollis bullocki* Mateo, Cordova & Guzman, 1982 (número de acesso do BOLD GBSP8561-13), espécie não aceita por Amin (2013) não sendo nem mesmo citada em sua revisão. *Profilicollis bullocki* também não é listada no site ITIS (consultado em 23/04/2021), sendo indicada como possível sinônimo júnior de *Profilicollis altmani* (Perry, 1942) em alguns estudos (Balboa et al., 2009). Uma análise visual dessa sequência única mostrou uma enorme semelhança com

as sequências disponíveis anotadas como *Profilicollis chasmagnathi* (Holcman-Spector, Mane-Garzon and Dei-Cas, 1978) e *Profilicollis antarcticus* Zdzitowiecki, 1985.

Outra espécie de taxonomia complicada é *Pseudoacanthocephalus lucidus*. Tal nomenclatura foi sugerida num artigo recente (Nakao, 2016) para a espécie já descrita *Acanthocephalus lucidus* Van Cleave, 1925, sobre a qual não parece haver um consenso na literatura. Amin (2013) a considera, em sua revisão, como sinônimo júnior de *Acanthocephalus artatus* Van Cleave, 1925 (escrito "atratus"), porém outros autores consideram *A. lucidus* como nome válido e *A. artatus* como sinônimo júnior (Nagasawa, 2014). Em nossas análises, as sequências foram tratadas como pertencentes ao gênero *Pseudoacanthocephalus*, mantendo a anotação indicada no BOLD.

Os três gêneros mais diversos do filo Acanthocephala, segundo informações do site ITIS, são: *Neoechinorhynchus* Stiles and Hassall, 1905, com 116 espécies descritas atualmente, *Echinorhynchus* Zoega in Müller, 1776, com 94 espécies, e *Centrorhynchus*, com 93. Porém a representatividade de tais gêneros no banco de dados do BOLD não é proporcional à sua diversidade, pois ao fim da seleção das sequências restaram apenas 22, nove e três sequências de *cox1* disponíveis para cada um deles respectivamente, enquanto outros gêneros passaram para a etapa de análises no R com mais de 100 sequências. A grande maioria das sequências disponíveis pertencia aos gêneros *Pomphorhynchus* e *Profilicollis*, cujas espécies têm ciclos de vida aquáticos ou semi-aquáticos e infectam peixes, seja como HDs ou como hospedeiro paratênicos (García-Varela et al., 2013; Lewisch et al., 2020; Reier et al., 2020). Portanto, são espécies com potencial importância econômica, na medida em que podem afetar as populações de peixes que lhes servem de hospedeiros.

Gêneros de acantocéfalos muito diversos tendem a estar presentes na helmintofauna de um grande número de espécies de animais silvestres, o que somado ao fato de terem ciclos de vida complexos, faz deles grandes fontes de conhecimento ecológico (Calegari-Marques & Amato, 2010). Tomando-se como exemplo o gênero *Centrorhynchus*, o terceiro gênero mais diverso do filo, os acantocéfalos das suas várias espécies têm como

hospedeiros definitivos aves ou mamíferos, artrópodes como intermediários e anfíbios, répteis ou mamíferos insetívoros como hospedeiros paratênicos (Lunaschi & Drago, 2010; Petrochenko, 1971). Porém, são espécies geralmente consideradas de pouca ou nenhuma importância econômica, o que pode ser um dos motivos da grande lacuna de sequências de *barcode* disponíveis. É um gênero bastante antigo, com grande quantidade de espécies muito semelhantes entre si e cuja coleta muitas vezes provém de material não fresco (Richardson & Nickol, 1995), fatores que dificultam o trabalho de identificação dos novos achados, especialmente no caso de larvas, as quais comumente terminam não sendo identificadas a nível de espécie (Santos & Amato, 2010; Torres & Puga, 1996). As três sequências de *cox1* anotadas para *Centrorhynchus* disponíveis no BOLD certamente não são suficientes para tirar conclusões acerca do potencial do marcador na taxonomia do gênero mas, levando em conta o padrão geral dos resultados deste trabalho, estudos futuros focados na taxonomia integrativa destas espécies por meio de *barcoding* poderão se mostrar de grande valor.

Os achados sugerem que, havendo um banco de sequências abundantes e bem identificadas, o *barcode* com *cox1* pode ser uma ferramenta muito útil e confiável no auxílio das pesquisas taxonômicas com acantocéfalos. Em vista dos esforços recentes para ampliar os bancos de dados disponíveis, a sua utilização pode tornar-se cada vez mais frequente e cada vez mais possível.

Referências

- Alcántar-Escalera, F. J., García-Varela, M., Vázquez-Domínguez, E., & León, G. P.-P. de. (2013). Using DNA barcoding to link cystacanths and adults of the acanthocephalan *Polymorphus brevis* in central Mexico. *Molecular Ecology Resources*, 13(6), 1116–1124. <https://doi.org/10.1111/1755-0998.12090>
- Amin, O. M. (2013). Classification of the Acanthocephala. *Folia Parasitologica*, 60(4), 273–305. <https://doi.org/10.14411/fp.2013.031>

- Amin, O. M., Heckmann, R. A., Wilson, E., Keele, B., & Khan, A. (2015). The description of *Centrorhynchus globirostris* n. sp. (Acanthocephala: Centrorhynchidae) from the pheasant crow, *Centropus sinensis* (Stephens) in Pakistan, with gene sequence analysis and emendation of the family diagnosis. *Parasitology Research*, *114*(6), 2291–2299. <https://doi.org/10.1007/s00436-015-4424-0>
- Badotti, F., Silva de Oliveira, F., Fernando, C., Vaz, A., Fonseca, P. L., Fonseca, C., Nahum, L., Oliveira, G., & Góes-Neto, A. (2017). Effectiveness of ITS and sub-regions as DNA barcode markers for the identification of Basidiomycota (Fungi). *BMC Microbiology*, *17*, 42. <https://doi.org/10.1186/s12866-017-0958-x>
- Balboa, L., Hinojosa, A., Riquelme, C., Rodríguez, S., Bustos, J., & George-Nascimento, M. (2009). Alloxic distribution of cystacanths of two *Profilicollis* species in sympatric crustacean hosts in Chile. *The Journal of Parasitology*, *95*(5), 1205–1208. <https://doi.org/10.1645/GE-1965.1>
- Bianchi, F. M., & Gonçalves, L. T. (2021). Borrowing the Pentatomomorpha tome from the DNA Barcode library: Scanning the overall performance of *cox1* as a tool. *Journal of Zoological Systematics and Evolutionary Research*. https://doi.org/10.1111/jzs.124_76
- Brown, A. F. (1987). Anatomical variability and secondary sexual characteristics in *Pomphorhynchus laevis* (Müller, 1776) (Acanthocephala). *Systematic Parasitology*, *9*(3), 213–219. <https://doi.org/10.1007/BF00010856>
- Brown, S. D. J., Collins, R. A., Boyer, S., Lefort, M.-C., Malumbres-Olarte, J., Vink, C. J., & Cruickshank, R. H. (2012). Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, *12*(3), 562–565. <https://doi.org/10.1111/j.1755-0998.2011.03108.x>
- Calegario-Marques, C., & Amato, S. B. (2010). Parasites as secret files of the trophic interactions of hosts: The case of the rufous-bellied thrush. *Revista Mexicana de Biodiversidad*, *81*(3), 801–811.

- Campos, P., & Gilbert, T. (2012). DNA Extraction from Formalin-Fixed Material. *Methods in molecular biology (Clifton, N.J.)*, 840, 81–85. https://doi.org/10.1007/978-1-61779-516-9_11
- Crompton, D. W. T., & Nickol, B. B. (1985). *Biology of the Acanthocephala*. Cambridge University Press.
- Fontaneto, D., & Jondelius, U. (2011). Broad taxonomic sampling of mitochondrial cytochrome c oxidase subunit I does not solve the relationships between Rotifera and Acanthocephala. *Zoologischer Anzeiger - A Journal of Comparative Zoology*, 250(1), 80–85. <https://doi.org/10.1016/j.jcz.2010.11.005>
- García-Varela, M., Pérez-Ponce de León, G., Aznar, F. J., & Nadler, S. A. (2013). Phylogenetic relationship among genera of Polymorphidae (Acanthocephala), inferred from nuclear and mitochondrial gene sequences. *Molecular Phylogenetics and Evolution*, 68(2), 176–184. <https://doi.org/10.1016/j.ympev.2013.03.029>
- Garey, J. R., Near, T. J., Nonnemacher, M. R., & Nadler, S. A. (1996). Molecular evidence for Acanthocephala as a subtaxon of Rotifera. *Journal of Molecular Evolution*, 43(3), 287–292. <https://doi.org/10.1007/BF02338837>
- Gazi, M., Sultana, T., Min, G.-S., Park, Y. C., García-Varela, M., Nadler, S. A., & Park, J.-K. (2012). The complete mitochondrial genome sequence of *Oncicola luehei* (Acanthocephala: Archiacanthocephala) and its phylogenetic position within Syndermata. *Parasitology International*, 61(2), 307–316. <https://doi.org/10.1016/j.parint.2011.12.001>
- Hajibabaei, M., deWaard, J. R., Ivanova, N. V., Ratnasingham, S., Dooh, R. T., Kirk, S. L., Mackie, P. M., & Hebert, P. D. N. (2005). Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1462), 1959–1967. <https://doi.org/10.1098/rstb.2005.1727>
- Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1512), 313–321. <https://doi.org/10.1098/rspb.2002.2218>

- Hebert, P. D. N., Ratnasingham, S., & de Waard, J. R. (2003). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(suppl_1), S96–S99. <https://doi.org/10.1098/rsbl.2003.0025>
- Hollingsworth, M. L., Clark, A., Forrest, L. L., Richardson, J., Pennington, R., Long, D. G., Cowan, R., Chase, M. W., Gaudeul, M., & Hollingsworth, P. M. (2009). Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants: DNA BARCODING. *Molecular ecology resources*, 9, 439–457. <https://doi.org/10.1111/j.1755-0998.2008.02439.x>
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4), 1160–1166. <https://doi.org/10.1093/bib/bbx108>
- Kirton, L. G. (2005). The importance of accurate termite taxonomy in the broader perspective of termite management. *Proceedings of 5th International Conference on Urban Pests, 10–13th July, 2005, Singapore*, 571, 1–7.
- Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics (Oxford, England)*, 30. <https://doi.org/10.1093/bioinformatics/btu531>
- Lewis, E., Solymos, V., Waldner, K., Vloedt, L. van der, Harl, J., Bakran-Lebl, K., El-Matbouli, M., & Fuehrer, H.-P. (2020). Acanthocephalan parasites collected from Austrian fishes: Molecular barcoding and pathological observations. *Diseases of Aquatic Organisms*, 139, 103–111. <https://doi.org/10.3354/dao03471>
- Lunaschi, L., & Drago, F. (2010). A new species of *Centrorhynchus* (Acanthocephala, Centrorhynchidae) endoparasite of *Guira guira* (Aves, Cuculidae) from Argentina. *Helminthologia*, 47(1), 38–47. <https://doi.org/10.2478/s11687-010-0007-x>

- Muhammad, N., Suleman, Khan, M. S., Li, L., Zhao, Q., Ullah, H., Zhu, X.-Q., & Ma, J. (2020). Characterization of the complete mitogenome of *Centrorhynchus clitorideus* (Meyer, 1931) (Palaeacanthocephala: Centrorhynchidae), the largest mitochondrial genome in Acanthocephala, and its phylogenetic implications. *Molecular and Biochemical Parasitology*, 237, 111274. <https://doi.org/10.1016/j.molbiopara.2020.111274>
- Muhammad, N., Suleman, Ma, J., Khan, M. S., Wu, S.-S., Zhu, X.-Q., & Li, L. (2019). Characterization of the complete mitochondrial genome of *Centrorhynchus milvus* Ward, 1956 (Acanthocephala: Polymorphida) and its phylogenetic implications. *Infection, Genetics and Evolution*, 75, 103946. <https://doi.org/10.1016/j.meegid.2019.103946>
- Nagasawa, K. (2014). An amphibian acanthocephalan, *Acanthocephalus lucidus* (Echinorhynchida: Echinorhynchidae), infecting a fish, *Salvelinus leucomaenis leucomaenis* (Salmoniformes: Salmonidae). *Species Diversity*, 19, 151–156. <https://doi.org/10.12782/sd.19.2.151>
- Nakao, M. (2016). *Pseudoacanthocephalus toshimai* sp. nov. (Palaeacanthocephala: Echinorhynchidae), a common acanthocephalan of anuran and urodelan amphibians in Hokkaido, Japan, with a finding of its intermediate host. *Parasitology International*, 65(4), 323–332. <https://doi.org/10.1016/j.parint.2016.03.011>
- Near, T. J. (2002). Acanthocephalan Phylogeny and the Evolution of Parasitism. *Integrative and Comparative Biology*, 42(3), 668–677. <https://doi.org/10.1093/icb/42.3.668>
- Petrochenko, V. I. (1971). *Acanthocephala of Domestic and Wild Animals (Akantotsefaly (skrebni) Domashnikhi Dikikh Zhivotnykh)*. Israel Program for Scientific Translations; [Available from the National Technical Information Service, Springfield, Va.].
- Ratnasingham, S., & Hebert, P. D. N. (2007). bold: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular Ecology Notes*, 7(3), 355–364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>

- Reier, S., Sattmann, H., Schwaha, T., Fuehrer, H.-P., & Haring, E. (2020). Unravelling the hidden biodiversity – the establishment of DNA barcodes of fish-parasitizing Acanthocephala Koehltreuther, 1771 in view of taxonomic misidentifications, intraspecific variability and possible cryptic species. *Parasitology*, 1–10.
<https://doi.org/10.1017/S0031182020001316>
- Reier, S., Sattmann, H., Schwaha, T., Harl, J., Konecny, R., & Haring, E. (2019). An integrative taxonomic approach to reveal the status of the genus Pomphorhynchus Monticelli, 1905 (Acanthocephala: Pomphorhynchidae) in Austria. *International Journal for Parasitology: Parasites and Wildlife*, 8, 145–155.
<https://doi.org/10.1016/j.ijppaw.2019.01.009>
- Richardson, D. J., & Nickol, B. B. (1995). The Genus *Centrorhynchus* (Acanthocephala) in North America with Description of *Centrorhynchus robustus* n. Sp., Redescription of *Centrorhynchus conspectus*, and a Key to Species. *The Journal of Parasitology*, 81(5), 767–772. <https://doi.org/10.2307/3283971>
- Santos, V. G. T. dos, & Amato, S. B. (2010). *Rhinella fernandezae* (Anura, Bufonidae), hospedero paraténico de *Centrorhynchus* sp. (Acanthocephala, Centrorhynchidae) en Brasil. *Revista Mexicana de Biodiversidad*, 81(001), Article 001.
<https://doi.org/10.22201/ib.20078706e.2010.001.211>
- Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E., & Crozier, R. H. (2010). Integrative Taxonomy: A Multisource Approach to Exploring Biodiversity. *Annual Review of Entomology*, 55(1), 421–438. <https://doi.org/10.1146/annurev-ento-112408-085432>
- Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W., & Consortium, F. B. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, 109(16), 6241–6246.
<https://doi.org/10.1073/pnas.1117018109>

- Sielaff, M., Schmidt, H., Struck, T. H., Rosenkranz, D., Mark Welch, D. B., Hankeln, T., & Herlyn, H. (2016). Phylogeny of Syndermata (syn. Rotifera): Mitochondrial gene order verifies epizoic Seisonidea as sister to endoparasitic Acanthocephala within monophyletic Hemirotifera. *Molecular Phylogenetics and Evolution*, *96*, 79–92. <https://doi.org/10.1016/j.ympev.2015.11.017>
- Song, Z., & Proctor, H. (2020). Application of DNA barcoding and morphometric analysis in differentiating cystacanths of Polymorphus species (Acanthocephala: Polymorphidae) from central Alberta, Canada. *Parasitology Research*, *119*(10), 3359–3368. <https://doi.org/10.1007/s00436-020-06870-6>
- Steinauer, M. L., Nickol, B. B., & Ortí, G. (2007). Cryptic speciation and patterns of phenotypic variation of a highly variable acanthocephalan parasite. *Molecular Ecology*, *16*(19). <https://doi.org/10.1111/j.1365-294x.2007.03462.x>
- Suwannasai, N., Martín, M. P., Phosri, C., Sihanonth, P., Whalley, A. J. S., & Spouge, J. L. (2013). Fungi in Thailand: A Case Study of the Efficacy of an ITS Barcode for Automatically Identifying Species within the Annulohyphoxylon and Hypoxylon Genera. *PLoS ONE*, *8*(2). <https://doi.org/10.1371/journal.pone.0054529>
- Torres, P., & Puga, S. (1996). Occurrence of cystacanths of Centrorhynchus sp. (Acanthocephala: Centrorhynchidae) in toads of the genus Eupsophus in Chile. *Memórias Do Instituto Oswaldo Cruz*, *91*(6), 717–719. <https://doi.org/10.1590/S0074-02761996000600011>
- Zhang, A., Hao, M., Yang, C., & Shi, Z. (2017). BarcodingR: An integrated R package for species identification using DNA barcodes. *Methods in Ecology and Evolution*, *8*(5), 627–634. <https://doi.org/10.1111/2041-210X.12682>

Apêndice 1

Script utilizado para calcular o PCI dos gêneros analisados (com anotações):

```

library(spider)
library(ape)

#####ATENÇÃO#####
# Ao rodar o script, certifique-se que dentro do diretório de trabalho #
# não há arquivos .txt com os nomes usados nas funções write() e      #
# write.table().                #
#####

#Escolhe o diretório de trabalho
setwd("C:/Users/Usuário/Documents/Pasta")

#Cria um objeto contendo os arquivos
files <- list.files(path="C:/Users/Usuário/Documents/Bia/Pasta",
                    pattern="*.fasta", full.names=TRUE, recursive=FALSE)

#Inicia o loop
lapply(files, function(x) {

#Lê o arquivo FASTA
data <- read.FASTA(x, type = "DNA")
names(data) <- sub(" ", "_", names(data))

#Faz a distribuição das distâncias e cria o vetor de espécies (formato BOLD)
distribution <- dist.dna(data, pairwise.deletion = TRUE, model = "raw")
species <- strsplit(names(data), split="[\\_ ]")
species <- sapply(species, function(x) paste(x[2], x[3], sep="_"))

#Armazena o nome do gênero
genus <- sapply(strsplit(species, split="_"), function(x) x[1])[1]

#Maximum intraspecific distance - Distância máxima intraespecífica
max.intra <- maxInDist(distribution, species)
table1 <- data.frame(species, max.intra)
table1 <- table1[order(table1$max.intra), ]
mi <- table1[!duplicated(table1$species), ]
mi <- mi[order(mi$species), ]

#Nearest non-conspecific - Distância ao vizinho mais próximo
near.neigh <- nonConDist(distribution, species)
table2 <- data.frame(species, near.neigh)
table2 <- table2[order(table2$near.neigh), ]

```

```
nn <- table2[!duplicated(table2$species), ]
nn <- nn[order(nn$species), ]

#Faz o dataset total das distâncias e remove comparações inválidas (NA)
total <- merge(mi, nn, by="species")
total <- total[complete.cases(total), ]

#Escreve os resultados das distâncias em um arquivo de texto
write.table(total, file="distancias.txt", sep = ";", append=TRUE,
            row.names=FALSE, col.names = !file.exists("distancias.txt"))

#Verifica se a distância ao vizinho mais próximo é maior que a máxima
#distância intraespecífica e calcula o PCI
total$count <- ifelse(total$near.neigh > total$max.intra, 1, 0)
pci <- round((sum(total$count) / length(total$count) * 100), digits = 2)

#Escreve o pci de cada gênero em um arquivo de output
write(paste(genus, pci), file="pci_generos.txt", append=TRUE)

#Fecha o loop
})

#Lê o arquivo com as distâncias para plotar
dataset <- read.table("distancias.txt", sep = ";", header = TRUE)
```

Apêndice 2

Script adaptado para avaliar um único marcador de *barcode*, a partir da função *barcodes.eval* (com anotações):

```
library(BarcodingR)
library(ape)

#Escolhe o diretório de trabalho
setwd("C:/Users/Usuário/Documents/Pasta2")

#Escreve o cabeçalho do arquivo final
write(paste("genus", "success"), file="barcode_evaluation.txt")

#Cria um objeto contendo os arquivos
files <- list.files(path="C:/Users/Usuário/Documents/Pasta2",
                    pattern="*.fasta", full.names=TRUE, recursive=FALSE)

#Inicia o loop
lapply(files, function(x) {

data <- read.FASTA(x, type = "DNA")
names(data) <- sub(" ", "_", names(data))

species <- strsplit(names(data), split="[\\|_]")
species <- sapply(species, function(x) paste(x[2], x[3], sep="_"), sep = ",")

#Armazena o nome do gênero
genus <- as.character(strsplit(names(data), split="[\\|_]")[[1]][2])

data <- setNames(data, species)

kmer1 <- optimize.kmer(data, max.kmer=10)[1]
barcode1 <- data

evaluation <- function (barcode1, kmer1 = kmer1)
{
  set.seed(7)
  NAMES <- function(seqs) {
    if (mode(seqs) == "raw") {
      SeqNames <- attr(seqs, "dimnames")[[1]]
    }
    else {
      SeqNames <- names(seqs)
    }
    if (length(SeqNames) == 0) {
```

```

    stop("the mode(seqs) is wrong!")
  }
  else {
    return(SeqNames)
  }
}
barcode1.IDs <- NAMES(barcode1)
barcode1 <- del.gaps(barcode1)
names(barcode1) <- barcode1.IDs
strings.equal <- function(str1, str2) {
  ifelse(str1 == str2, 1, 0)
}
c2s <- function(chars = c("m", "e", "r", "g", "e", "d")) {
  return(paste(chars, collapse = ""))
}
DNABin2kmerFreqMatrix <- function(ref, kmer = kmer) {
  if (class(ref) != "DNABin")
    stop("seqs should be in DNABin format!")
  if (class(kmer) != "integer")
    kmer <- as.integer(kmer)
  seqs.as.char <- as.character(ref)
  ifelse(is.vector(seqs.as.char), seqs.as.str.vector <- lapply(seqs.as.char,
    FUN = c2s), seqs.as.str.vector <-
    apply(seqs.as.char,
                                                MARGIN = 1, FUN = c2s))
  seqs.unique.as.str.vector <- unique(seqs.as.str.vector)
  s <- seqs.unique.as.str.vector
  u.s0 <- unique(substring(s, 1, kmer))
  u.s <- u.s0
  for (i in 2:(max(nchar(s)) - kmer + 1)) {
    u.s <- c(u.s, unique(substring(s, i, kmer + i -
      1)))
    u.s <- unique(u.s)
    n.char <- nchar(u.s)
    size.kmer.exact <- n.char == kmer
    u.s <- subset(u.s, size.kmer.exact)
  }
  kmer.freq.one.seq <- sapply(u.s, gregexpr, seqs.as.str.vector[1])
  kmer.freq.one.seq2 <- numeric(length(kmer.freq.one.seq))
  for (i in 1:length(kmer.freq.one.seq)) {
    ifelse(kmer.freq.one.seq[[i]] == -1, kmer.freq.one.seq2[i] <- 0,
    kmer.freq.one.seq2[i] <- length(kmer.freq.one.seq[[i]]))
  }
  kmer.freq.one.seq3 <- kmer.freq.one.seq2
  for (k in 2:length(seqs.as.str.vector)) {
    kmer.freq.one.seq <- sapply(u.s, gregexpr, seqs.as.str.vector[k])
    kmer.freq.one.seq2 <- numeric(length(kmer.freq.one.seq))
    for (i in 1:length(kmer.freq.one.seq)) {

```

```

    ifelse(kmer.freq.one.seq[[i]] == -1, kmer.freq.one.seq2[i] <- 0,
    kmer.freq.one.seq2[i] <- length(kmer.freq.one.seq[[i]]))
  }
  kmer.freq.one.seq3 <- c(kmer.freq.one.seq3, kmer.freq.one.seq2)
  }
  kmer.freq.matrix <- t(array(kmer.freq.one.seq3, dim = c(length(kmer.freq.one.seq),
  length(kmer.freq.one.seq3)%/length(kmer.freq.one.seq))))
  ifelse(length(names(ref)) != 0, rownames(kmer.freq.matrix) <- names(ref),
  rownames(kmer.freq.matrix) <- rownames(ref))
  out <- list(unique.str = u.s, kmer.Freq.ref = kmer.freq.matrix)
  return(kmer.freq.matrix)
}
success.b1 <- 0
morph.spe <- gsub(".", "", barcode1.IDs)
no.samples.b1 <- length(barcode1)
barcode0 <- barcode1
for (i in 1:no.samples.b1) {
  barcode1 <- barcode0
  morph.spe.minus1 <- morph.spe[-i]
  kmer.freq.b0 <- DNABin2kmerFreqMatrix(barcode1, kmer = kmer1)
  test <- kmer.freq.b0[i, ]
  kmer.freq.b0.minus1 <- kmer.freq.b0[-i, ]
  Spp2 <- as.factor(morph.spe.minus1)
  knn1 <- knn(kmer.freq.b0.minus1, test, cl = Spp2, k = 1,
  l = 0, prob = FALSE, use.all = TRUE)
  spe.Identified <- as.character(knn1)
  success.b1 <- strings.equal(morph.spe[i], spe.Identified) +
  success.b1
}

success <- success.b1
total <- no.samples.b1
out <- (success.b1/no.samples.b1)*100
return(out)
}

result <- evaluation(barcode1, kmer1 = kmer1)

#Escreve a porcentagem de acertos de cada gênero em um arquivo de output
write(paste(genus, round(result, digits = 2)), file="barcode_evaluation.txt", append=TRUE)

#Fecha o loop
})

```