



UNIVERSIDADE FEDERAL DE GOIÁS  
INSTITUTO DE CIÊNCIAS BIOLÓGICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA  
MOLECULAR

---



**Eliane Cotrim Batista**

**CARACTERIZAÇÃO DA PRODUÇÃO CIENTÍFICA SOBRE DNA  
*BARCODING* NO ESTUDO DE PLANTAS**

Orientadora: **Dr.<sup>a</sup> Mariana Pires de Campos Telles**

Coorientadora: **Dr.<sup>a</sup> Ariany Rosa Gonçalves**

ABRIL – 2021



UNIVERSIDADE FEDERAL DE GOIÁS  
INSTITUTO DE CIÊNCIAS BIOLÓGICAS

## TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

### E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

#### 1. Identificação do material bibliográfico

Dissertação       Tese

#### 2. Nome completo do autor

Eliane Cotrim Batista

#### 3. Título do trabalho

Caracterização da produção científica sobre DNA *barcoding* no estudo de plantas.

#### 4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento  SIM       NÃO<sup>1</sup>

**[1]** Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

- a)** consulta ao(à) autor(a) e ao(à) orientador(a);
- b)** novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

**Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.**



Documento assinado eletronicamente por **Mariana Pires De Campos Telles**,



**Professor do Magistério Superior**, em 01/05/2021, às 13:00, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).

---



Documento assinado eletronicamente por **ELIANE COTRIM BATISTA, Discente**, em 03/05/2021, às 18:23, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).

---



A autenticidade deste documento pode ser conferida no site [https://sei.ufg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **2035078** e o código CRC **1484CAA0**.

---

**Referência:** Processo nº 23070.017215/2021-74

SEI nº 2035078

**ELIANE COTRIM BATISTA**

**CARACTERIZAÇÃO DA PRODUÇÃO CIENTÍFICA SOBRE DNA  
*BARCODING* NO ESTUDO DE PLANTAS**

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Federal de Goiás, como parte dos requisitos para obtenção do título de mestre em Genética e Biologia Molecular.

Orientadora: **Dr.<sup>a</sup> Mariana Pires de Campos Telles**

Coorientadora: **Dr.<sup>a</sup> Ariany Rosa Gonçalves**

GOIÂNIA – GO

2021

Ficha de identificação da obra elaborada pelo autor, através do  
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Batista, Eliane Cotrim

Caracterização da produção científica sobre DNA barcoding no estudo  
de plantas [manuscrito] / Eliane Cotrim Batista. - 2021.  
lxi, 61 f.

Orientador: Profa. Dra. Mariana Pires de Campos Telles; co  
orientadora Dra. Ariany Rosa Gonçalves.

Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto  
de Ciências Biológicas (ICB), Programa de Pós-Graduação em Genética  
e Biologia Molecular, Goiânia, 2021.

Bibliografia.

Inclui tabelas, lista de figuras, lista de tabelas.

1. Infometria. 2. DNA barcode. 3. ITS. 4. matK, psbA-trnH. 5. rbcL.  
I. Telles, Mariana Pires de Campos , orient. II. Título.

CDU 575



UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

## ATA DE DEFESA DE DISSERTAÇÃO

Ata nº 89 da sessão de Defesa de Dissertação de **Eliane Cotrim Batista**, que confere o título de Mestre(a) em **Genética e Biologia Molecular**, na área de concentração em **Genética e Biologia Molecular**.

Ao/s **vinte e dois dias do mês de abril de dois mil e vinte e um**, a partir da(s) **09h00**, por videoconferência, seguindo portaria CAPES no. 36 de 16 de março de 2020 e recomendação da UFG realizou-se a sessão pública de Defesa de Dissertação intitulada “Caracterização da produção científica sobre DNA barcoding no estudo de plantas”. Os trabalhos foram instalados pelo(a) Orientador(a), Professor(a) Doutor(a) **Mariana Pires de Campos Telles** com a participação dos demais membros da Banca Examinadora: Professor(a) Doutor(a) **Rhewter Nunes**, membro titular externo; Professor(a) Doutor(a) **Ramila dos Santos Braga**, membro titular interno. Durante a arguição os membros da banca **não** fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido(a) o(a) candidato(a) **aprovado(a)** pelos seus membros. Proclamados os resultados pelo(a) Professor(a) Doutor(a) **Mariana Pires de Campos Telles**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, ao(s) **vinte e dois dias do mês de abril de dois mil e vinte e um**.

## TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Mariana Pires De Campos Telles, Professor do Magistério Superior**, em 22/04/2021, às 15:46, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **RHEWTER NUNES, Usuário Externo**, em 22/04/2021, às 15:48, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **RAMILLA DOS SANTOS BRAGA, Usuário Externo**, em 23/04/2021, às 18:15, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **2018449** e o código CRC **F115E6D2**.

Referência: Processo nº 23070.017215/2021-74

SEI nº 2018449

A todos aqueles que foram minha rede de apoio durante esses intensos dois anos: família, meu companheiro, amigos de sempre e para sempre, às novas amizades que chegaram com a pós-graduação, ao pequenino Samuel que ainda carrego no ventre, mas que já ocupa todos os espaços em mim.

*Dedico.*

## AGRADECIMENTOS

Inicialmente agradeço ao apoio logístico e financeiro oferecidos pelo Instituto Nacional de Ciência e Tecnologia em Ecologia, Evolução e Conservação da Biodiversidade (INCT EECBio) e o Núcleo de Excelência em Ecologia Molecular, Genética Forense e Conservação (BioEMFoCo), financiado dentro do Programa de Apoio a Núcleos de Excelência PRONEX/FAPEG/CNPq (aprovado na Chamada Pública nº 06/2016).

Aproveito também para agradecer à Universidade Federal de Goiás e ao Programa de Genética e Biologia Molecular, na figura dos professores e funcionários pelo suporte e oferta de educação pública de qualidade.

Agradeço ainda ao Governo do Estado de Goiás que, por meio da Secretaria de Estado da Educação, concedeu um ano de afastamento para meu aprimoramento. Esse tipo de incentivo é crucial para que o servidor continue se especializando e aprimorando para retornar para a rede mais capacitado, podendo oferecer um trabalho de maior qualidade, como a educação pública tanto necessita e merece.

Retornar à vida acadêmica depois de um hiato de dez anos não foi fácil e apesar da segurança de regressar a um lugar que foi meu lar ao longo de toda a graduação, o recomeço é sempre preenchido pelo inesperado. Agradeço imensamente a Professora Mariana Pires de Campos Telles e a toda a equipe do Laboratório de Genética & Biodiversidade (professores e estudantes) por terem me acolhido de forma tão fraterna e terem acreditado no meu potencial. Agradeço à Mariana também por sempre buscar desenvolver em seus estudantes habilidades que vão além do conhecimento científico específico de nossa área, promovendo não só a nossa formação profissional, mas também humana. Gratidão!

Agradeço à Professora Thannya do Nascimento Soares, pelas palavras certas em nossos raros encontros de família e que me levaram a fazer a escolha de fazer Iniciação Científica ao longo da graduação e retornar à pós-graduação, mesmo depois de muitos anos afastada da vida acadêmica. Mesmo com nosso “jeito desajeitado” de demonstrar nossos sentimentos, saiba que sou imensamente grata pelas palavras de apoio ditas sempre em momentos importantes para mim, sou grata pelo modo que sempre me acolheu. Você mora no meu coração!

Agradeço aos velhos e novos amigos de caminhada do LGBio, especialmente à Ariany, minha coorientadora, que desde o início segurou na minha mão, me amparou e me ensinou muito. Admiro muito a forma com que se compromete em absolutamente tudo que se propõe, admiro a sua empatia, a sua dedicação e cuidado ao ensinar. Obrigada também pelas longas horas de conversas sobre nossas impressões sobre a vida, com certeza você é imensamente especial para mim. Agradeço à Andrezza, Cíntia, Fernanda, Lucas Donizetti, Natácia, Ramilla e Rhewter pelos sorrisos, abraços, dedos de prosa, cafezinhos compartilhados e o acolhimento tão sincero. Ariany, Ramilla e Rhewter foram muito importantes para que eu “resetasse meu HD” e me apropriasse da nova proposta de projeto em meio ao caos psicológico e emocional de uma pandemia sem fim.

Agradeço aos meus colegas de turma pelas horas de disciplinas e ensinamentos compartilhados. Agradeço à Aliane e Camila pelas trocas no estágio de docência e especialmente à Camila pelos cafezinhos com chocolate, nossas conversas afetuosas sobre a vida e pelo companheirismo. Agradeço à Amanda, Larissa e Sara sempre tão queridas e divertidas. Foram vários perrengues, choros, frustrações, disciplinas sem fim, mas também gargalhadas, abraços, almoços, lanches, cafés, ombros amigos, colo, conversas fiadas e sérias, que levarei para sempre na minha memória. Todas vocês têm um potencial incrível e se tornarão profissionais diferenciadas em qualquer coisa que se proponham.

Por fim agradeço à minha rede de apoio familiar sempre tão presente e importante em qualquer aspecto da minha vida. Agradeço à Luciana Siqueira Silva, minha irmã da barriga de outra mãe, por ser tão presente mesmo com a distância, pelos inúmeros momentos de trocas frente aos desafios da pós-graduação (eu de cá no mestrado e ela de lá, no doutorado). Mas a minha imensa gratidão vai para o meu companheiro Danilo, sempre lado a lado comigo, me amparando mesmo com o seu silêncio, para que eu me regulasse e voltasse ao centro. Em meio ao ano turbulento de 2019 em que me desdobrei para trabalhar e estudar ou ao, sem comentários 2020, você nunca soltou a minha mão. A certeza de que estamos no caminho certo hoje, cresce dentro do meu “barrigão”! Mesmo com tantos medos e incertezas, um novo ciclo se iniciará para nós com a chegada do Samuel e eu só posso dizer que te amo.

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	16
<b>2. OBJETIVOS</b> .....	18
<b>3. REVISÃO DE LITERATURA</b> .....	19
<b>3.1 Tecnologias de Sequenciamento de DNA</b> .....	19
<b>3.2 DNA <i>barcoding</i></b> .....	21
<b>3.3 DNA <i>barcoding</i> para plantas</b> .....	24
<b>3.4 Aplicações de DNA <i>barcoding</i> no estudo de plantas</b> .....	28
<b>3.5 Ferramentas para avaliação do desenvolvimento científico</b> .....	30
<b>4. MATERIAL E MÉTODOS</b> .....	32
<b>4.1 Análise Cienciométrica</b> .....	32
<b>4.2 Categorização</b> .....	35
<b>5. RESULTADOS E DISCUSSÃO</b> .....	37
<b>5.1 Análise Cienciométrica</b> .....	37
<b>5.2 Categorização</b> .....	52
<b>6. CONCLUSÃO</b> .....	56
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	57

## LISTA DE FIGURAS

- Figura 1.** Histórico da tecnologia de sequenciamento genético, com início na década de 1970 com o método proposto por Sanger et al. (1977), passando pelo Projeto Genoma Humano e o desenvolvimento do Sequenciamento de Alto Rendimento. Adaptado de: Yang et al. (2020). ..... 20
- Figura 2.** Fluxo de trabalho para desenvolvimento e aplicações de DNA *barcodes*, partindo da coleta de amostras até a publicação e acesso aos dados. Adaptado de: Fajarningsih (2016). ..... 22
- Figura 3.** Representação do *barcoding gap* com as respectivas distribuições da variação intraespecífica, mostrada em vermelho e da intraespecífica em amarelo. **A)** Mundo ideal para DNA *barcoding*, com distribuições discretas e sem sobreposições (*overlap*). **B)** Uma versão alternativa do mundo ideal com sobreposição significativa e sem *gap*. Adaptado de: Meyer & Paulay (2005). ..... 23
- Figura 4.** Representação das etapas de levantamento e refinamento para obtenção do conjunto de dados de estudos com aplicação de DNA *barcoding* no estudo de plantas e as respectivas etapas de análise cienciométrica e de categorização. Dentro da etapa “Levantamento de dados”, as setas em vermelho representam os critérios de exclusão adotados, enquanto a seta verde representa os critérios de inclusão. .... 33
- Figura 5.** Proposta de categorização dos artigos com base nas informações sobre uso e/ou aplicação da técnica de DNA *barcoding* no estudo de plantas. .... 35
- Figura 6.** Estudos encontrados utilizando a técnica de DNA *barcoding* em plantas, considerando uma janela de tempo de 16 anos (2005-2020). **A)** Produção anual de documentos, considerando a janela de tempo de 2005-2020. **B)** Média global de citações do conjunto de documentos produzidos dentro de cada ano, considerando a janela de tempo de 2005-2020. .... 39
- Figura 7.** Gráfico de três campos apresentando a relação das 10 palavras-chave mais utilizadas pelos 10 autores que mais publicaram e os respectivos 10 periódicos de maior relevância para produções na área de DNA *barcoding* em plantas. .... 40
- Figura 8.** *Ranking* dos 10 autores mais citados globalmente, considerando as publicações sobre a técnica de DNA *barcoding* em plantas. .... 42
- Figura 9.** Distribuição da produtividade dos autores mais relevantes sobre a aplicação da técnica de DNA *barcoding* em plantas, ao longo da janela de tempo (2005-2020). O tamanho e a intensidade da cor de cada círculo representam a quantidade de publicações e número de citações por ano em determinado período, respectivamente. .... 43

<b>Figura 10.</b> Representação da Lei de Lotka envolvendo a produtividade dos autores com relação a pesquisas sobre aplicação de DNA barcoding em plantas. A linha pontilhada laranja representa o modelo esperado para a Lei de Lotka enquanto a linha azul demonstra o comportamento do conjunto de dados considerado.....	44
<b>Figura 11.</b> Frequência da produção científica mundial relacionada a aplicação da técnica de DNA <i>barcoding</i> em plantas com destaque para China (1.479 artigos publicados), Estados Unidos (392 artigos publicados) e Índia 332 (artigos publicados). .....	45
<b>Figura 12.</b> <i>Ranking</i> dos 10 trabalhos mais citados dentro do conjunto de dados sobre aplicação de DNA <i>barcoding</i> em plantas, considerando a janela de tempo de 2005-2020. ....	46
<b>Figura 13.</b> Representação esquemática da Lei de Bradford em que periódicos são agrupados por zonas, que condensam maior número de documentos especializados considerando a temática estudada.....	49
<b>Figura 14.</b> Lista de periódicos pertencentes à Zona 1 segundo a Lei de Bradford, que apresentaram a maior número de publicações sobre a temática estudada. ....	50
<b>Figura 15.</b> Rede de coocorrência das palavras-chave empregadas pelos autores em artigos relacionados com a aplicação de DNA <i>barcoding</i> em plantas. ....	51
<b>Figura 16.</b> Rede de colaboração entre os 30 principais autores de diferentes nacionalidades quanto a produção relativa à aplicação de DNA <i>barcoding</i> no estudo de plantas.....	52
<b>Figura 17.</b> Respective usos e/ou aplicações atribuídos aos 1.073 artigos encontrados <i>na Web of Science</i> relacionados a estudos com DNA <i>barcoding</i> em plantas.....	52
<b>Figura 18.</b> Magnitude da produção científica por categoria de uso e/ou aplicação da técnica de DNA <i>barcoding</i> no estudo de plantas ao longo da janela de tempo de 2005-2020. ....	54
<b>Figura 19.</b> Número de publicações de estudos em plantas, utilizando as técnicas de: <b>A)</b> DNA barcoding e <b>B)</b> metabarcoding, ao longo (2005-2020). <b>C)</b> Categorização dos estudos do banco de dados, considerando o uso e/ou aplicação de DNA <i>barcoding</i> no estudo de plantas. <b>D)</b> Categorização dos trabalhos do banco de dados, considerando o uso e/ou aplicação de <i>metabarcoding</i> no estudo de plantas.....	55

## LISTA DE TABELAS

- Tabela 1.** Características dos diferentes marcadores que foram rotineiramente incluídos em estudos de códigos de barras de plantas. Adaptado de: Hollingsworth et al. (2011).  
..... 25
- Tabela 2.** Descrição geral das principais informações encontradas na busca realizada no banco de dados da *Web of Science*. ..... 37
- Tabela 3.** *Ranking* dos 10 autores que mais publicaram sobre aplicação de DNA *barcoding* em plantas, relacionados aos respectivos: índice *h*, país de origem e instituição de filiação. .... 41
- Tabela 4.** *Ranking* dos países com maior número de publicações e que apresentaram maior relevância quanto ao número de citações sobre a aplicação de DNA *barcoding* em plantas. .... 46
- Tabela 5.** *Ranking* dos 10 periódicos de maior destaque em relação ao número de documentos publicados, número de citações e índice *h*. .... 48
- Tabela 6.** *Ranking* das 10 palavras-chave mais usadas pelos autores em publicações relacionadas a aplicação da técnica de DNA *barcoding* no estudo de plantas. .... 50

## RESUMO

A aplicação da técnica de DNA *barcoding* tem se mostrado bem-sucedida em estudos de levantamento da biodiversidade, biomonitoramento de espécies invasoras, análises forenses para avaliação de segurança alimentar e de medicamentos à base de plantas medicinais, monitoramento do comércio ilegal de espécies ameaçadas, estudos sobre ecologia alimentar, iniciativas de conservação etc. O presente estudo teve como objetivo principal levantar e sistematizar a produção científica sobre o uso de DNA *barcoding* em plantas. Para tanto, foi realizada uma busca na base de dados *Web of Science*, utilizando como palavras-chave combinadas as regiões anotadas no genoma plastidial e nuclear de *Arabidopsis thaliana* (espécie modelo), que são utilizadas como marcadores de DNA *barcoding*, associadas às expressões: “DNA *barcoding*” OR “*barcoding*” OR “DNA *barcode*” OR “*barcode*” OR “*metabarcoding*”. A busca foi realizada abrangendo o período completo de publicação da base de dados até 2020. Os arquivos recuperados passaram por filtragem, sendo considerados apenas artigos originais, excluindo-se artigos de revisão, bem como os que não correspondiam ao escopo do estudo. Um total de 1.897 artigos foram recuperados, dos quais 1.073 estavam dentro dos critérios de inclusão. Como o primeiro artigo foi publicado em 2005, a análise abrangeu o período entre 2005 e 2020. Os estudos foram publicados em um total de 309 periódicos, com uma média geral de publicação por ano igual a 5,21. A produção científica variou ao longo do período, mas sempre com uma tendência crescente. A média de autores por documento foi de 3,49 visto que, cada artigo apresenta em sua grande maioria mais de um autor, o que explica a média de coautores por documento (5,34). No entanto, de um total de 3.747 autores registrados, o número de artigos por autor foi baixo, com uma média igual a 0,29. A China se destaca como principal país, de acordo com a origem dos autores, que tem contribuído com trabalhos envolvendo a aplicação da técnica de DNA *barcoding* no estudo de plantas, sendo também o país mais citado. Os autores mais relevantes, com base no número de publicações foram os chineses: Chen SL, Song J, Yao H, sendo o artigo de Chen et al. (2010) o segundo mais citado globalmente. Os periódicos que apresentaram maior destaque em relação ao número de artigos publicados, considerando o assunto, foram respectivamente: *Plos One*, *Molecular Ecology Resources* e *Scientific Reports*. Foi possível caracterizar quatro grupos de documentos quanto ao uso/aplicação da técnica, respectivamente: Descrição/delimitação de espécies (545), Genética Forense (275), Desenvolvimento/metodologia (163), Ecologia (90) e 75 artigos foram categorizados em até dois grupos diferentes. Os estudos agrupados na categoria de maior destaque (Descrição/delimitação de espécies), abordam desde delimitação de espécies de grupos complexos, até a validação de espécies com propriedades de interesse econômico e de espécies já catalogadas em herbários. As possibilidades de aplicação vêm sendo ampliadas graças ao desenvolvimento das tecnologias de Sequenciamento de Alto Rendimento, abrindo portas para estudos em Ecologia, Evolução e Genética Forense e que utilizam além de DNA *barcoding*, o *metabarcoding*.

**Palavras-chave:** Cienciometria, DNA *barcode*, *ITS*, *matK*, *psbA-trnH*, *rbcL*.

## ABSTRACT

The application of the DNA barcoding has been shown to be successful in studies of biodiversity surveys, biomonitoring of invasive species, forensic analysis to assess food security and medicinal products based on medicinal plants, monitoring the illegal trade in threatened species, studies about food ecology, conservation initiatives, etc. The present study had as main objective to raise and systematize the scientific production on the use of DNA barcoding in plants. For this purpose, a search was carried out in the *Web of Science* database, using as keywords combined the regions annotated in the chloroplast and nuclear genome of *Arabidopsis thaliana* (model species), which are used as DNA barcoding markers, associated with the expressions: “DNA barcoding” OR “barcoding” OR “DNA barcode” OR “barcode” OR “metabarcoding”. The search was carried out covering the complete period of publication of the database until 2020. The recovered files were filtered, being considered only original articles, excluding review articles, as well as those that did not correspond to the scope of the study. A total of 1,897 articles were retrieved, of which 1,073 were within the criteria’s inclusion. As the first article was published in 2005, the analysis covered the period between 2005 and 2020. The studies were published in a total of 309 sources, with a general average of publication per year equal to 5.21. Scientific production varied over the period, but always with an increasing trend. The average number of authors per document was 3.49, since each article mostly has more than one author, which explains the average number of coauthors per document (5.34). However, out of a total of 3,747 registered authors, the number of articles per author was low, with an average equal to 0.29. China stands out as the main country, according to the origin of the authors, who has contributed with works involving the application of the DNA barcoding technique in the study of plants, being also the most cited country. The most relevant authors, based on the number of publications, were the Chinese: Chen SL, Song J, Yao H, with the article by Chen et al. (2010) the second most cited globally. The journals that showed greater prominence in relation to the number of articles published, considering the subject, were respectively: *Plos One*, *Molecular Ecology Resources* and *Scientific Reports*. It was possible to characterize four groups of documents regarding the use / application of the technique, respectively: Description/delimitation of species (545), Forensic Genetics (275), Development/methodology (163), Ecology (90) and 75 articles were categorized in up to two different groups. The studies grouped in the most prominent category (Description/delimitation of species), range from delimiting species from complex groups to the validation of species with properties of economic interest and species already cataloged in herbariums. The application possibilities have been expanded due to the development of High Throughput Sequencing technologies, opening doors for studies applied to Ecology, Evolution and Forensic Genetics and which use in addition to DNA barcoding, the metabarcoding.

**Keywords:** Scientrometrics, DNA barcode, ITS, matK, psbA-trnH, rbcL.

## 1. INTRODUÇÃO

A técnica de identificação molecular denominada DNA *barcoding* (Código de Barras de DNA) foi proposta por Hebert e colaboradores (2003), com o intuito de colaborar com a discriminação da vida, através da utilização de uma sequência padrão de DNA semelhante a um código de barras comercial (Hebert et al., 2003). Apesar de ser considerada alternativa aos métodos taxonômicos clássicos, a identificação molecular na verdade, é um complemento para cobrir lacunas observadas na identificação clássica (morfologia, acústica etc.), na identificação baseada em análises químicas e bioquímicas de moléculas, na identificação molecular cromossômica (tamanho e número de cromossomos) e etc. (Hebert et al., 2003; Hebert & Gregory, 2005).

Segundo Meyer & Paulay (2005), a delimitação de espécies por meio do DNA *barcoding* acontece pela existência de uma lacuna (*barcoding gap*), usada para verificar os limites entre variação intraespecífica (variação genética observada entre os indivíduos de uma mesma espécie) da divergência interespecífica (variação genética observada entre espécies diferentes). Desse modo quanto mais sobreposição (*overlap*) houver entre a variação genética intraespecífica e a divergência interespecífica, menos eficaz o código de barras se torna. Em síntese, a região de DNA escolhida deve conter variabilidade e divergências genéticas significativas em nível de espécie e ser amplamente distribuída nos *taxa*, ser passível de sequenciamento com iniciadores universais e ser pequena o suficiente para ser sequenciada com facilidade (Hebert et al., 2003).

Inicialmente, a técnica de DNA *barcoding* foi aplicada em animais, utilizando parte da sequência do gene da citocromo oxidase I (*COI*), devido às características intrínsecas ao genoma mitocondrial. No entanto, a aplicação desta técnica no estudo das plantas não apresentou o mesmo sucesso, visto que o genoma mitocondrial nesses organismos apresenta uma estrutura instável, com divergências significativas nas regiões codificantes do gene *COI*, principalmente entre as angiospermas (Hollingsworth et al., 2009). Assim, o Consórcio para o Código de Barras da Vida (*Consortium for the Barcode of Life – CBOL*), por meio do grupo de trabalho voltado para plantas, sugeriu três regiões genômicas plastidiais (genes *matK* e *rbcL* e a região espaçadora *trnH-psbA*) e uma nuclear (região *ITS* do genoma nuclear ribossomal) para serem utilizadas como códigos de barra em plantas (CBOL, 2009; CBOL, 2011). Embora a combinação *rbcL* + *matK* tenha sido sugerida como sendo o código de barras padrão para plantas e as regiões *ITS* e *trnH-psbA* como suplementares, isoladamente

ou numa perspectiva de multilocos, nenhuma das regiões propostas pelo CBOL conseguiu o sucesso esperado para todos os grupos de plantas (Li et al., 2015).

A construção de uma biblioteca global de regiões de referência, com a finalidade de mapear toda a diversidade vegetal existente, ainda está em processo. Porém, o campo de aplicação potencial do código de barras do DNA, em botânica e serviços especializados relacionados, acabou extrapolando o foco primário de descoberta da biodiversidade (Hollingsworth et al., 2011; Kress, 2017). Uma rápida identificação em nível de espécie é crucial em muitas aplicações de importância econômica e social (Pecnikar & Buzan, 2013). A aplicação da técnica de DNA *barcoding* tem se mostrado bem-sucedida em estudos de levantamento da biodiversidade, biomonitoramento de espécies invasoras, análises forenses para avaliação de segurança alimentar e de medicamentos à base de plantas medicinais, monitoramento do comércio ilegal de espécies ameaçadas, estudos sobre ecologia alimentar, iniciativas de conservação etc. (Hollingsworth et al., 2011; Kress, 2017). No entanto, o estabelecimento de uma biblioteca de referência apropriada ainda é um pré-requisito crítico para as inúmeras aplicações da técnica, o que requer a geração de dados de códigos de barras de DNA a partir de amostras bem identificadas e comprovadas que representem a diversidade botânica da terra (Hollingsworth et al., 2011).

Diante da grande quantidade de informações produzidas desde o advento da técnica de DNA *barcoding* até o presente momento, é necessário compreender a dinâmica da produção desse conhecimento. Diferentes métodos de síntese de conhecimento estão sendo utilizados para o acompanhamento, avaliação e descrição de avanços nas mais diversas áreas, viabilizando novas pesquisas e contribuindo para o seu desenvolvimento (Vanti, 2002). Dentre as diferentes abordagens qualitativas e quantitativas de síntese, a cienciometria viabiliza a compreensão da organização das tendências da produção do conhecimento científico sobre determinado assunto (Spinak, 1996; Macias-Chapula, 1998; Vanti, 2002).

Portanto, buscando compreender as lacunas de pesquisa acerca do uso da técnica de DNA *barcoding* nos estudos de plantas, este trabalho tem o intuito de levantar e sistematizar a produção científica sobre a aplicação de DNA *barcoding* no estudo de plantas e, com isso, identificar as tendências e avanços científicos acumulados ao longo do tempo.

## **2. OBJETIVOS**

### **Geral**

Levantar e sistematizar a produção científica sobre o uso de DNA *barcoding* em plantas.

### **Específicos**

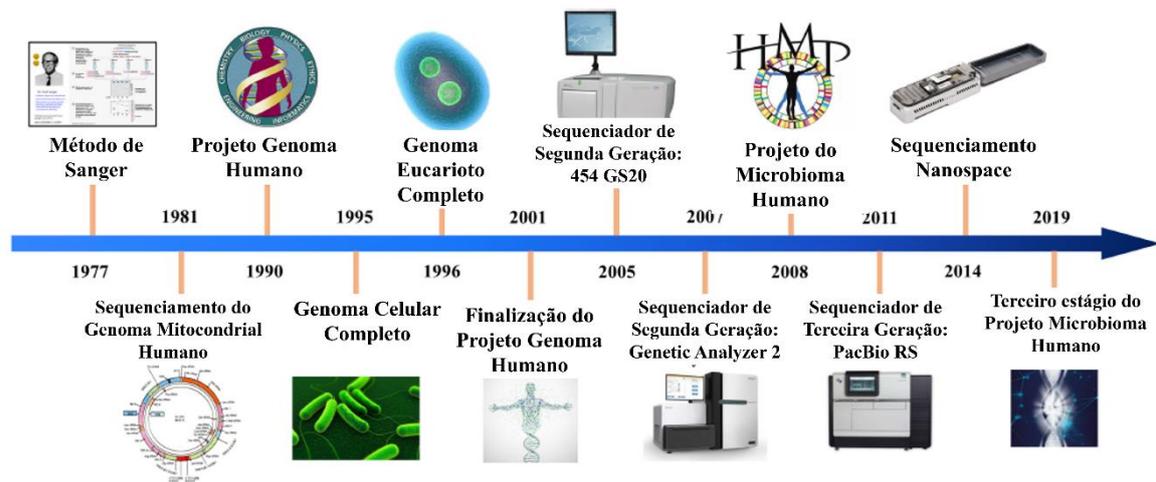
- Identificar e descrever padrões globais de produção do conhecimento científico, relacionados à aplicação da técnica de DNA *barcoding* em estudos de plantas ao longo do tempo;
- Identificar a estrutura da comunidade científica a partir das redes de colaboração estabelecidas entre pesquisadores, instituições de pesquisa e países de origem;
- Categorizar os artigos quanto ao tipo de uso/aplicação da técnica em estudos de plantas;
- Levantar os diferentes usos e aplicações da técnica e suas contribuições para o estudo das plantas.

### **3. REVISÃO DE LITERATURA**

#### **3.1 Tecnologias de Sequenciamento de DNA**

A partir da descoberta da estrutura do DNA, grandes avanços foram feitos no entendimento da complexidade e diversidade dos genomas. Na década de 1970, os primeiros métodos de sequenciamento de DNA foram criados, dando início à era de “Sequenciamento de Primeira Geração”, através do desenvolvimento de dois protocolos: o método de sequenciamento químico de Alan Maxam e Walter Gilbert e o sistema Alan Coulson e Frederick Sanger (Goodwin et al., 2016). O procedimento de Maxam e Gilbert determina as sequências de bases da fita de DNA através da clivagem química da terminação 5' dos fragmentos de DNA marcados radioativamente (Maxam & Gilbert, 1976). Já na metodologia de Sanger e Coulson, uma molécula de DNA cuja sequência deve ser determinada, é convertida em fitas simples que, por sua vez, são utilizadas como molde para sintetizar uma série de fitas complementares. Cada uma dessas fitas termina aleatoriamente em um nucleotídeo específico diferente (Sanger et al., 1977). A série resultante de fragmentos de DNA em ambos os métodos, é separada por eletroforese e analisada para revelar a sequência do DNA (Goodwin et al., 2016).

A técnica desenvolvida por Sanger et al. (1977) passou a ser comumente referida como Sequenciamento de Sanger e tornou-se o método de sequenciamento prevalecente nos próximos anos, sendo amplamente utilizada até os dias de hoje, mesmo com algumas limitações (Figura 1). O Sequenciamento de Sanger transformou a biologia ao fornecer ferramentas para decifrar genes completos e, mais tarde, genomas inteiros, oferecendo a oportunidade para explorar a variação genética dos organismos (Figura 1) (Kircher & Kelso, 2010; Morthie et al., 2011, Yang et al., 2020). A partir, então, do que se chamou de “Sequenciamento de Primeira Geração”, uma infinidade de inovações apoiou o início do Projeto Genoma Humano na década de 1990 e sua conclusão revelou a necessidade de tecnologias e conjuntos de dados maiores e mais avançados para responder às complexas questões biológicas que foram surgindo (Figura 1) (Shendure & Aiden, 2012; Goodwin et al., 2016). Desde a conclusão do Projeto Genoma Humano em 2003, um progresso extraordinário foi observado nas tecnologias de sequenciamento do genoma, o que levou a uma diminuição do custo por megabase e um aumento do número de diversidade de genomas sequenciados (Figura 1) (Dijk et al., 2014; Goodwin et al., 2016; Yang et al., 2020).



**Figura 1.** Histórico da tecnologia de sequenciamento genético, com início na década de 1970 com o método proposto por Sanger et al. (1977), passando pelo Projeto Genoma Humano e o desenvolvimento do Sequenciamento de Alto Rendimento. Adaptado de: Yang et al. (2020).

O lançamento da primeira plataforma de sequenciamento de alto rendimento, em meados dos anos 2000, significou uma queda simbólica nos custos do processo, denominado sequenciamento de nova geração (*Next-generation sequencing – NGS*) (Dijk et al., 2014). Os métodos de Sequenciamento de Alto Rendimento (*High Throughput Sequencing – HTS*) são caracterizados pela capacidade de sequenciamento de milhares de moléculas simultaneamente, favorecendo uma maior profundidade da cobertura de sequência (profundidade de sequenciamento). Quanto maior a cobertura de sequenciamento, maior a chance de evitar erros durante a montagem de todo o genoma ou da região de interesse (Shneyer & Rodionov, 2019). Não é uma coincidência que a técnica de *DNA barcoding* tenha se desenvolvido em conjunto com as investigações em genômica, uma vez que é uma ferramenta para identificação rápida de espécies com base em sequências de DNA, enquanto a genômica compara a estrutura e a função de genes. Ambas compartilham, então, uma ênfase na aquisição de dados genéticos em grande escala, que oferece novas respostas a questões antes fora do alcance de muitas disciplinas com dados limitados (Hebert et al., 2003; Kress & Erickson, 2008).

Novas plataformas como as da Oxford Nanopore e PacBio (Figura 1), vem sendo sugeridas como alternativas econômicas ao sequenciamento Sanger para a geração de códigos de barras de DNA (Kennedy et al., 2020). Ambas permitem que o sequenciamento alcance comprimentos de leitura superiores a qualquer tecnologia anterior, além de permitirem o processamento de milhares de códigos de barras de DNA em uma única

execução de sequenciamento. A plataforma de DNA Nanopore é um exemplo de tecnologia em desenvolvimento, baseada em nanoporos que oferece grandes vantagens potenciais no contexto da pesquisa de tecnologias de sequenciamento, apresentando como exemplo o MinION, um dispositivo portátil que pode ser usado por grupos de pesquisa em ambientes de campo, não ficando restrito a ambientes que necessitam de estrutura e logística apropriados (Figura 1) (Menegon et al., 2017, Pomerantz et al., 2017).

A disponibilidade de plataformas de sequenciamento de alto rendimento cada vez mais modernas e a necessidade de identificação de táxons sob diferentes condições, inclusive de logística de trabalho, viabilizaram o surgimento do *metabarcoding* (Taberlet et al., 2012). *Metabarcoding* é um caso especial de DNA *barcoding* que utiliza o mesmo banco de dados de referência que o código de barras de DNA, mas permite a identificação automática de várias espécies, a partir de uma amostra global contendo organismos inteiros ou de única amostra ambiental contendo DNA degradado (solo, água, fezes etc.) (Dormontt et al., 2020). O Sequenciamento de Alto Rendimento, particularmente associado ao *metabarcoding*, permite análises de diversidade e interações em estudos de comunidades em escalas sem precedentes e com reduções de custo cada vez mais consideráveis (Kennedy et al., 2020).

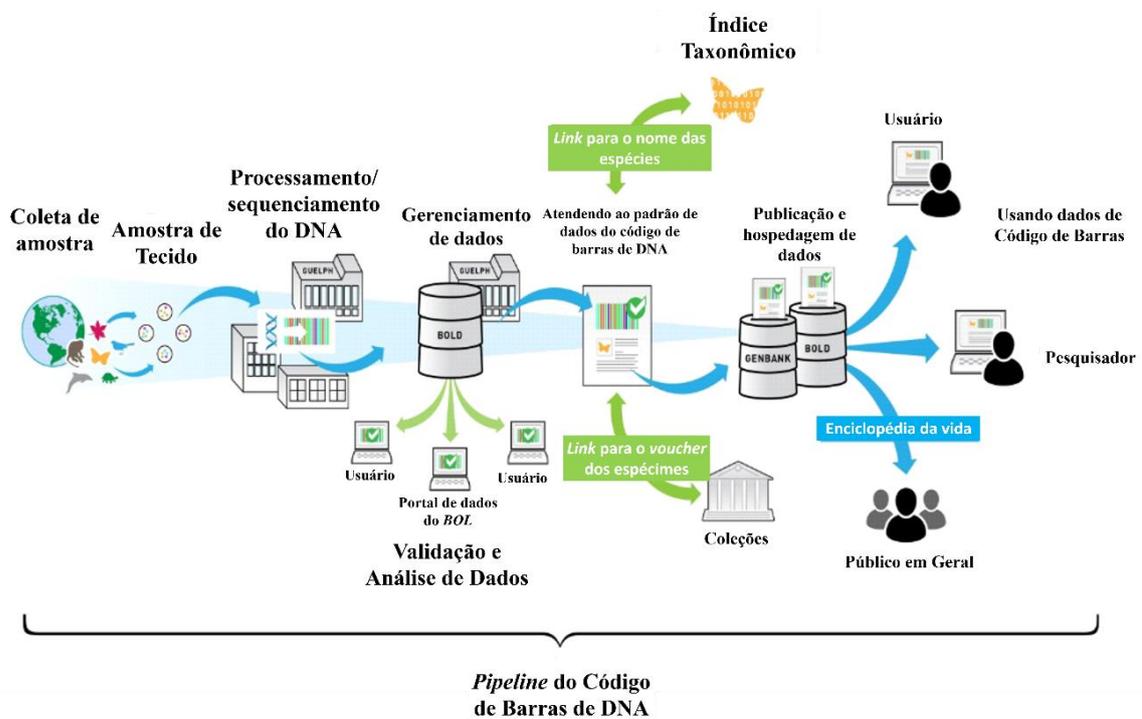
### 3.2 DNA *barcoding*

Em 2003, Hebert e colaboradores propuseram a técnica baseada em uma abordagem molecular para identificação de espécies denominada DNA *barcoding*, que consiste no uso de uma sequência curta de DNA (400 a 800 pares de base), a partir de uma região padrão e comum no genoma, utilizada como um “código de barras” para identificação dos organismos (Hebert et al., 2003). Ao combinar os pontos fortes da genética molecular, das tecnologias de sequenciamento e da bioinformática, esses marcadores oferecem um meio preciso para reconhecer espécies, descritas e nomeadas anteriormente, recuperando informações sobre elas, além de complementar métodos clássicos de identificação de novas espécies (Hebert et al., 2003; Hebert & Gregory, 2005; Kress, 2017).

Para coordenar os trabalhos com DNA *barcoding* em eucariotos, foi organizado em 2004 o *Consortium for the Barcode of Life – CBOL* (Consórcio Internacional para o Código de Barras da Vida), que incluiu organizações e pesquisadores que trabalham na estrutura desta abordagem. A região selecionada como DNA *barcode*, bem como os protocolos de seu uso, deverão ser aprovados e homologados pelo consórcio. O processo de geração e aplicação de DNA *barcodes*, para fins de identificação, envolve etapas básicas como: i) a

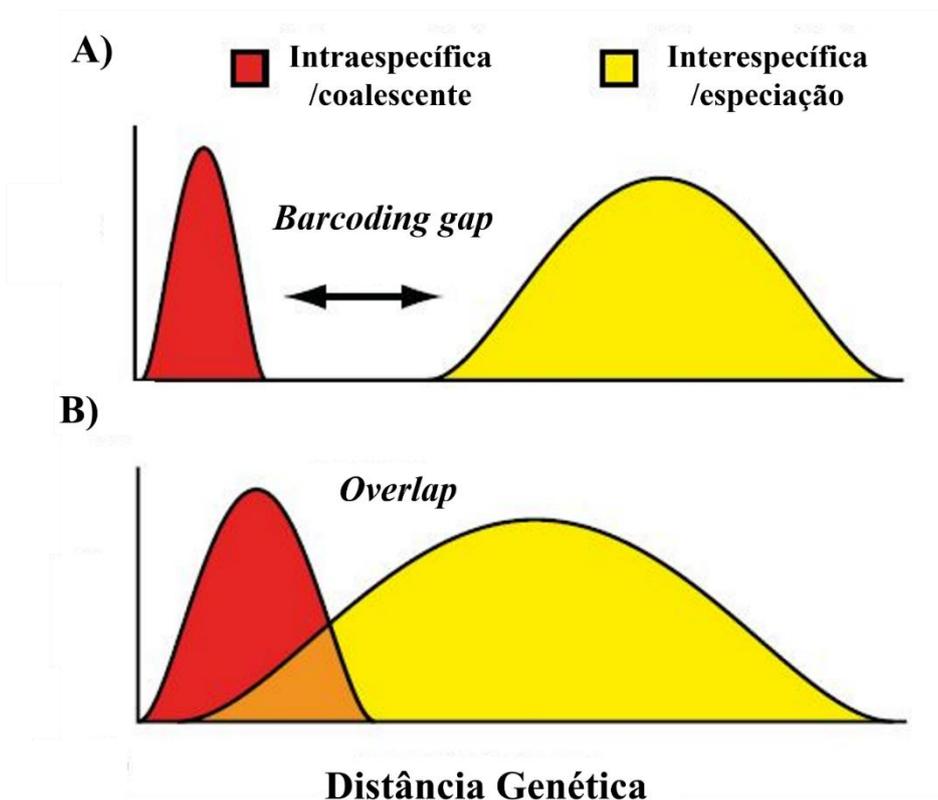
construção de uma biblioteca de marcadores de espécies conhecidas e *ii*) a combinação da sequência do DNA *barcode*, de uma amostra desconhecida, com a biblioteca de referência, a partir de um algoritmo de correspondência (Figura 2). Essas etapas ainda representam um desafio, considerando que a maior parte da biodiversidade biológica ainda não foi documentada (Meyer & Paulay, 2005; Kress, 2017).

O volume potencial de dados produzidos pela iniciativa do *CBOL* despertou a comunidade científica envolvida para a necessidade de um *software* para oferecer suporte a novos aspectos ligados ao código de barras de DNA, o que motivou o desenvolvimento do Sistema de Códigos de Barras da Vida (*Barcode of Life Data System – BOLD*). O *BOLD* – [www.barcodinglife.org](http://www.barcodinglife.org) – é uma plataforma de bioinformática integrada que oferece suporte desde a coleta de espécimes até a biblioteca de códigos de barras devidamente validada, se caracteriza por ser: primeiro um repositório para o espécime e registros de sequência que forma a unidade de dados básica de todos os estudos de código de barras; em segundo lugar é uma “bancada” de trabalho que auxilia no gerenciamento, garantia de qualidade e análise de dados de códigos de barras; e em terceiro fornece um veículo para colaboração entre comunidades de pesquisa geograficamente dispersas, combinando segurança flexível e recursos de entrada de dados com entrega baseado na *web* (Ratnasingham & Hebert, 2007).



**Figura 2.** Fluxo de trabalho para desenvolvimento e aplicações de DNA *barcodes*, partindo da coleta de amostras até a publicação e acesso aos dados. Adaptado de: Fajarningsih (2016).

Dentro do processo de estabelecimento de uma biblioteca de DNA *barcodes*, o ponto crítico para o desenvolvimento de um código de barras capaz de identificar os organismos é a precisão, que depende especialmente da extensão e da distância entre a variação intraespecífica e a divergência interespecífica oferecida pelo marcador selecionado (*barcoding gap*) (Figura 3A). Quanto mais sobreposição (*overlap*) houver entre a variação genética dentro das espécies e a divergência que separa espécies-irmãs, menos eficaz o *barcode* se torna (Figura 3B) (Meyer & Paulay, 2005). Desse modo, para ser prática como um “código de barras”, a região de interesse deve apresentar baixa diversidade entre indivíduos da mesma espécie além variabilidade e divergência genéticas significativas entre espécies, possuir regiões flanqueadoras conservadas para o anelamento de *primers* universais, ser pequena o suficiente para que possa ser isolada de amostradas danificadas e ser sequenciada com facilidade (não ultrapassando 700-800 pares de base), além de apresentarem poucas inserções e deleções (*indels*), para facilitar o alinhamento (Hebert & Gregory, 2005; Kress & Erickson, 2008).



**Figura 3.** Representação do *barcoding gap* com as respectivas distribuições da variação intraespecífica, mostrada em vermelho e da interespecífica em amarelo. **A)** Mundo ideal para DNA *barcoding*, com distribuições discretas e sem sobreposições (*overlap*). **B)** Uma versão alternativa do mundo ideal com sobreposição significativa e sem *gap*. Adaptado de: Meyer & Paulay (2005).

Hebert e colaboradores (2003) propuseram, inicialmente, um fragmento da região 5' do gene mitocondrial do citocromo oxidase *I* (*COI*), com cerca de 650 pares de base (pb) como região alvo para estudos em animais, devido à sua herança predominantemente uniparental, ser um genoma pequeno e com poucos genes, apresentar baixas taxas de recombinação e taxas elevadas de substituição de nucleotídeos e possibilidade de recuperação eficaz de informações gênicas de amostras em estados inadequados de preservação. Em contraste ao observado com a aplicação de DNA *barcoding* em animais, a técnica aplicada em plantas se mostrou mais complexa, visto que o fragmento do gene *COI* mitocondrial apresenta taxas de substituição em geral muito baixas e o rearranjo estrutural do genoma mitocondrial variando muito entre os diferentes grupos de plantas (Hollingsworth et al., 2009; Shneyer & Rodionov, 2019).

### 3.3 DNA *barcoding* para plantas

O Grupo de Trabalho de Plantas (*Plant Working Group – PWG*) é um dos cinco grupos de trabalho que compõem o *CBOL*. Foi constituído em 2005 e organizado para pesquisar códigos de barras de DNA de plantas, sob coordenação do *Kew Royal Botanical Garden* de Londres, composto por pesquisadores de sete países diferentes, de onze instituições participantes (Shneyer, 2009). Após a constatação de que o gene *COI* não teria sucesso como DNA *barcode* aplicado ao estudo de plantas, novas possibilidades passaram a ser consideradas a partir dos genomas plastidial e nuclear (CBOL, 2009, 2011).

A partir de investigações iniciais de regiões plastidiais, sete sequências de DNA plastidial foram propostas como promissoras, em estudo publicado pelo *PWG CBOL* em 2009, que incluía: *i*) quatro sequências codificadoras (gene maturase K – *matK*; gene da subunidade  $\beta$  da RNA polimerase – *rpoB*; gene da subunidade  $\beta'$  da RNA polimerase – *rpoC1*; a grande subunidade de ribulose-bifosfato carboxilase – *rbcL*); *ii*) três sequências não codificantes (espaçadores *trnH-psbA*, *atpF-atpH*, *psbK-psbI*); e *iii*) a região *ITS* dos genes nucleares ribossômicos, que foi adicionada às sequências candidatas posteriormente (Tabela 1) (CBOL, 2009; Hollingsworth et al., 2009; CBOL, 2011).

**Tabela 1.** Características dos diferentes marcadores que foram rotineiramente incluídos em estudos de códigos de barras de plantas. Adaptado de: Hollingsworth et al. (2011).

Marcador	Fonte genômica	Tipo	Número aproximado de espécies (GenBank)	Comprimento médio do <i>amplicon</i> <sup>1</sup> (bases)
<i>nrITS</i>	Nuclear	Íntron e gene 5.8S	52.450	705
<i>nrITS2</i>	Nuclear	Íntron	57.579	494
<i>atpF-H</i>	Plastidial	Íntron	664	669
<i>matK</i>	Plastidial	Codificação de proteínas	22.701	889
<i>psbK-I</i>	Plastidial	Intergênico	626	468
<i>rbcL</i>	Plastidial	Codificação de proteínas	20.374	654
<i>rpoB</i>	Plastidial	Codificação de proteínas	1.970	548
<i>rpoCl</i>	Plastidial	Codificação de proteínas	3.075	616
<i>trnH-psbA</i>	Plastidial	Intergênico	11.539	509

<sup>1</sup> Estimativas do tamanho do *amplicon* (incluindo regiões de *priming*) foram feitas a partir de todos os genomas plastidiais anotados de plantas terrestres depositados no GenBank e, para *nrITS*, a partir de um conjunto com curadoria de sequências nrITS.

As regiões descritas foram testadas, conforme os critérios e diretrizes do *CBOL* para a seleção de locos, em diferentes grupos de plantas, a fim de avaliar a qualidade de amplificação com *primers* universais e o nível de discriminação de espécies vegetais (*CBOL*, 2009). Dentre o conjunto de locos testado, observou-se que as regiões codificantes apresentaram vantagens em relação às regiões não codificantes, uma vez que são facilmente alinháveis e os pseudogenes são identificados com mais facilidade. No primeiro estágio, constatou-se que as regiões não codificantes apresentaram variabilidade insuficiente e as sequências suficientemente variáveis apresentaram problemas com o comprimento, ultrapassando o tamanho ideal e inviabilizando sua utilidade como possível código de barras (*CBOL*, 2009; Shneyer, 2009).

Das quatro regiões codificadoras iniciais, tanto *rpoCl* quanto *rpoB* apresentaram bom desempenho em relação à universalidade/qualidade de sequência, mas tiveram baixo poder discriminatório, ao passo que, entre as regiões não codificadoras foram descartadas: *atpF-atpH*, que não apresentou boa resolução de espécies como código de barras simples devido comprometimento na recuperação de sequências bidirecionais e também a região *psbK-psbI* que, apesar de apresentar bom padrão discriminatório, teve o menor sucesso de sequenciamento, além de problemas na geração de leituras bidirecionais (*CBOL*, 2009).

Embora tenham apresentado atributos desejáveis, *rbcL*, *matK* e *psbA-trnH* não atenderam individualmente a todos os critérios previstos pelo *CBOL*. Segundo Kress & Erickson (2007), o espaçador *trnH-psbA* demonstrou boa amplificação em plantas terrestres

com um único par de *primers* além de altos níveis de discriminação em nível de espécie. Entretanto, as principais limitações desse loco são os problemas na obtenção de sequências bidirecionais de alta qualidade, além de apresentar ampla variação de comprimento (superior a 1.000 pb), dependendo do grupo de plantas estudado, como observado em monocotiledôneas e coníferas (Erickson et al., 2008).

Entre as regiões gênicas plastidais selecionadas, *rbcL* é a mais bem caracterizada. Melhorias ao longo do desenvolvimento de *primers*, tornaram essa região adequada para a recuperação de sequências bidirecionais de alta qualidade e, embora não seja a região mais variável, é um componente frequente das combinações multiloco (Fazekas et al., 2008). A região gênica plastidial *matK* é a que mais se assemelha ao gene *COI* em termos de taxa de evolução, demonstrando altos níveis de discriminação para angiospermas. No entanto, vários estudos sobre a universalidade dos *primers* indicaram variações quanto ao êxito de rotina e a recuperação irregular, apresentando maior sucesso em angiospermas e menor em gimnospermas e criptógamas, respectivamente (Kress & Erickson, 2007; Fazekas et al., 2008; Lahaye et al., 2008).

Em resumo, *rbcL* oferece alta universalidade, apesar de não apresentar um poder discriminatório excepcional, enquanto *matK* e *psbA-trnH* oferecem maior resolução, porém exigem mais esforço para desenvolvimento dos *primers*. Logo, nenhum loco único atendeu aos padrões e diretrizes determinados pelo *CBOL*, não sendo capaz de servir como “código de barras” para todos os grupos de plantas e, para isso, seria necessário encontrar uma combinação de regiões (*CBOL*, 2009). Dentre as três combinações de locos destacadas, o *CBOL* chegou à conclusão de que a melhor opção seria *rbcL+matK*, visto que essa combinação evitaria o aumento de custos para sequenciar três locos ao invés de dois, em conjuntos de amostras muito grandes, além de impedir atrasos adicionais na implementação de um código de barras padrão para plantas terrestres, já que o prazo de finalização previsto era 2007 (*CBOL*, 2009; Shneyer, 2009).

O *CBOL* (2009) destaca que essa combinação de dois locos foi a escolha majoritária entre os pesquisadores do *PWG* visto que as sequências de alta qualidade de *rbcL* são facilmente recuperáveis em linhagens filogeneticamente divergentes e apresentam bom desempenho em testes de discriminação associados a outros locos, enquanto que, o desenvolvimento de estratégias de amplificação para *matK*, melhorou sua recuperação em angiospermas, gerando perspectivas de melhorias adicionais dentro deste grupo e para os demais grupos de plantas terrestres. Portanto *rbcL+matK*, como código de barras padrão

para plantas terrestres, representou uma solução, com considerações práticas, que buscou associar universalidade, qualidade de sequência, discriminação e custo, oferecendo a oportunidade de aproveitar tecnologias de sequenciamento automatizadas de alto rendimento, para estabelecer uma poderosa estrutura universal para identificação de plantas. Apesar de recomendar *rbcL+matK* como código de barras padrão para as plantas terrestres, o *CBOL* ainda teve e tem como desafio estudos em que é necessário aumentar o poder de resolução e universalidade dessa combinação (CBOL, 2009, 2011).

Durante a 3ª Conferência Internacional para o Código de Barras da Vida (*Third International Barcode of Life Conference*), que aconteceu no México em 2009, o *CBOL* sugeriu que a combinação *rbcL+matK* deveria contar com regiões complementares já testadas, tais como: *psbA-trnH*, *atpF-atpH* e *psbK-psbI*, o íntron *trnL*, utilizado em situações envolvendo tecidos com níveis diferentes de degradação, e o espaçador de transcrito interno do DNA ribossomal nuclear – *ITS* (composto pelas subunidades *ITS1.58S.ITS2*) (Hollingsworth et al., 2011). Em 2011, o *CBOL* avaliou a eficácia e a universalidade dos marcadores *trnH-psbA* e *ITS* como possíveis regiões complementares à combinação (*rbcL+matK*) (CBOL, 2011). Essas regiões já haviam sido testadas em outros estudos, como em Kress & Erickson (2007) e Hollingsworth et al. (2009).

Em sua publicação oficial, o *CBOL* (2011) demonstrou que: *i*) os marcadores plastidiais apresentaram altos níveis de universalidade (87,1-92,7%), enquanto o *ITS* teve um desempenho relativamente bom em angiospermas (79%), porém o mesmo não foi observado em gimnospermas, sendo identificada ainda uma frequência modesta de sequências de cópias múltiplas e casos de contaminação fúngica; *ii*) em grupos taxonômicos para os quais o sequenciamento direto do marcador é possível, *ITS* demonstrou o maior poder discriminatório dentre os quatro marcadores e, em combinação com qualquer um dos marcadores plastidiais, foi capaz de discriminar 69,9-79,1% das espécies, em comparação com apenas 49,7% do código de barras principal, corroborando estudos (Kress & Erickson, 2007; Yao et al., 2010; Chen et al., 2010), que indicaram rápida evolução nessa região, o que leva a mudanças genéticas que podem diferenciar espécies intimamente relacionadas; *iii*) em testes com vários indivíduos de espécies morfologicamente definidas, as atribuições baseadas em *ITS* e regiões plastidiais foram incongruentes em alguns gêneros amostrados (45,2% para gêneros com mais de uma espécie amostrada), o que destaca a importância de amostrar vários indivíduos e usar marcadores com diferentes modos de herança (CBOL, 2011).

A incongruência de atribuição constatada entre marcadores plastidiais e *ITS* pode resultar de hibridação e introgressão ou classificação incompleta, fenômenos que acontecem com frequência em plantas (CBOL, 2011). Tal fato deixa claro que usar apenas marcadores plastidiais pode não permitir a discriminação entre espécies intimamente relacionadas, e estabelecer um banco de dados usando apenas um indivíduo, de cada espécie, associado apenas a marcadores plastidiais, pode levar a conclusões equivocadas. Os maiores problemas com *ITS* foram encontrados em gimnospermas, em que a grande variabilidade no comprimento e a falta de *primers* universais dificultaram o sucesso da *PCR* (reação em cadeia da polimerase) e do sequenciamento. Além disso as inúmeras cópias da região no genoma nuclear podem apresentar divergências de sequência no mesmo indivíduo, principalmente em organismos que apresentam poliploidia como em diversos grupos de plantas, o que não representa um problema significativo na sua aplicação em fungos visto que muitas espécies apresentam etapas haploides de vida bem determinadas. Em casos em que *ITS* é difícil de amplificar, *ITS2* representa uma alternativa útil para gimnospermas, ou mesmo para outras plantas com sementes, devido à facilidade de amplificação com um único conjunto de *primers* universais (CBOL, 2009; CBOL, 2011).

Até aqui, observa-se que o caminho para estabelecer um código de barras padrão para as plantas não se demonstrou fácil e ainda não há uma conclusão definitiva. A atuação de muitas profissões relacionadas ao estudo de plantas, envolve estabelecer e/ou usar protocolos de identificação sendo necessário combinar a disponibilidade, com o poder discriminatório da técnica. Para as plantas, diferentemente do gene *COI*, que foi bem estabelecido como loco único para os animais, a viabilidade da aplicação da ferramenta DNA *barcode* envolve o estabelecimento de uma combinação mínima viável (financeiramente e experimentalmente) para sua utilização em escala comercial.

### **3.4 Aplicações de DNA *barcoding* no estudo de plantas**

A construção de um banco de dados global de regiões de referências, com a finalidade de mapear toda a diversidade vegetal existente, ainda está em processo. Porém, a identificação em nível de espécie é crucial em muitas aplicações de importância econômica e social. Desse modo, o campo de aplicação potencial do código de barras do DNA em botânica e serviços especializados relacionados, acabou extrapolando o foco inicial de descoberta e descrição da biodiversidade (Pecnikar & Buzan, 2013; Kress, 2017). O leque

de aplicações de DNA *barcoding* no estudo de plantas inclui: inventários de biodiversidade, avaliação filogenética, biossegurança e saúde pública, avaliação de conservação e preservação ambiental, interações entre espécies e redes ecológicas, informações sobre diversidade críptica, metadados de DNA *barcoding*, análises de ecologia forense, montagem de comunidades, tráfico de espécies ameaçadas e monitoramento de produtos comerciais (Hollingswoth et al., 2011; Kress, 2017).

Atualmente, a capacidade do código de barras do DNA em distinguir espécies de uma variedade de *taxa* e revelar espécies crípticas está bem documentada (Pecnikar & Buzan, 2013). O DNA *barcoding* tem se mostrado útil no estudo de grupos taxonomicamente difíceis, como o gênero de leguminosas *Dumasia*, que apresenta espécies com poucas diferenças morfológicas e as plantas sem flores ou frutos são difíceis de se identificar com precisão. Neste estudo, as análises filogenéticas indicaram que a espécie (*Dumaria yunnanensis* não é monofilética e é separada por dois ramos independentes, que podem resultar de diferenciação críptica, demonstrando que dados moleculares podem aprofundar a compreensão taxonômica e fornecer uma abordagem eficiente para a identificação de espécies (Jiang et al., 2020).

A aplicação do código de barras de DNA tem se mostrado bem-sucedida em estudos de avaliação rápida da biodiversidade a partir de amostras ambientais, uma vez que fragmentos curtos de DNA persistem no meio ambiente e permitem uma avaliação da biodiversidade local do solo ou da água e estabelecimento da composição da dieta, conforme revisado por Valentini et al. (2009) e ainda a avaliação da estrutura de comunidades filogenéticas em florestas tropicais (Heckenhauer et al., 2017). O uso de DNA *barcodes* para realização de determinações taxonômicas precisas é uma importante ferramenta para o monitoramento, conservação e avaliação de habitats ameaçados, especialmente em biomas tropicais onde a biodiversidade é pouco conhecida (Costion et al., 2011; Kress, 2017).

O biomonitoramento de espécies vegetais exóticas, como o caso das lentilhas-d'água em ambientes lacustres na Itália, é uma ação importante para o estabelecimento de alertas precoces de invasão a habitats vulneráveis. Uma vez identificadas as espécies invasoras, é possível estabelecer por exemplo, a fonte e forma de entrada dessas espécies no habitat considerado, bem como executar ações de manejo das mesmas para que não garantam riscos às espécies endêmicas (Marconi et al., 2019). O biomonitoramento permite ainda a investigação forense do comércio ilegal de espécies ameaçadas e autenticação da qualidade produtos alimentícios ou terapêuticos de origem vegetal, normalmente oriundos de amostras

processadas ou descaracterizadas o que impede uma primeira identificação com base nas características físicas do material (Pecnikar & Buzan, 2013).

O uso de *metabarcoding* para análises de amostras complexas (processadas e/ou descaracterizadas), tem ganhado campo graças à evolução das tecnologias de sequenciamento de alto rendimento, permitindo a análise de milhares de amostras simultaneamente, a fim de avaliar sua autenticidade e segurança (Palhares et al., 2021). O declínio de polinizadores em todo o mundo, leva a déficits de polinização tanto em safras comerciais quanto em plantas selvagens. Trabalhos de investigação da riqueza de pólen, associando o *metabarcoding* e a microscopia eletrônica, são importantes para um melhor planejamento na oferta de espécies de plantas com flores em jardins urbanos, por exemplo, como destacado por Bansch et al. (2020) e Potter et al. (2019).

Xin e colaboradores (2018), destacam a preocupação global quanto ao perigo potencial de medicamentos tradicionais à base de plantas (*Traditional Herbal Medicinal Products – THMPs*), visto que produtos abaixo do padrão e falsificados são uma ameaça potencial à saúde pública. O uso de DNA *barcoding* e especialmente de *metabarcoding*, em inúmeros estudos, determinou que os métodos de controle de qualidade atuais não são suficientes para garantir a presença de ingredientes fitoterápicos e a detecção de contaminantes/adulterantes, sendo necessária a utilização de métodos complementares como o código de barras de DNA (Xin et al., 2018). Palhares e colaboradores (2021) também salientam a importância de países como o Brasil, com um longo histórico tradicional de utilização de produtos fitoterápicos, de se adotar protocolos para autenticação baseados em códigos de barras de DNA.

### **3.5 Ferramentas para avaliação do desenvolvimento científico**

Diante da grande quantidade de informações produzidas desde o advento da técnica de DNA *barcoding* até o presente momento, faz-se necessário compreender a dinâmica da produção desse conhecimento. O mapeamento da produção científica tem se tornado uma atividade essencial para estudiosos nas mais diversas áreas, uma vez que, à medida que o número de publicações continua a se expandir em taxas crescentes e as publicações se desenvolvem de forma fragmentada, a tarefa de acumular e organizar o conhecimento se torna mais complexa. A determinação da estrutura intelectual e da frente de pesquisa dos

domínios científicos são importantes, não apenas para a pesquisa, mas também para a formulação de políticas e novas práticas (Aria & Cuccurullo, 2017).

Diferentes métodos de síntese de conhecimento, estão sendo utilizados para o acompanhamento, avaliação e descrição de avanços nas mais diversas áreas, viabilizando novas pesquisas e contribuindo para o seu desenvolvimento (Vanti, 2002). Dentre as abordagens de síntese do conhecimento existentes, a cienciometria é um método quantitativo que viabiliza a compreensão da organização das tendências da produção do conhecimento científico sobre determinado assunto, do ponto de vista de sua produção ou comunicação (Bufem & Prates, 2005; Silveira & Ferreira, 2019).

A produção de indicadores quantitativos mais representativos se tornou uma realidade concreta em função da criação, manutenção e informatização de dados para armazenamento e consulta de informações científicas (Silveira & Ferreira, 2019), através de plataformas eletrônicas como a *Web of Science (WoS)*. A *WoS* é uma base de dados multidisciplinar que indexa somente os periódicos mais citados em suas respectivas áreas, contando, atualmente, com cerca de 165 milhões de artigos indexados em 21.300 periódicos. Por meio da *WoS*, estão disponíveis ferramentas para análise de citações, referências, índice *h*, além do acesso a um índice de citações, informando, para cada artigo, os documentos por ele citados e os documentos que o citaram (Fonte: <https://www.clarivate.com>). O portal de periódicos CAPES por meio de assinatura com a editora *Clarivate Analytics*, responsável pela *WoS*, permite acesso a referências e resumos em todas as áreas do conhecimento (Fonte: <http://www-periodicos-capes-gov-br>). Pesquisas cienciométricas utilizando bancos de dados como a *WoS*, buscam estabelecer as contribuições a respeito de assuntos de interesse e compreender os caminhos da ciência visando identificar as lacunas do conhecimento e propor novas agendas de pesquisa.

## 4. MATERIAL E MÉTODOS

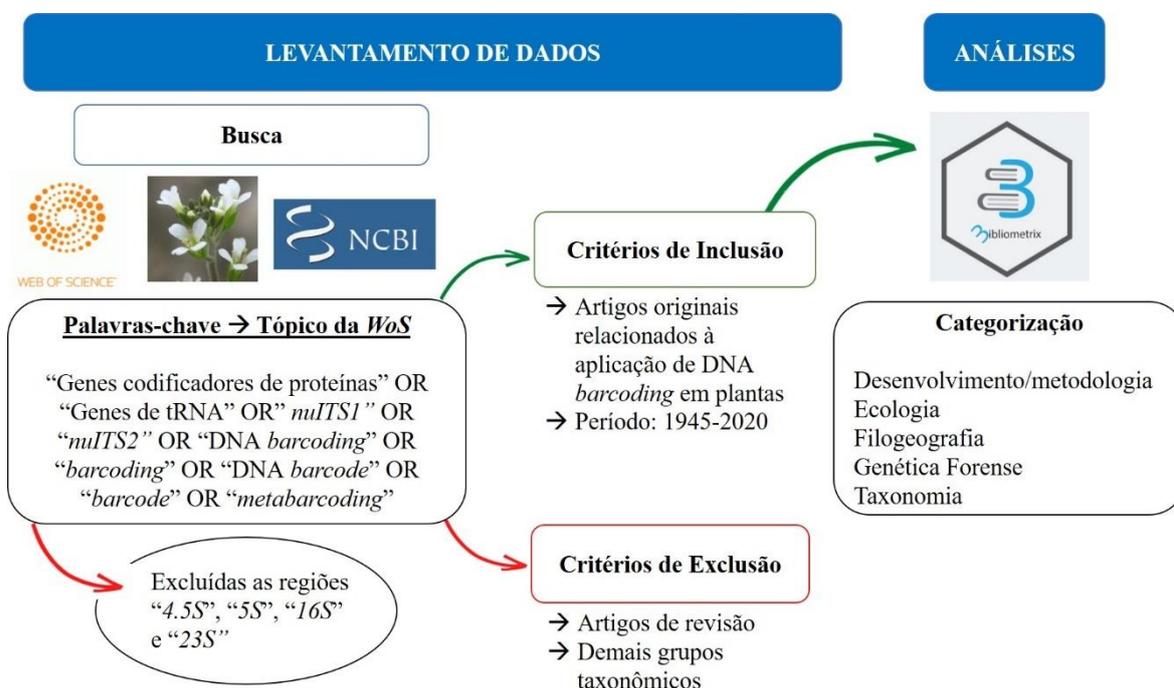
### 4.1 Análise Cienciométrica

Para acessar a literatura existente sobre a aplicação da técnica de DNA *barcoding* em plantas, inicialmente foram realizadas buscas nas bases de dados *Web of Science (WoS)* e *Scopus*. Apesar de um maior número de artigos ter sido recuperado através de *Scopus*, a base de dados da *WoS* foi escolhida para o presente trabalho, uma vez que foram encontrados muitos artigos sobrepostos entre ambas as bases, além de um número significativo de publicações fora do escopo do trabalho envolvendo principalmente microrganismos em *Scopus*. Optou-se por buscar pelo nome específico das regiões utilizadas como marcadores em plantas, visto que, quando testadas apenas as expressões *Plant AND “DNA barcoding” AND e/ou OR “DNA barcode”*, foi recuperado um número significativo de artigos cujos objetos de estudos eram bactérias, fungos e animais.

Foram utilizadas como palavras-chave combinadas, o nome das regiões de genes codificadores de proteínas e genes de tRNA do genoma plastidial da espécie modelo *Arabidopsis thaliana*, conforme anotação do genoma disponibilizada no *NCBI* (Disponível em: [https://www.ncbi.nlm.nih.gov/gene?LinkName=nucore\\_gene&from\\_uid=7525012](https://www.ncbi.nlm.nih.gov/gene?LinkName=nucore_gene&from_uid=7525012)). Além das regiões do genoma plastidial, foram inseridas as palavras-chave “*nuITS*” OR “*nuITS1*” OR “*ITS2*” OR “*ITS2*”, que são regiões genômicas nucleares do DNA ribossomal, amplamente utilizadas como marcadores complementares no estudo de plantas. O nome das regiões genômicas foi associado às expressões: “DNA *barcoding*” OR “*barcoding*” OR “DNA *barcode*” OR “*barcode*” OR “*metabarcoding*”. Foram retiradas das buscas as regiões codificadores das subunidades ribossomais “4.5S”, “5S”, “16S” e “23S” por serem regiões também presentes no genoma de bactérias, favorecendo a recuperação de muitos artigos relacionados a esses microrganismos (Figura 4).

Como critério de busca, foi utilizado o item “Tópico” da *WoS*, que procura automaticamente pelas palavras-chave de interesse nos resumos, títulos e palavras-chave dos documentos. A busca foi realizada abrangendo o período completo de publicação da base de dados até o ano mais recente, portanto de 1945 a dezembro de 2020. Após a busca foi realizada a leitura dos resumos dos artigos recuperados, selecionando-se apenas artigos originais e excluindo-se artigos de revisão, bem como os trabalhos relacionados a grupos taxonômicos que não correspondessem ao grupo das plantas. Como não foi definido um idioma para a busca, a *WoS* resgata todos os artigos relacionados ao tema que possuam pelo

menos o resumo em inglês. Os trabalhos de revisão foram desconsiderados, pois o intuito da busca era encontrar aplicações práticas de DNA *barcoding* no estudo das plantas (Figura 4).



**Figura 4.** Representação das etapas de levantamento e refinamento para obtenção do conjunto de dados de estudos com aplicação de DNA *barcoding* no estudo de plantas e as respectivas etapas de análise cienciométrica e de categorização. Dentro da etapa “Levantamento de dados”, as setas em vermelho representam os critérios de exclusão adotados, enquanto a seta verde representa os critérios de inclusão.

A partir do conjunto de dados obtido, foram realizadas análises estatísticas descritivas, conduzidas a partir do pacote *Bibliometrix* v. 3.0.4 (Aria & Cuccurullo, 2017), implementado no programa R v. 4.0.2 (R Core Team, 2020). Esse pacote importa dados de plataformas como a *WoS*, tornando possível a descrição das tendências da produção científica de interesse. Para tanto, o *input* do arquivo de dados foi exportado em formato *bibtex* (.bib) (Figura 4).

Do conjunto final de dados foram obtidos: *i*) o número total de publicações; *ii*) o ano da primeira publicação; *iii*) a quantidade de periódicos que publicaram e a verificação das zonas da Lei de Bradford (Bradford, 1961), com base na frequência de ocorrência dos documentos sobre os periódicos; *iv*) a média de citações por documento; *v*) o número de artigos por autor; *vi*) os autores mais produtivos; *vii*) a produção dos principais autores ao longo do tempo; *viii*) a produtividade dos autores avaliada com base no índice *h*, que trata-

se de uma medida de relevância que leva em consideração o número de publicações por autor e o número de citações por documento (Hirsch, 2005); *ix*) produtividade científica estimando a Lei de Lotka (Lotka, 1926), através da frequência de publicação dos autores; e *x*) as publicações mais citadas.

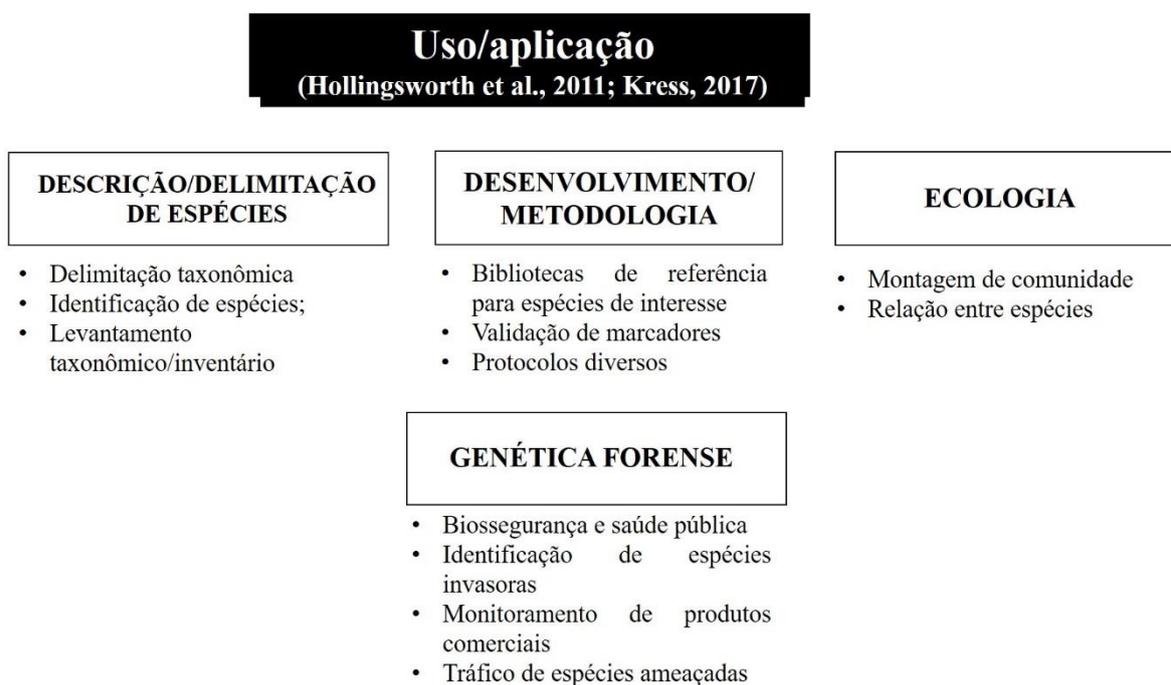
Periódicos com maior publicação de artigos sobre determinado assunto tendem a estabelecer supostamente um núcleo de qualidade superior e maior relevância, de forma que os artigos iniciais de um determinado assunto são submetidos a um número restrito de periódicos. Com o aumento de interesse sobre o assunto e seu respectivo desenvolvimento, torna-se possível o estabelecimento de outros núcleos de periódicos segundo a produtividade na área de interesse, constituindo-se nesse contexto um conjunto de três zonas segundo a Lei de Bradford, cada qual com um terço do total de artigos relevantes. A primeira zona contém um pequeno número de periódicos altamente produtivos, a segunda contém um número maior de periódicos menos produtivos, enquanto a terceira zona inclui um volume ainda maior de periódicos com reduzida produtividade. Assim, por meio da medida da produtividade dos periódicos, é possível estabelecer o núcleo e as áreas de dispersão (zonas) sobre determinado assunto em um mesmo conjunto de periódicos (Bradford, 1961; Junior et al., 2016).

Segundo a Lei de Lotka, o número de autores que fazem “n” contribuições em um determinado campo científico é aproximadamente  $1/n^2$  daqueles que fazem apenas uma contribuição, sendo que a proporção daqueles que fazem apenas uma contribuição é de aproximadamente 60% (Lotka, 1926). De modo geral, pode-se afirmar que a Lei de Lotka está fundamentada na premissa básica de que grande parte da produção científica se concentra nas pesquisas de poucos autores e que um pequeno número de autores produz mais sobre o assunto em estudo. A sua aplicabilidade é verificada na avaliação da produtividade de pesquisadores, na identificação dos centros de pesquisas mais desenvolvidos e no reconhecimento da solidez de uma área científica. Quanto mais sólida estiver uma ciência, maior a probabilidade de seus autores produzirem múltiplos artigos em um dado período do tempo (Alvarado, 2002).

Para analisar a estrutura da ciência em estudos utilizando DNA *barcoding* em plantas, foi gerada uma análise de rede colaborativa entre autores. Um gráfico de três campos foi gerado para visualizar as relações entre os 10 principais autores, as 10 palavras-chave mais utilizadas e os 10 periódicos que mais publicaram, também usando o *Bibliometrix* v. 3.0.4 (Aria & Cuccurullo, 2017).

## 4.2 Categorização

A partir da leitura dos resumos dos artigos selecionados, foram sintetizadas informações sobre o uso e/ou aplicação da técnica no estudo de plantas, a partir da categorização dos documentos, nas seguintes áreas do conhecimento: Descrição/delimitação de espécies, Desenvolvimento/metodologia, Ecologia, Genética Forense. Vale ressaltar que o mesmo artigo poderia ser inserido em mais de um grupo temático, dependendo do escopo (Figura 5).



**Figura 5.** Proposta de categorização dos artigos com base nas informações sobre uso e/ou aplicação da técnica de DNA *barcoding* no estudo de plantas.

Conforme descrição de aplicações da técnica de DNA *barcoding* nos estudos de plantas, em artigos de referência, como Hollingsworth et al. (2011) e Kress (2017), foram estabelecidos os seguintes critérios para categorização: i) em “Descrição/delimitação de espécies”, foram considerados trabalhos com foco em delimitação de espécies em grupos complexos, estabelecimento de filogenias, identificação de espécies de interesse, relacionadas a estudos farmacológicos e bioquímicos e o estabelecimento de inventário botânico em regiões de interesse como parques e reservas; ii) em “Desenvolvimento/metodologia”, foram alocados os estudos destinados ao desenvolvimento de bibliotecas de referência para espécies de interesse, bem como trabalhos que descrevem protocolos

diversos, relacionados ao aprimoramento da utilização da técnica de DNA *barcoding* em plantas e o teste de regiões propostas como marcadores; *iii*) para a categoria “Ecologia”, foram considerados estudos de montagem de comunidades de plantas, bem como os estudos de compreensão de relações ecológicas estabelecidas entre animais e plantas; *iv*) para “Genética Forense”, foram considerados estudos de biossegurança e/ou monitoramento de produtos comerciais de origem vegetal, além de estudos para identificação de espécies invasoras e monitoramento de espécies ameaçadas.

Desde a implementação das tecnologias do Sequenciamento de Alto Rendimento, o uso de *metabarcoding* como uma ferramenta de biodiversidade tem atraído imenso interesse (Creer et al., 2016). A utilização de *metabarcoding* se baseia nos mesmos dados de referência que o DNA *barcoding*, mas permite a identificação virtual de diferentes táxons dentro de uma amostra ou em amostras mistas (Dormontt et al., 2018). Pensando nisso, além da categorização geral da produção do conjunto de dados com base no uso e/ou aplicação da técnica de DNA *barcoding* no estudo de plantas, também foi discriminada a produção de trabalhos tanto DNA *barcoding* ou *metabarcoding* em cada categoria.

## 5. RESULTADOS E DISCUSSÃO

### 5.1 Análise Cienciométrica

Um total de 1.897 artigos originais foram recuperados, dos quais 1.073 estavam de acordo com os critérios de filtragem. Como o primeiro artigo sobre o tema foi publicado apenas em 2005, as análises contemplaram um período de 16 anos (de 2005 a 2020). Conforme resumo das principais informações sobre os dados apresentados na Tabela 2, os artigos foram publicados em 309 periódicos, com uma média geral de publicação por ano de 5,21 e média de citações para todos os documentos de 21,67. Cerca de 2.550 palavras-chave diferentes foram utilizadas nos estudos relacionados a DNA *barcoding* em plantas (Tabela 2).

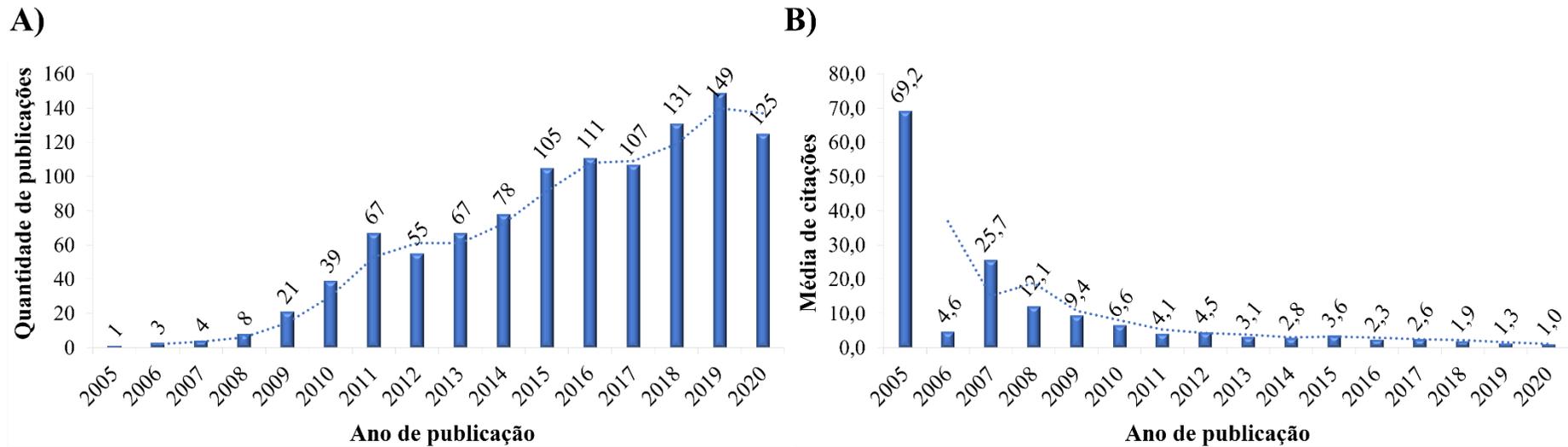
**Tabela 2.** Descrição geral das principais informações encontradas na busca realizada no banco de dados da *Web of Science*.

Descrição	Resultados
Período	2005-2020
Documentos	1.073
Periódicos	309
Média de publicações por ano	5,21
Média de citações por documentos	21,67
Palavras-chave do autor	2.550
Autores	3.747
Aparições do autor	5.728
Documentos de autoria única	16
Média de documentos por autor	0,29
Média de autores por documento	3,49
Coautores por documento	5,34

A produção científica apresentou variação ao longo da janela de tempo considerada (2005-2020), mas sempre com uma tendência crescente de trabalhos publicados, com destaque para os últimos cinco anos que comportam o maior índice de publicações, correspondendo a 67,8% do total (Figura 6A). Vale destacar o fato de que existem publicações em todos os anos no espaço de tempo considerado. Observa-se que o ano com maior média global de citações por documento (2005) (Figura 6B) não coincide com o ano de maior publicação (2019) (Figura 6A). O ano de 2005 apresenta a maior média global de

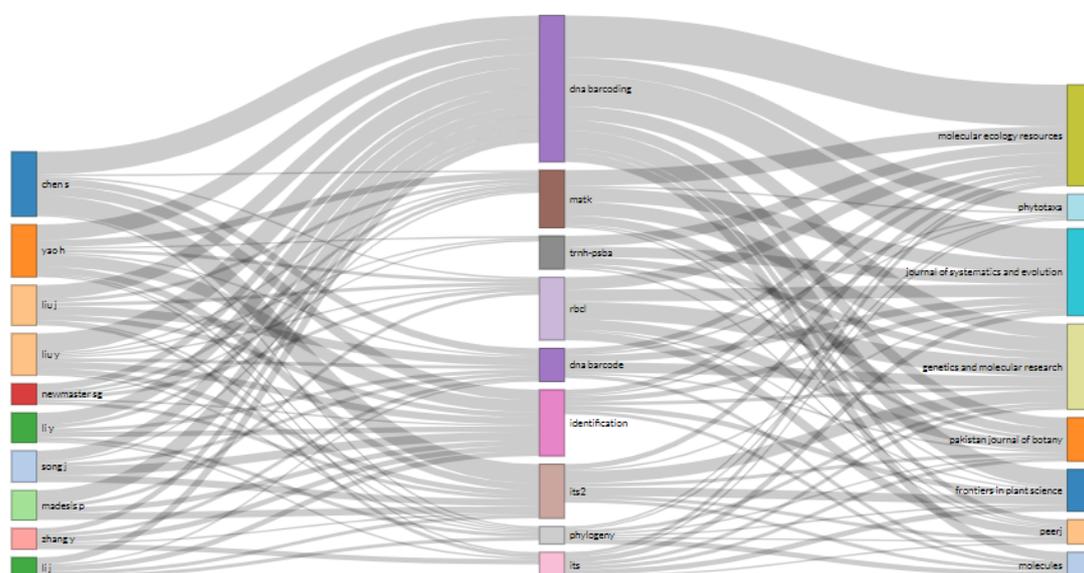
citações por documento, com somente um documento sobre a temática publicado (Figura 6B).

O primeiro trabalho publicado e o mais citado globalmente sobre a aplicação de DNA *barcoding* no estudo de plantas foi produzido por Kress et al. (2005). No artigo, o autor sugere regiões alternativas (*ITS* e *trnH-psbA*) ao gene *COI* para o estudo em plantas, visto que essa região mitocondrial não é apropriada para a maioria das espécies de plantas por apresentar estrutura instável, baixa taxa de substituição de nucleotídeos e divergências das regiões codificantes para este gene entre famílias de plantas com flores (angiospermas). Este trabalho tem grande destaque por ser pioneiro na busca de novas regiões para serem utilizadas como *barcode*, considerando as possibilidades encontradas nos genomas nuclear e plastidial nas plantas, o que justifica sua citação como referência global para o tema.



**Figura 6.** Estudos encontrados utilizando a técnica de DNA *barcoding* em plantas, considerando uma janela de tempo de 16 anos (2005-2020). **A)** Produção anual de documentos, considerando a janela de tempo de 2005-2020. **B)** Média global de citações do conjunto de documentos produzidos dentro de cada ano, considerando a janela de tempo de 2005-2020.

A organização do conhecimento científico foi resumida em um gráfico de três campos no qual é possível visualizar as relações entre os 10 principais autores, as 10 palavras-chave mais utilizadas e os 10 periódicos mais relevantes (Figura 7). A partir desse gráfico, é possível constatar que a produção científica se concentra na utilização das regiões mais recorrentes como marcadores e seu papel na identificação de espécies de interesse, visto que “*matK*”, “*trnH-psbA*”, *rbcL*”, “*ITS2*” estão entre as palavras-chave mais utilizadas em periódicos que se destacam por serem generalistas. Diante destas contatações é possível sugerir que a técnica e DNA *barcoding* em plantas vem sendo utilizada para identificação de espécies de interesse para diversos fins (usos e/ou aplicações).



**Figura 7.** Gráfico de três campos apresentando a relação das 10 palavras-chave mais utilizadas pelos 10 autores que mais publicaram e os respectivos 10 periódicos de maior relevância para produções na área de DNA *barcoding* em plantas.

Os autores mais relevantes, com base no número de publicações com seus respectivos índices *h*, foram os chineses: Shilin Chen ( $n = 37$ ;  $h = 18$ ); Jingyuan Song ( $n = 25$ ;  $h = 18$ ) Hui Yao ( $n = 25$ ;  $h = 13$ ). Esses autores apresentam como características em comum, serem colaboradores entre si e pertencerem, ao mesmo país e a mesma instituição de pesquisa a *Chinese Academy of Sciences/Peking Union Medical College*. As instituições de afiliação mais relevantes também são as chinesas, *Peking Union Medical College* e *Chinese Academy of Sciences*, e juntas lideram a medicina moderna no país há mais de um século, produzindo educação e pesquisa de impacto mundial,

assumindo a responsabilidade de assessorar o governo chinês nas reformas essenciais da saúde e da educação médica (Tabela 3) (Fonte: <http://english.cams.cn/index.html>).

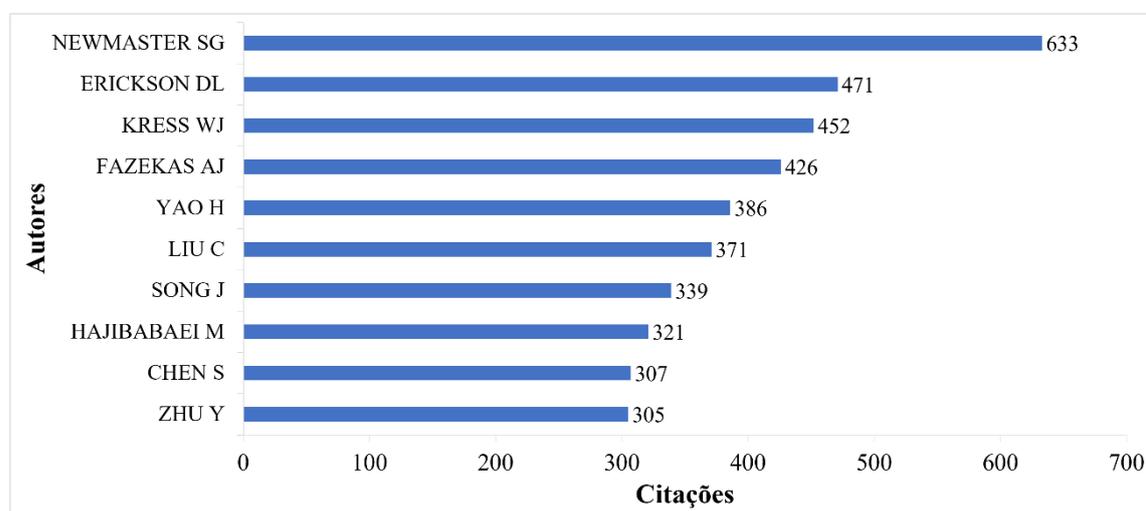
O artigo de Chen et al. (2010) é um exemplo de estudo realizado em colaboração entre os três autores mais relevantes já citados anteriormente (Tabela 3). Nele, são comparadas sete regiões candidatas como DNA *barcode* em espécies de plantas medicinais (*psbA-trnH*, *matK*, *rbcL*, *rpoC1*, *ycf5*, *ITS2* e *ITS*). Foram incluídos como critérios de classificação: a eficiência na amplificação, divergências intra e interespecíficas e as lacunas no código de barras. A região *ITS2* representou a maior adequação para ser utilizada como códigos de barras, sendo sua capacidade de discriminação testada em mais 6.600 amostras de plantas, pertencentes a 4.800 espécies de 753 gêneros distintos, com uma taxa de identificação bem-sucedida de 92,7% em nível de espécie (Chen et al., 2010).

**Tabela 3.** *Ranking* dos 10 autores que mais publicaram sobre aplicação de DNA *barcoding* em plantas, relacionados aos respectivos: índice *h*, país de origem e instituição de filiação.

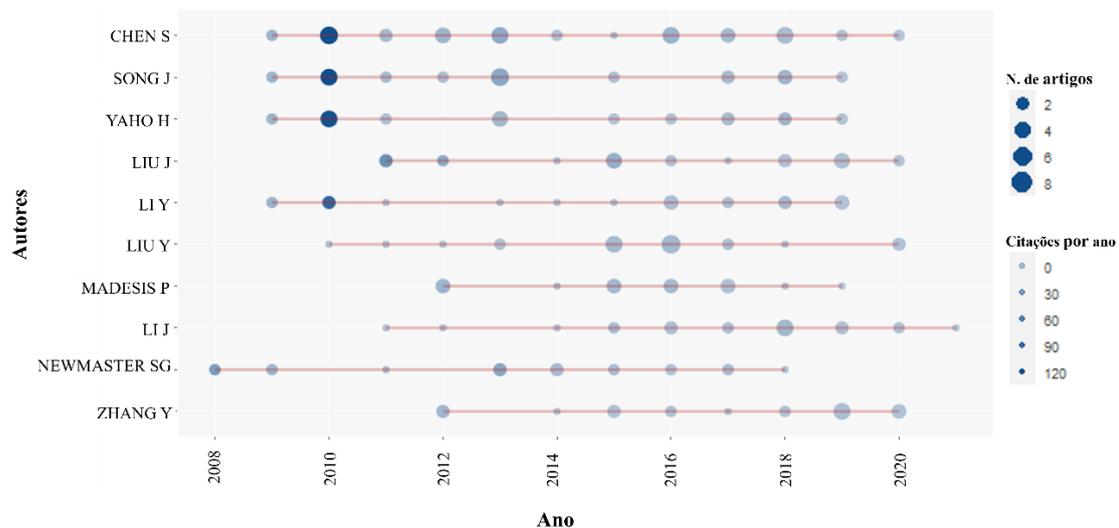
Ranking	Autores	Artigos	Índice <i>h</i>	País	Instituição
1	Chen, SL	37	18	China	Chinese Academy of Sciences/Peking Union Medical College
2	Song, J	25	18	China	
3	Yao, H	25	13	China	Kunming Institute of Botany
4	Liu, J	24	10	China	
5	Liu, Y	21	9	China	Chinese Academy of Sciences/Peking Union Medical College
6	Li, Y	20	10	China	
7	Madesis, P	19	11	Grécia	Aristotle University of Thessaloniki
8	Li, J	18	8	China	Key Laboratory of Bio-Resources and Eco-Environment
9	Newmaster, SG	18	15	Canadá	University of Guelph
10	Zhang, Y	18	6	China	Chengdu University of Traditional Chinese Medicine

Vale ressaltar que os autores mais citados globalmente são Steven G. Newmaster (633 citações); David L. Erickson (471 citações) e Walter Jhon Kress (452 citações) (Figura 8) e, apesar de serem autores de referência, são caracterizados por produções sazonais, não produzindo com certa constância. Newmaster é um botânico canadense especializado em diversidade vegetal, plantas medicinais e taxonomia molecular. Seu laboratório de pesquisa está situado no *Center for Biodiversity Genomics (CBG)* no

*Biodiversity Institute of Ontario*. (Fonte: <https://www.uoguelph.ca/ib/newmaster>). Erickson é especializado em genética e evolução e atualmente lidera o laboratório *DNA4*, fundado em 2015 com o objetivo de fornecer à indústria de alimentos e medicamentos o que há de mais moderno em serviços genômicos e de bioinformática, com foco na identificação de espécies e validação de produtos (Fonte: <https://www.dna4tech.com/about/>). O botânico norte americano Kress é o curador de botânica do Museu Nacional de História Natural de Smithsonian, localizado em Washington DC. Atualmente, seu foco de pesquisa é genômica da biodiversidade e conservação (Fonte: <https://arboretum.harvard.edu/people/w-john-kress/>). Kress produziu trabalhos significativos aplicando a técnica de DNA *barcoding* em plantas, inclusive em parceria com Erickson, e ambos se dedicam a diversos estudos botânicos e colaboram com inúmeros países. De maneira geral, os três autores contribuíram com os primeiros trabalhos relacionados à aplicação da técnica de DNA *barcoding* no estudo de plantas. No entanto, constata-se uma sazonalidade de autoria, ao observar o baixo número de publicações e a distribuição esparsa dessa produção ao longo do tempo, considerando Newmaster, o único autor dentre os três que aparece no *ranking* (Figuras 8 e 9).



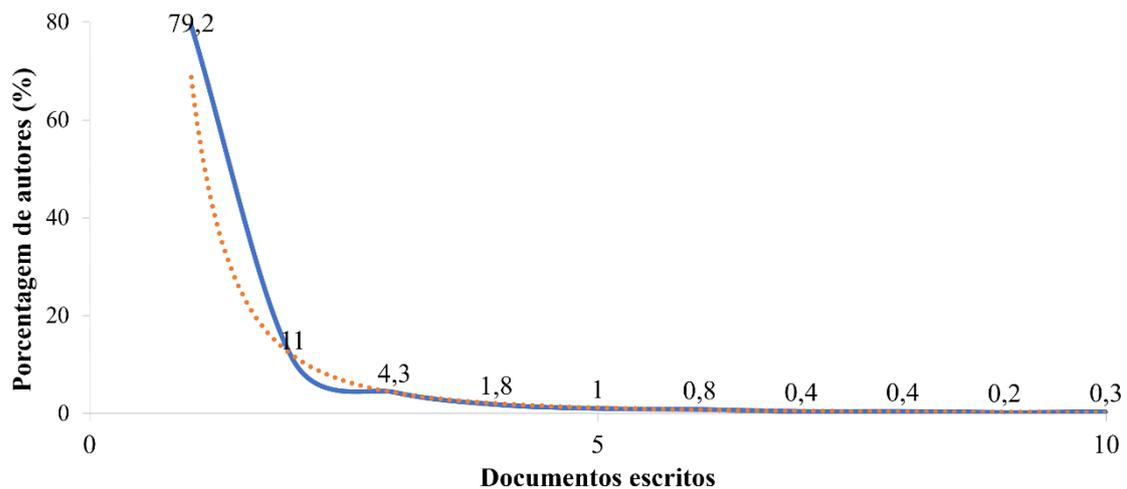
**Figura 8.** *Ranking* dos 10 autores mais citados globalmente, considerando as publicações sobre a técnica de DNA *barcoding* em plantas.



**Figura 9.** Distribuição da produtividade dos autores mais relevantes sobre a aplicação da técnica de DNA *barcoding* em plantas, ao longo da janela de tempo (2005-2020). O tamanho e a intensidade da cor de cada círculo representam a quantidade de publicações e número de citações por ano em determinado período, respectivamente.

Cerca de 79,2% dos autores produziram apenas um artigo relacionado ao tema, enquanto 90% dos autores produziram até dois artigos (Figura 10). Esses dados validam as médias descritas na Tabela 2, uma vez que a média de documentos por autor (0,29) é menor que a média de autores por documento (3,49). A média de autores por documento também explica o número de coautores por documento (5,34) e o baixo número de documentos com autoria única (16). A produtividade dos autores se adequa à Lei de Lotka, uma vez que o conjunto de dados (linha azul) apresenta aderência quando considerado o esperado para a Lei de Lotka (linha laranja pontilhada), sugerindo que a maior parte do conhecimento científico sobre a aplicação de DNA *barcoding* em estudo de plantas é produzida por um pequeno número de autores, visto que a maioria destes participa de estudos pontuais envolvendo o tema.

Shilin Chen, Jingyuan Song e Hui Yao, respectivamente os três autores de maior relevância, segundo o índice *h*, mantiveram um nível de produtividade contínuo ao longo do tempo (Figura 9), enquanto os demais autores apresentaram diferentes níveis de sazonalidade, o que reforça o constatado pela Lei de Lotka, uma vez que a maior parte do conhecimento científico, referente à aplicação de DNA *barcoding* em plantas, está concentrada em determinados autores. Tais observações nos levam à hipótese de uma tendência de fortificação da comunidade científica em relação a essa temática, considerando os últimos 10 anos, com destaque para a China.

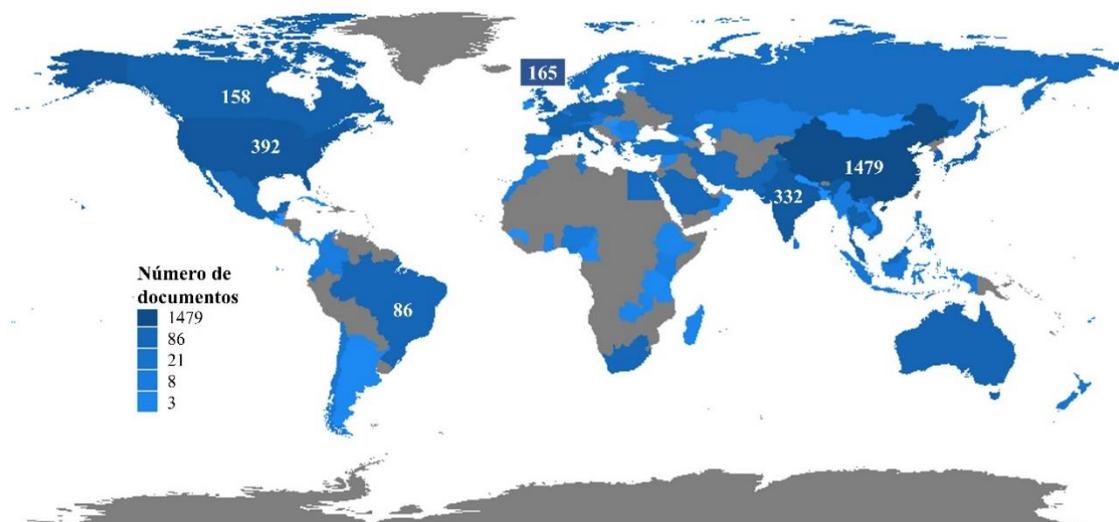


**Figura 10.** Representação da Lei de Lotka envolvendo a produtividade dos autores com relação a pesquisas sobre aplicação de DNA *barcoding* em plantas. A linha pontilhada laranja representa o modelo esperado para a Lei de Lotka enquanto a linha azul demonstra o comportamento do conjunto de dados considerado.

Os países que apresentaram a maior frequência de artigos publicados foram China (1.479), EUA (392) e Índia (322) (Figura 11). Os países que apresentaram maior relevância em número de citações foram China (6.924), EUA (4.271) e Canadá (1.788) (Tabela 4). Durante a Terceira Conferência Internacional para o *Barcode of Life*, o *CBOL PWG* convocou a comunidade internacional de pesquisadores envolvidos nesse processo, para que, em um esforço coletivo, conseguisse avaliar melhor os códigos de barras propostos até aquele momento, dentro dos 18 meses subsequentes, a fim de encontrar um *barcode* padrão que fosse consenso para todos os grupos de plantas (CBOL, 2011).

Em resposta a esta solicitação, a China mobilizou os principais grupos de pesquisa do país para que pudessem colaborar, utilizando como fonte de estudo a sua própria megadiversidade botânica. O país abriga, em seu território, quatro dos 34 *hotspots* globais de biodiversidade reconhecidos (montanhas da Ásia Central, os Himalaias, a região Indo-Miamar e as montanhas do sudoeste do país), além de ser o centro de distribuição de muitos gêneros endêmicos, de regiões de clima temperado, e abrigar florestas únicas (Li, 2008; Raven, 2011). O projeto chinês envolveu 46 grupos de pesquisa de 17 instituições e Universidade do país, todos com longa experiência em taxonomia e extensas coleções de materiais vegetais (CBOL, 2011). Todo o esforço coordenado pela China é de grande importância global e reflete nos parâmetros de

produção científica analisados, se destacando pelo número de autores e instituições mais relevantes, por exemplo (Tabela 4). O Canadá aparece como um dos países mais citados (Tabela 4), visto que Paul Hebert, pesquisador filiado à Universidade de Guelph, publicou o trabalho pioneiro de utilização da técnica de DNA *barcoding* na expectativa de criar uma metodologia de identificação sustentável, como complementar ao diagnóstico de espécies da taxonomia tradicional. Hebert et al. (2003) estabeleceram o gene mitocondrial do citocromo oxidase I (*COI*), como núcleo de um sistema de bioidentificação global para animais. Desde então, o uso da técnica não se restringe apenas ao estudo de animais e não apenas como alternativa à taxonomia clássica, sendo também aplicada nas mais diferentes áreas como Sistemática Molecular, Genética Forense e Ecologia (Hollingsworth et al., 2011; Kress et al., 2017).

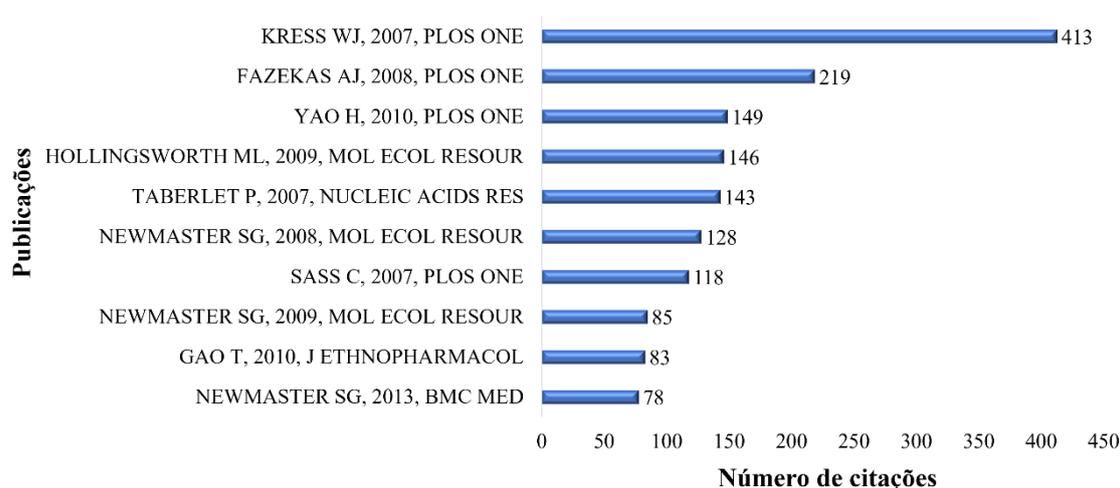


**Figura 11.** Frequência da produção científica mundial relacionada a aplicação da técnica de DNA *barcoding* em plantas com destaque para China (1.479 artigos publicados), Estados Unidos (392 artigos publicados) e Índia 332 (artigos publicados).

**Tabela 4.** *Ranking* dos países com maior número de publicações e que apresentaram maior relevância quanto ao número de citações sobre a aplicação de DNA *barcoding* em plantas.

<i>Ranking</i>	País	Frequência	<i>Ranking</i>	País	Total de citações
1	CHINA	1.479	1	CHINA	6.924
2	EUA	392	2	EUA	4.271
3	ÍNDIA	322	3	CANADÁ	1.788
4	REINO UNIDO	165	4	REINO UNIDO	1.492
5	CANADÁ	158	5	FRANÇA	1.456
6	TAILÂNDIA	149	6	ÍNDIA	872
7	COREIA DO SUL	130	7	ITÁLIA	709
8	ITÁLIA	125	8	ALEMANHA	530
9	ALEMANHA	117	9	AUSTRÁLIA	416
10	FRANÇA	115	10	HOLANDA	410

A partir do *ranking* de 10 documentos mais citados localmente (dentro do conjunto de dados), observa-se um número variável de citações, apresentando um mínimo de 78 e máximo de 413 citações, dentro do período considerado (Figura 12). Os artigos mais citados são, em geral, os que mais contribuíram para trabalhos posteriores, pois são alguns dos estudos considerados pioneiros, oferecendo embasamento para os posteriores. Os três primeiros trabalhos apresentam propostas gerais de regiões do genoma plastidial e nuclear que demonstraram grande potencial para identificação e delimitação de espécies (Figura 12).



**Figura 12.** *Ranking* dos 10 trabalhos mais citados dentro do conjunto de dados sobre aplicação de DNA *barcoding* em plantas, considerando a janela de tempo de 2005-2020.

O trabalho mais citado dentro do conjunto de dados considerado para este estudo foi o de Kress et al. (2007), que testa um sistema de código de barras global (respectivamente: oito regiões plastidiais *trnH-psbA*, *rbcL-a*, *rpoC1*, *rpoB2*, *accD*, *ycf5*, *ndhJ*, *matK* e uma nuclear *ITS1*), avaliando a sua universalidade e grau de divergência de sequência, em pares e em conjunto em uma amostra filogeneticamente diversa de 48 gêneros. Foi constatado que nenhum loco único poderia discriminar entre as espécies, mas a porcentagem de discriminação aumentou quando se associou *trnH-psbA* à região codificadora *rbcL-a*. A região *rbcL-a* fornece uma forte âncora de reconhecimento colocando um espécime não identificado dentro de uma família, gênero e às vezes espécie, enquanto *trnH-psbA* estreita ainda mais a identificação correta, especialmente em gêneros de angiospermas ricos em espécies (Kress et al., 2007).

O segundo trabalho mais citado foi o de Fazekas et al. (2008), que compara oito regiões plastidiais (*rpoB*, *rpoC1*, *rbcL*, *matK*, *23S*, *trnH-psbA*, *atpF-atpH* e *psbK-psbI*) e uma região mitocondrial, testando o poder de discriminar a monofilia de 92 espécies em 32 gêneros diversos de plantas terrestres. Este trabalho destaca a importância de se trazer maior atenção às questões práticas relacionadas à facilidade de recuperação de sequência, facilidade e confiabilidade no alinhamento global e redundância de marcadores em sistemas de códigos de barras multilocos para plantas (Fazekas et al., 2008). O terceiro trabalho mais citado, de Yao et al. (2010), caracteriza a região *ITS2* do DNA ribossômico nuclear como uma grande candidata, uma vez que possui uma série de características valiosas, como a disponibilidade de regiões conservadas para projetar *primers* universais, além da facilidade de amplificação e variabilidade suficientes para distinguir até mesmo espécies estreitamente relacionadas. Neste trabalho, foram observadas taxas de sucesso de identificação para dicotiledôneas, monocotiledôneas, gimnospermas, samambaias e musgos, de: 76,1%, 74,2%, 67,1%, 88,1% e 77,4%, respectivamente (Yao et al., 2010).

Os periódicos que apresentaram maior destaque em relação ao número de artigos publicados, com os respectivos valores de índice *h*, foram: *PLOS ONE* ( $n = 103$ ;  $h = 38$ ), *Molecular Ecology Resources* ( $n = 52$ ;  $h = 30$ ), *Scientific Reports* e *Journal of Systematics and Evolution* (com  $n = 32$  e  $n = 24$ , respectivamente e  $h = 14$  para ambos). Além de ser a revista que mais publicou trabalhos relacionados ao assunto e a que apresenta maior índice *h*, a *Plos One* também é o periódico mais citado, com 3.653 citações (Tabela 5).

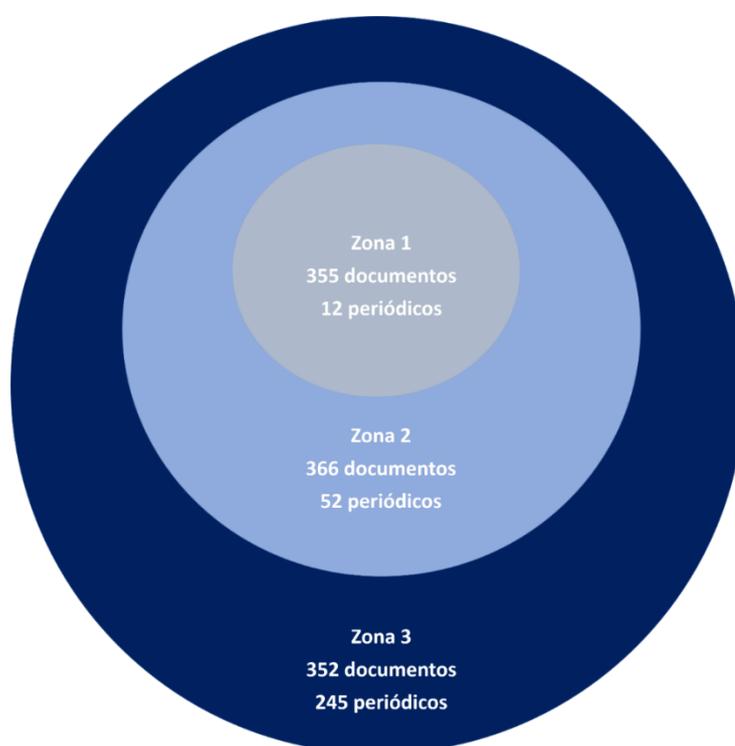
O periódico *Plos One* se destacou dentre os demais, publicando praticamente o dobro de documentos da segunda colocada *Molecular Ecology Resources* (Tabela 5). Trata-se de uma revista científica norte americana, multidisciplinar e de acesso aberto, que tem publicado artigos desde 2006 pela editora *PLoS (Public Library of Science)*, exclusivamente online e que cobre principalmente pesquisas básicas nas áreas de Ciências e Medicina (Fonte: <https://plos.org/about/>). Vale salientar que alguns desses periódicos, como a *Plos One*, *Scientific Reports* e *PEERJ*, podem ser classificados como generalistas visto que não apresentam uma área de interesse específica de publicação, o que sugere que o tema sobre aplicação de DNA *barcoding* em plantas é de interesse geral da comunidade científica e não apenas de grupos específicos de pesquisa.

**Tabela 5.** *Ranking* dos 10 periódicos de maior destaque em relação ao número de documentos publicados, número de citações e índice *h*.

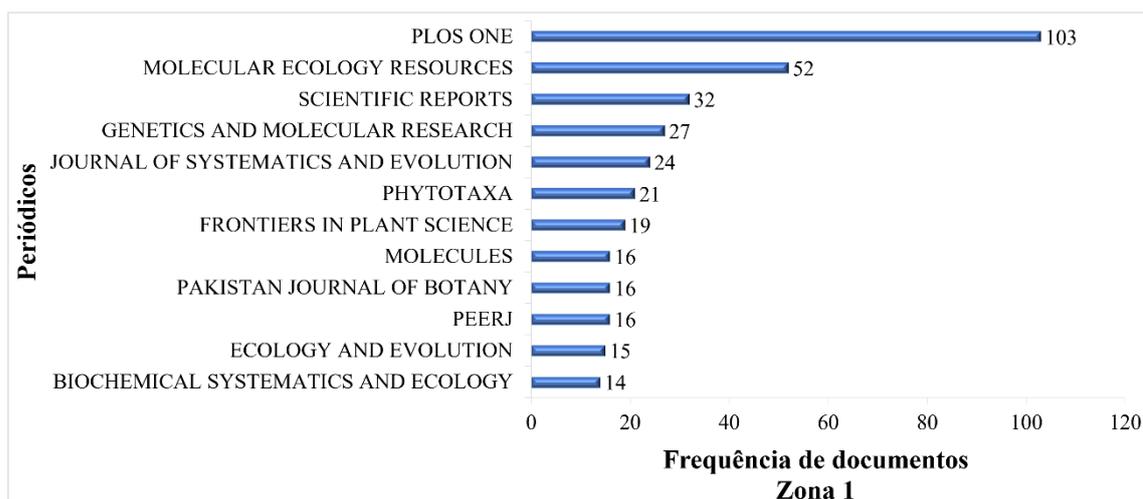
<i>Ranking</i>	Periódicos	Número de publicações ( <i>n</i> )	Número de citações	Índice <i>h</i>
1	<i>Plos One</i>	103	3.653	38
2	<i>Molecular Ecology Resources</i>	52	1.765	30
3	<i>Scientific Reports</i>	32	475	14
4	<i>Genetics and Molecular Research</i>	27	135	9
5	<i>Journal of Systematics and Evolution</i>	24	435	14
6	<i>Phytotaxa</i>	21	98	4
7	<i>Frontiers in Plant Science</i>	19	275	11
8	<i>Molecules</i>	16	116	8
9	<i>Pakistan Journal of Botany</i>	16	87	5
10	<i>PeerJ</i>	16	55	4

A Lei de Bradford prevê a divisão do número total de artigos que compõem o conjunto de dados considerado, em três zonas com tamanhos equivalentes. Essas zonas, por sua vez, correspondem aos periódicos que abrigam os artigos segundo o assunto de interesse. Como o número total de artigos do presente estudo (1.073) não retorna números inteiros quando dividido por três, as zonas não contemplaram o mesmo número de artigos. A partir dessa Lei, os artigos foram organizados nas seguintes zonas: Zona 1, com 355 documentos, divididos em 12 periódicos; Zona 2, com 366 documentos, distribuídos em 52 periódicos e, por fim; a Zona 3, que engloba 352 documentos, distribuídos em 245 periódicos diferentes (Figura 13).

A Zona 1 apresenta uma maior quantidade de documentos concentrados em uma pequena quantidade de periódicos, destacando-se: *Plos One*, *Molecular Ecology Resources*, *Scientific Reports*, revistas que mais produziram e apresentam índice *h* mais relevante (Figura 14). A Zona 2 apresenta aumento, tanto na frequência de documentos produzidos, quanto no número de revistas, enquanto a Zona 3 apresenta uma queda na frequência de documentos, porém um aumento significativo de periódicos. O aumento do número de periódicos que produzem documentos relacionados com o tema demonstra que áreas de interesse diferentes estão aplicando a técnica de DNA *barcoding* em plantas para os mais diferentes fins (Figura 14).



**Figura 13.** Representação esquemática da Lei de Bradford em que periódicos são agrupados por zonas, que condensam maior número de documentos especializados considerando a temática estudada.



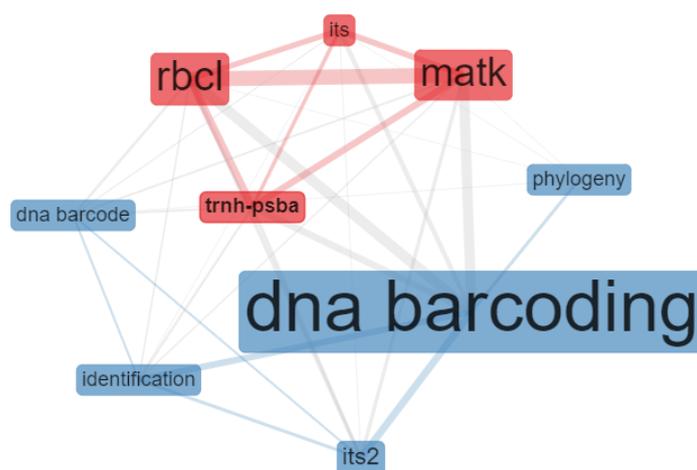
**Figura 14.** Lista de periódicos pertencentes à Zona 1 segundo a Lei de Bradford, que apresentaram a maior número de publicações sobre a temática estudada.

Diferentemente da aplicação da técnica em animais, que conseguiu sucesso utilizando o gene *COI* para a grande maioria dos grupos taxonômicos, nas plantas não existe um marcador consenso único. Desse modo, normalmente são utilizados múltiplos marcadores para oferecerem as respostas desejadas para cada tipo de estudo (CBOL, 2009; CBOL, 2011). Tal fato justifica a citação de mais de um marcador como palavra-chave pelos autores (Tabela 6).

**Tabela 6.** Ranking das 10 palavras-chave mais usadas pelos autores em publicações relacionadas a aplicação da técnica de DNA *barcoding* no estudo de plantas.

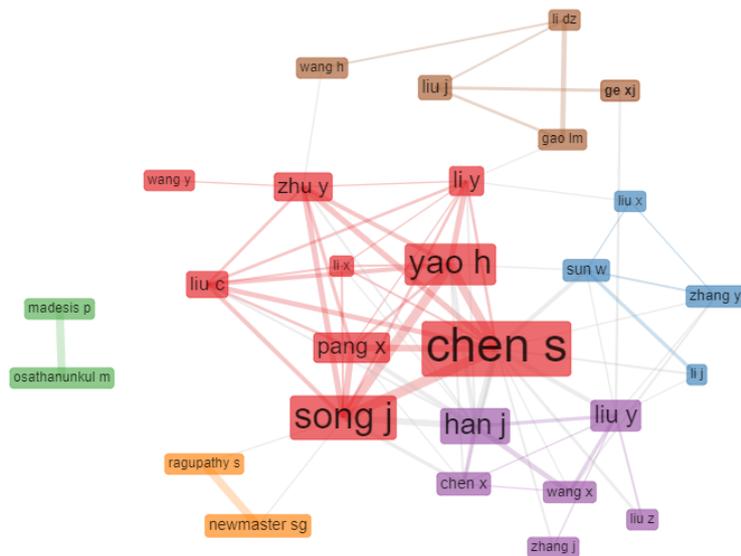
Ranking	Palavras-chave	Ocorrência
1	<i>DNA barcoding</i>	336
2	<i>rbcL</i>	128
3	<i>matK</i>	121
4	<i>DNA barcode</i>	84
5	<i>ITS2</i>	80
6	<i>Identification</i>	69
7	<i>ITS</i>	67
8	<i>trnH-psbA</i>	59
9	<i>Phylogeny</i>	54
10	<i>Species identification</i>	53

As palavras-chave podem estar presentes em mais de um documento, relacionando-se entre si (Figura 15). O tamanho das palavras demonstra aquelas que foram mais usadas e as linhas representam aquelas que estão relacionadas em um mesmo documento. Pode-se observar que o nome das principais regiões utilizadas como marcadores aparecem juntas nos mesmos documentos com destaque para: *rbcL* e *matK*, regiões consideradas centrais na aplicação da técnica de DNA *barcoding* no estudo de plantas (Figura 15) (CBOL, 2009).



**Figura 15.** Rede de coocorrência das palavras-chave empregadas pelos autores em artigos relacionados com a aplicação de DNA *barcoding* em plantas.

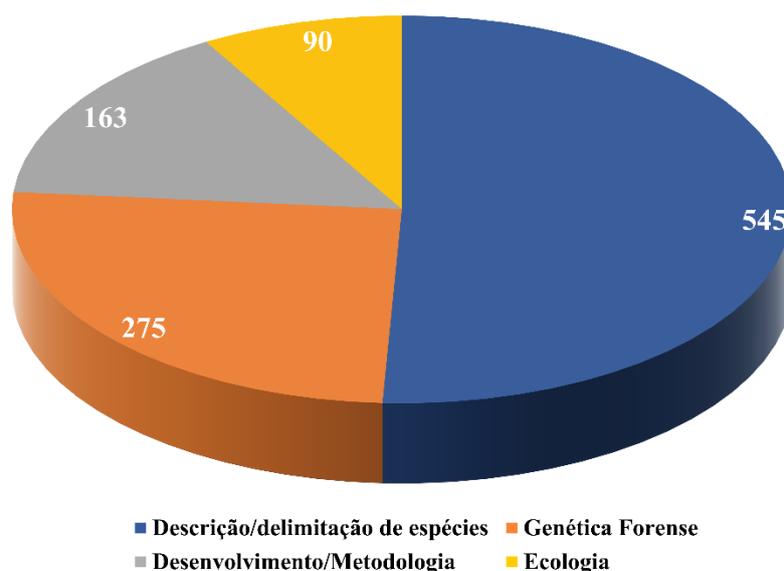
Na rede de colaboração entre os 30 principais autores pode-se observar uma rede maior que agrupa cinco *clusters* representados em marrom, vermelho, azul, lilás e laranja representando autores de diversas nacionalidades e um *cluster* verde isolado, formado por autores europeus (Figura 16). Observa-se que a colaboração entre os autores asiáticos é intensa, principalmente a partir de 2010, ano em que a produção de documentos na China se torna significativa ao longo do tempo (como observado na Figura 9). Ligados a essa rede maior formada por autores asiáticos (*clusters* marrom, vermelho, azul, lilás), aparecem autores canadenses (em laranja) filiados à mesma instituição (Universidade de Guelph), que publicam há bastante tempo sobre o tema e colaboram, de forma contínua, tanto entre si como em trabalhos com os principais pesquisadores da China, com uma pausa a partir de 2018 (Figura 9).



**Figura 16.** Rede de colaboração entre os 30 principais autores de diferentes nacionalidades quanto a produção relativa à aplicação de DNA *barcoding* no estudo de plantas.

## 5.2 Categorização

Após a leitura dos 1.073 resumos, os documentos foram categorizados em cinco grupos quanto ao uso e/ou aplicação da técnica. As áreas com maior produção foram, respectivamente: Descrição/delimitação de espécies (545 artigos), Genética Forense (275), Desenvolvimento/metodologia (163), Ecologia (90), sendo que 75 artigos foram categorizados em até dois grupos diferentes (Figura 17).



**Figura 17.** Usos e/ou aplicações atribuídos aos 1.073 artigos, encontrados na *Web of Science*, relacionados a estudos com DNA *barcoding* em plantas.

Embora inicialmente proposta como uma técnica “alternativa” à taxonomia clássica, principalmente no que tange à descrição de espécies ainda não reconhecidas pela Ciência (Hebert et al., 2003; Hebert et al., 2005), o que se observa ao longo do tempo é a aplicação de DNA *barcoding* no estudo de plantas para resolução de questões biológicas cada vez mais específicas (Kress, 2017). Os trabalhos agrupados na categoria de maior destaque (Descrição/delimitação de espécies) abordaram a delimitação de espécies de grupos complexos através da associação de dados morfológicos integrados à análise de dados moleculares, bem como a certificação de espécies em coleções de herbários e de espécies com propriedades bioquímicas de interesse que precisam passar por confirmação taxonômica e, por fim, a realização de inventários taxonômicos em áreas de preservação, por exemplo.

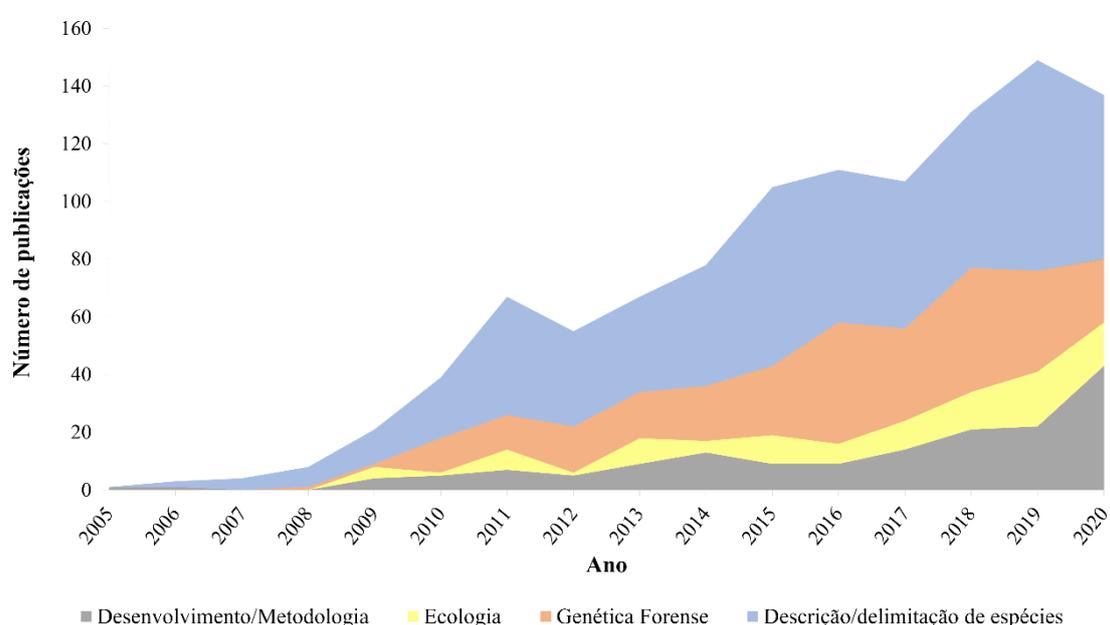
Trabalhos utilizando DNA *barcoding* para fins forenses ocuparam a segunda posição de destaque (Figura 17), e tem como foco biossegurança e saúde pública a partir do monitoramento de produtos comerciais, como a autenticação de fitoterápicos, plantas de uso medicinal e alimentício, amplamente utilizadas na cultura de países asiáticos como China (Alter, 2005), principalmente como tratamento médico alternativo. Nesta categoria, também se destacam os trabalhos que envolvem o monitoramento do tráfico ilegal de espécies ameaçadas de extinção e identificação de espécies invasoras.

Na categoria Desenvolvimento/metodologia foram inseridos trabalhos de desenvolvimento de bibliotecas de sequências de referência para espécies com algum interesse (econômico, ecológico, etc.), que ainda não possuem informações de sequenciamento disponíveis, trabalhos em que são realizados testes de amplificação de regiões já disponíveis, protocolos de identificação de material processado e sequenciamento de genomas plastidiais completos para desenvolvimento de marcadores *barcode* ou serem utilizados como *super barcode*.

A categoria Ecologia agrupa trabalhos em que a técnica é utilizada para acessar informações sobre redes de interação entre espécies, através da análise da dieta de herbívoros e insetos polinizadores, utilizando, em muitos casos, marcadores para *metabarcoding*, além de trabalhos com foco em montagem de espécies em comunidades vegetais e determinação de habitats evolutivamente mais diversificados para conservação.

A necessidade de se obter informações genéticas para diversos fins, o avanço da tecnologia e outras demandas impulsionaram o desenvolvimento de novas metodologias

de sequenciamento com melhor qualidade, menor custo, maior rapidez e maior capacidade de geração de informações, contribuindo para que a identificação de espécies a partir de fragmentos de DNA e/ou genomas completos se tornasse possível (Goodwin et al., 2016). A partir de 2011, observa-se um aumento significativo na produção científica (Figura 18), que coincide com incentivos promovidos pelo CBOL (2011), para estabelecimento de regiões padrão para o estudo de plantas, e com a evolução das tecnologias de Sequenciamento de Alto Rendimento implementadas no início dos anos 2000, a partir do desenvolvimento do Projeto Genoma Humano (Yang et al., 2020).

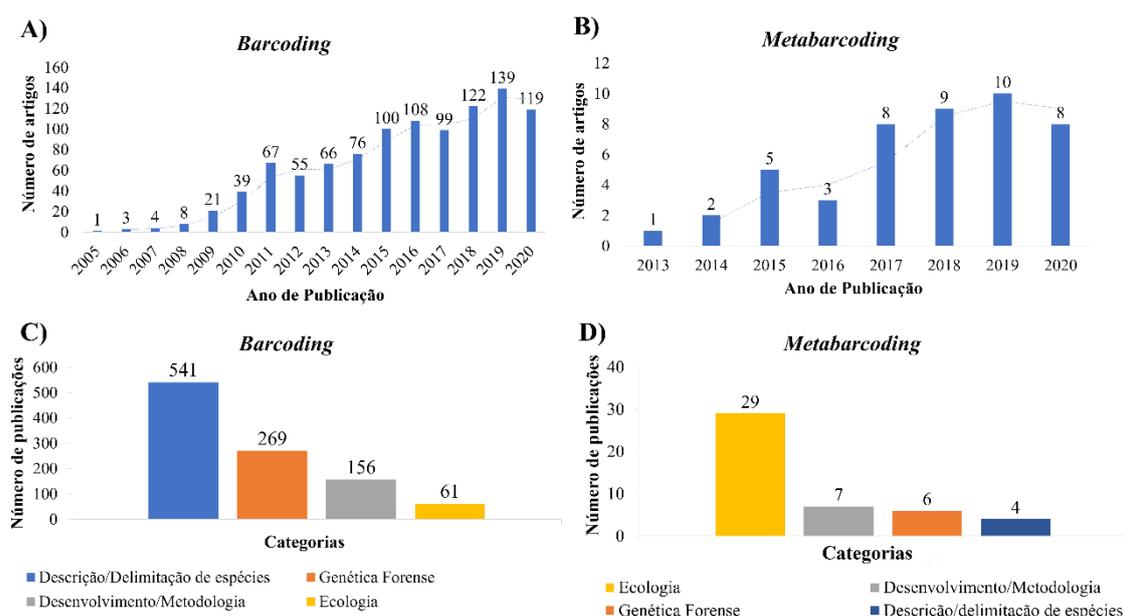


**Figura 18.** Magnitude da produção científica por categoria de uso e/ou aplicação da técnica de DNA *barcoding* no estudo de plantas ao longo da janela de tempo de 2005-2020.

O uso de marcadores baseados em DNA *barcoding*, para responder mais diversas questões biológicas no estudo de plantas, é crescente e está presente em todos os anos do período avaliado, com pico em 2019, contendo 139 artigos (Figura 19A). Já a utilização de marcadores baseados em *metabarcoding* é mais recente, para o estudo de plantas, sendo a primeira publicação realizada em 2013 (Figura 19B).

Vale destacar que o primeiro trabalho sobre aplicação de *metabarcoding* no estudo de plantas buscou compreender como a fragmentação de habitats influencia na plasticidade dietética de uma espécie ameaçada de lêmure em Madagascar, utilizando

amostras ambientais de fezes de 96 indivíduos e o marcador *trnL* (Quéméré et al., 2013). Paralelo a este estudo, foi construído um grande banco de dados de referência de espécies vegetais da região. Além disto, este estudo também destaca o principal uso/aplicação de *metabarcoding* na categoria Ecologia (estabelecimento de redes de interação entre espécies), como pode ser observado na Figura 19D, enquanto a categoria Descrição/delimitação de espécies é preponderante entre as categorias de aplicação de DNA *barcoding* (Figura 19C).



**Figura 19.** Número de publicações de estudos em plantas, utilizando as técnicas de: **A)** DNA *barcoding* e **B)** *metabarcoding*, ao longo (2005-2020). **C)** Categorização dos estudos do banco de dados, considerando o uso e/ou aplicação de DNA *barcoding* no estudo de plantas. **D)** Categorização dos trabalhos do banco de dados, considerando o uso e/ou aplicação de *metabarcoding* no estudo de plantas.

Apesar da produção global sobre o estabelecimento de um código de barras central para o estudo de plantas, bem como sobre a diversificação de seus usos/aplicações apresentarem um *continuum* ao longo do tempo, ainda existem muitos aspectos a serem considerados. A consolidação de uma biblioteca global efetiva de DNA *barcode* de plantas e a adoção de tecnologias de sequenciamento genômico para um fluxo de trabalho mais eficiente e econômico, que viabilize a aplicação desses marcadores de identificação genética em campos adicionais de empreendimentos biológicos e comerciais, ainda são necessárias.

## 6. CONCLUSÃO

Neste estudo, foi possível descrever as tendências do conhecimento com relação à aplicação da técnica de DNA *barcoding* no estudo das plantas. Constatou-se que, mesmo o *Barcode of Life* sendo pensado desde o início como um grande projeto de colaboração mundial para gerar dados para identificação de espécies, a China é o país que mais vem se destacando tanto em produtividade quanto em incentivo às suas organizações científicas de forma contínua.

De fato, pôde-se constatar ao longo deste trabalho, que a rede de colaboração estabelecida entre os autores de destaque, em sua maioria chineses, é praticamente restrita ao país ou a poucos países asiáticos, com foco em espécies de interesse social e econômico nacional. Muito provavelmente este fato pode estar relacionado a falta de liderança do país em projetos internacionais pioneiros. Ainda considerando a China como país de destaque, a maior parte dos trabalhos se concentraram na verificação da autenticidade das espécies utilizadas na medicina tradicional chinesa, enquanto outras aplicações deram mais atenção à segurança alimentar, inspeção e quarentena e controle de pragas e espécies invasoras.

Quanto à categorização, foi possível verificar que a identificação de espécies apresenta uma grande gama de aplicações, o que também pode ser confirmado pelo tipo de periódicos predominantes. Estes podem ser classificados como generalistas, visto que não apresentam uma área de interesse específica de publicação, o que sugere que o tema sobre aplicação de DNA *barcoding* em plantas é de interesse geral da comunidade científica e não apenas de grupos específicos de pesquisa.

As possibilidades de aplicação vêm sendo ampliadas graças ao desenvolvimento das tecnologias de Sequenciamento de Alto Rendimento, abrindo portas para trabalhos utilizando *metabarcoding* e conseqüentemente para aplicações voltadas para estudos em Ecologia, Evolução e Genética Forense.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ALTER, J.S. **Asian medicine na globalization**. Philadelphia: University of Pennsylvania Press, 2005, 187 p.
- ALVARADO, R.U. A Lei de Lotka na bibliometria brasileira. **Ciência da Informação**, v.31, n.2, 2002.
- ARIA, M.; CUCCURULLO, C. *Bibliometrix*: An R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, v. 11, p. 959-975, 2017.
- BANSCH, S.; TSCHARNTKE, T.; WUNSCHIERS, R.; NETTER, L.; BRENIG, B.; GABRIEL, D.; WESTPHAL, C. Using ITS2 metabarcoding and microscopy to analyse shifts in pollen diets of honey bees and humble bees along a mass-flowering crop gradiente. **Molecular Ecology**, n. 29, p. 5003-5018, 2020.
- BICUDO, C. de M. Taxonomia. **Biota Neotropica**, v. 4, n. 1, 2004.
- BRADFORD, S.C. **Documentação**. Rio de Janeiro, Fundo de Cultura, p. 196-216, 1961.
- BUFREM, L.; PRATES, Y. O saber científico e as práticas de mensuração da informação. **Ciência da Informação**, v. 34, n. 2, p. 9-25, 2005.
- CAPES. **Portal de Periódicos da CAPES**. Disponível em: <<http://www.periodicos.capes.gov.br>>. Acessado em: 06 de Fev. 2021.
- CBOL Plant Working Group. A DNA barcode for land plants. **PNAS**, v. 106, n. 31, 2009.
- CBOL Plant Working Group. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. **PNAS**, v. 108, n. 49, p. 19641-19646, 2011.
- CHEN S.; YAO, H.; HÁ, J.; LIU, C.; SONG, J.; SHI, L.; ZHU, Y.; MA, X.; GAO, T.; PANG, X.; LUO, K.; LI, X.; JIA, X.; LIN, Y.; LEON, C. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. **Plos One**, v. 5, 2010.
- COISSAC, E.; HOLLINGSWORTH, P. M.; LAVERGNE, S.; TABERLET, P. From barcodes to genomes: extending the concept of DNA barcoding. **Molecular Ecology**, v. 25, p. 1423-1428, 2016.
- COLLEVATTI, R. G. Nós realmente precisamos de DNA Barcoding? **Natureza e Conservação**, v. 9 n. 1, p. 105-110, 2011.
- COSTION, C.; FORD, A.; CROSS, H.; CRAYN, D.; HARRINGTON, M.; LOWW, A. Plant DNA barcodes can accurately estimate species richness in poorly know flora. **Plos One**, v. 6, 2011.
- CREER, S.; DEINER, K.; FREY, S.; PORAZINSKA, L.; TABERLET, P.; THOMAS W. K.; POTTER C.; BIK, H. M. The ecologist's field guide to sequence-based identification of biodiversity. **Methods in Ecology and Evolution**, v. 7, 2016.
- DIJK, E.L.; AUGER, H.; JASZCZYSZYN, Y.; THERMES C. Ten Years of next-generation sequencing technology. **Trends in Genetics**, 2014.

DORMONTT, E. E.; DIJK, K.; BELL, K. L.; BIFFIN, E.; BREED, M. F.; BYRNE M.; RETALIC, S. C.; VISO, F. E.; NEVILL, P. G.; SHAPCOTT, A.; YOUNG, J. M.; WAYCOTT, M.; LOWE, A. Advancing DNA barcoding and metabarcoding applications for plants requires systematic analysis of herbarium collections – An Australian perspective. **Frontiers in Ecology and Evolution**, v. 6, 2018.

ERICKSON, S. L. DNA barcoding in land plants: developing standards to quantify and maximize success. **TAXON**, v. 57, n. 4, p. 1304-1316, 2008.

FAZEKAS, A. J.; BURGUESS K. S.; KESANAKURTI, P. R.; GRAHAM, S. W.; NEWMASTER, S.G.; BRIAN C. H.; PERCY, D. M.; HAJIBABAEI, M.; BARRETT, S. C. H. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. **Plos One**, v. 3, n. 7, 2008.

FAJARNINGSIH, N.D. Internal Transcribed spaces (ITS) as DNA barcoding to identify fungal species: a review. **Squalen Bulletin of Marine Fisheries Postharvest and Biotechnology**, v. 11, n.2. p. 34-44, 2016.

GOODWIN, S.; MCPHERSON, J.D.; MCCOMBIE W.R. Coming of age: ten Years of next-generation sequencing technologies. **Nature Reviews Genetics**, v.17, 2016.

HEBERT, P. D. N.; CYWINSKA, A.; BALL, S. L., DEWAARD, J. R. BIOLOGICAL identifications through DNA barcodes. **Proceedings of the The Royal Society**, v. 270, p. 313.-321, 2003.

HEBERT, P.D.N.; GREGORY, T.R. The promise of DNA Barcoding for taxonomy. **Systematic Biology**, v. 54, n. 5, p. 852-859, 2005.

HECKENHAUER, J.; SALIM, K. A.; CHASE, M. W.; DEXTER, K. G.; PENNINGTON, R. T.; TAN, S.; KAYE, M. E.; SAMUEL, R. Plant DNA barcodes and assessment of phylogenetic community structure of a tropical mixed dipterocarp forest in Brunei Darussalam (Borneo). **Plos One**, 2017.

HIRSCH, J.E. Na index to quantify an individual's scientific research output. **PNAS**, v. 102, n. 46, p.16569-16572, 2005.

HOLLINGSWORTH, M.L.; CLARK, A.A.; FORREST, L.L.; RICHARDON, J.; PENNINGTON, R.T.; LONG, D.G.; COWAN, R.; CHASE, M.W.; GAUDEUL, M.; HOLLINGSWORTH, P.M. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. **Molecular Ecology Resources**, v. 9, p. 439-457, 2009.

HOLLINGSWORTH, P.M.; GRAHAM, S.W.; LITTLE, D.P. Choosing and Using a Plant DNA Barcode. **PLoS ONE**, v. 6, n. 5, 2011.

JIANG, K.; ZHANG, R.; ZHANG Z.; PAN.; TIAN B. DNA barcoding and molecular phylogeny of *Dumasia* (Fabaceae: Phaseoleae) reveals a cryptic lineage. **Plant Diversity**, v. 42, p. 376-385, 2020.

KENNEDY, S. R.; PROST, S.; OVERCAST, I.; ROMINGER, A.; GILLESPIE, R. G.; KREHENWINKEL, H. High-throughput sequencing for community analysis: the promise of DNA barcoding to uncover diversity, relatedness, abundances and interactions in spider communities. **Den Genes Evol**, p. 185-210, 2020.

- KIRCHER, M.; KELSO, J. High-throughput DNA sequencing – concepts and limitations. **Methods, Models & Techniques**, n. 32, p. 524-536, 2010.
- KRESS, W.J.; WURDACK, K.J.; ZIMMER, E.A.; WEIGHT, L.A.; JANZEN, D.H. Use of DNA barcodes to identify flowering plants. **PNAS**, v. 102, n. 23, p. 8369-8374, 2005.
- KRESS, W.J.; ERCKSON, D. A two-loco global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. **Plos One**, 2007.
- KRESS, W.J.; ERCKSON, D. DNA barcodes: genes, genomics and bioinformatics. **PNAS**, v. 105, n. 8, p. 2761-2762, 2008.
- KRESS, W. J. Plant DNA *barcodings*: applications today and in the future. **Journal of Systematics and Evolution**, v. 55, n. 4, p. 291-307, 2017.
- LAHAYE, R.; van der BANK, M.; BOGARIN, D.; WARNER, J.; PUPULIN, F.; GIGOT, G.; MAURIN, O.; DUHOIT, S.; BARRACLOUGH, T.G.; SAVOLAINEN, V. DNA barcoding the floras of biodiversity hotspots. **PNAS**, v. 105, p. 2923-2928, 2008.
- LI, D.Z. Floristics and plant biogeography in China. **Journal of Integrative Plant Biology**, v. 50, p. 771-777, 2008.
- LI, X.; YANG, Y.; HENRY, R.; ROSSENTO, M.; WANG, Y.; CHEN, S. Plant DNA barcoding: from gene to genome. **Biological Reviews**, 2015.
- MACIAS-CHAPULA, C.A. O papel da informetria e da cineciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, v. 27, n. 2, p. 134-140, 1998.
- MARCONI, G.; LANDUCCI, F.; ROSELLINI, D. DNA barcoding as a tool for Early warning alien duckweeds (*Lemna* sp.pl.): the case of Central Italy. **Plant Biosystems**, v. 153, 2019.
- MAXAM, A. M.; GILBERT, W. A new method for sequencing DNA. **Proceedings of the National Academy of Science**, 1976.
- MENEGON, M.; CANTALONI, C.; RODRIGUEZ-PIETRO, A.; CENTOMO, A. A.; ROSSATO, M.; XUMERLE, L.; LOADER, S.; DELLEDONNE, M. On site DNA barcoding by nanopore sequencing. **Plos One**, n. 12, v. 10, 2017.
- MORTHIE, S.; MATTOCKS, C.J.; WRIGHT, C.F. Review of massively parallel DNA sequencing technologies. **HUGO J**, n. 5, p. 5-1, 2011.
- PALHARES, R. M.; BARATTO, L. C.; SCOPEL, M.; MUGGE, F. L. B.; BRANDÃO M. G. Medicinal plants and herbal products from Brazil: how can we improve quality? **Frontiers in Pharmacology**, v. 11, 2021.
- PECNIKAR, Z. F.; BUZAN, E. V. 20 years since the introduction of DNA barcoding: from theory to application. **J Appl Genetics**, 2013.
- POMERANTZ A.; PEÑAFIEL, N.; ARTEAGA, A.; BUSTAMANTE, L.; PICHARDO, F.; COLOMA, L. A.; BARRIO-AMORÓS, C. L.; SALAZAR-VALENZUELA, D.; PROST, S. Real-time DNA barcoding in a rainforest using

nanopore sequencing: opportunities for rapid nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. **GigaScience**, v. 7, p. 1-14, 2018.

POTTER, C.; DE VERE, N.; JONES, L. E.; FORD, C. R.; HEGARTY, M. J.; HODDER, K. H.; DIAZ, A.; FRANKLIN, E. L. Pollen metabarcoding reveals broad and species-specific resource use by urban bees. **PeerJ**, 2019.

QUÉMÉRE, E.; HILBERT, F.; MIQUEL, C.; LHUILLIER, E.; RASOLONDRAIBE, E.; CHAMPEAU, J.; RABARIVOLA, C.; NUSBAUMER, L.; CHATELAIN, C.; GAUTIER, L.; RANIRISON, P.; CROUAU-ROY, B.; TABERLT, P.; CHIKHI, L. A DNA metabarcoding study of a primate dietary diversity and plasticity across its entire fragmented range. **Plos One**, v. 8, 2013.

RATNASINGHAM, S.; HEBERT, P. D. N.; BOLD: The Barcode of Life Data System ([www.barcodinglife.org](http://www.barcodinglife.org)). **Molecular Ecology Notes**, 2007.

RAPINI, A. Modernizando a taxonomia. **Biota Neotropica**, v. 4, n. 1, 2004.

RAVEN, P.H. Plant conservation in the future: New challenges, new opportunities. **Plant Diversity Resources**, v. 33, p. 1-9, 2011.

RIESEBERG, L.H.; WOOD, T.E.; BAACK, E.J. The nature of plant species. **Nature**, p. 524-527.

SANGER, F.; AIR G.M.; BARREL BG.; BROWN N.L.; COULSON, A.R.; FIDDES, C.A.; HUTCHISON, C.A.; SLOCOMBE, P.M.; SMITH, M. Nucleotide sequence of bacteriophage phi X174 DNA. **Nature**, 1977.

SAVOLAINEN, V.; COWAN, R.S.; VOGLER, A.P. Towards writing the encyclopaedia of life: and introduction to DNA barcoding. **Philosophical Transactions of The Royal Society Botanical**, 2005.

SHENDURE, J.; AIDEN, E.L. The expanding scope of DNA sequencing. **Nat. Biotechnol.**, n. 30, 2012.

SILVEIRA, B.S.; FERREIRA, E.B. Avaliação cientométrica da produção científica mundial de 2013 a partir da base Scopus. **Sigmae**, v.8, n.2, p. 8-18, 2019.

SPINAK, E. **Dicionário enciclopédico de bibliometria, cienciometria e infometria**. Caracas: Unesco, 1996.

TEAM, R.C. R: A language and environment for statistical computing. An Introduction to R. Version 4.0.2. Disponível em <<https://www.r-project.org/>>. Acesso em: 22 Jun. 2020.

TREBACH, R.; CERQUEIRA, N.M.; MEDEIROS, S.R.; GUTIERREZ, H.J.P.; HERNANDEZ, N.O.; RODRIGUES, M.D.N. DNA barcoding: uma ferramenta de apoio molecular para identificação de espécies de peixes. **Scientia Agraria Paranaensis**, v. 14, n. 2, p. 77-81, 2015.

VALENTINI, A.; POMPANON, F.; TABERLET, P. DNA barcoding for ecologists. **Trends in Ecology and Evolution**, v. 24, n. 2, 2009.

VANTI, N.A.P. Da bibliometria à webometria; uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da Informação**, v. 31, n. 2., p. 152-162, 2002.

XIN, T.; XU, Z.; JIA, J.; LEON, C.; HU, S.; LIN, Y.; RAGUPATHY, S.; SONG, J.; NEWMASER, S. G. Biomonitoring for traditional herbal medicinal products using DNA metabarcoding and single molecule, real-time sequencing. **Acta Pharmaceutica Sinica B**, v. 8, n. 3, p. 488-497, 2018.

YANG A.; ZHANG W.; WANG, J.; YANG, K.; HAN, Y.; ZHANG, L. Review on the application of machine learning algorithms in the sequence data mining of DNA. **Frontiers in Bioengineering and Biotechnology**, v. 8, 2020.

YANG, C.; Lv, Q.; ZHANG, A. Sixteen years of DNA barcoding in China: What has been done? What can be done? **Frontiers in Ecology and Evolution**, v. 8, 2020.