



BASES DE DADOS E ESTRATÉGIAS DE RECUPERAÇÃO DE INFORMAÇÃO PARA ANÁLISE DE PATENTES DERIVADAS DA BIODIVERSIDADE BRASILEIRA:

estudo de caso para a flora nativa da Mata Atlântica

Celise Villa dos Santos¹
Leandro Innocentini Lopes de Faria²
Fabio Mascarenhas e Silva³

Resumo: Avalia estratégias para a recuperação de patentes derivadas da biodiversidade dos biomas brasileiros, com estudo de caso para a flora nativa da Mata Atlântica. Considera características, recursos e limitações das fontes de informação sobre o patrimônio genético brasileiro, e das bases de dados de patentes. Partindo do nome do bioma (Estratégia 1), e das espécies que ocorrem em um bioma (Estratégia 2), conclui que a estratégia de busca baseada no nome espécies permite identificar maior número de patentes, embora a estratégia baseada no nome do bioma pode ser útil no monitoramento de biopirataria.

Palavras-Chave: Biomas brasileiros. Patentometria. Patrimônio genético. Mata Atlântica.

1 INTRODUÇÃO

O Brasil, dada sua biodiversidade distribuída em seis biomas, possui enorme potencial de produção de tecnologias a partir de seu patrimônio genético. Essas tecnologias protegidas na forma de patentes fomentam setores econômicos como o farmacêutico, de cosméticos, alimentos, construção civil, mobiliário e agrícola. A proteção da propriedade industrial brasileira na área de biotecnologia, entretanto, envolve questões éticas e técnicas, como, por exemplo, na forma de acesso ao patrimônio genético e aos conhecimentos tradicionais associados, atualmente regulamentadas pela Lei nº 13.123 de 20/05/2015 (FAPESP, 2021; MMA, 2021).

A análise de patentes tem sido utilizada para monitorar como o patrimônio genético e os conhecimentos tradicionais de um país têm sido explorados para agregar valor em produtos e processos. E também como instrumento para reconhecer e monitorar a apropriação indevida destes recursos, e garantir o cumprimento das políticas de repartição de benefícios. Os

¹ Instituto Nacional da Mata Atlântica (INMA)

² Universidade Federal de São Carlos (UFSCar)

³ Universidade Federal de Pernambuco (UFPE)

procedimentos para identificação dessas patentes partem, em geral, da seleção de um conjunto de termos baseados em nomes de espécies que ocorrem em determinadas localizações geográficas, e/ou dos nomes das localizações geográficas. (OLDHAM; BARNES; HALL, 2015; SIMMONDS *et al.*, 2020; SORIA-LÓPEZ; FUENTES-PÁRAMO, 2016).

No Brasil há carência de identificação, organização e análise de informação sobre o conjunto de patentes relacionadas aos biomas brasileiros, e à biodiversidade como um todo. Os estudos, em geral, são limitados à análise de patentes para poucas espécies (AMARAL; FIERRO, 2013; CORDEIRO, 2021; YANAI, 2012).

Neste contexto, este trabalho teve como objetivo avaliar possibilidades de procedimentos para identificar patentes envolvendo o conjunto da biodiversidade dos biomas brasileiros, contribuindo para a proposição de bases teórico-metodológicas de monitoramento de informações de patentes.

Duas estratégias para a recuperação de patentes reportadas na literatura foram aplicadas e comparadas para a flora nativa do bioma Mata Atlântica. Uma baseada no nome do bioma (Estratégia 1), e outra no nome das espécies que ocorrem no bioma (Estratégia 2). Para a aplicação das estratégias foi necessário identificar e analisar recursos e limitações das fontes de informação sobre o patrimônio genético brasileiro, e dos bancos de dados de patentes.

2 METODOLOGIA

O procedimento para aplicação e avaliação das estratégias de busca de patentes consistiu em:

Estratégia 1: Busca pelo nome do bioma

- a) Desenvolvimento de expressão de busca (Tabela 1) representativa do bioma “Mata Atlântica” a partir da literatura (FREITAS *et al.*; 2021; GUERRA *et al.*; 2020; MMA, 2021).
- b) Recuperação de patentes com a terminologia selecionada em quatro bases de patentes: Derwent Innovation Index-Clarivate, Patentscope, Espacenet e Google. Na base comercial Derwent, os termos foram aplicados no campo de pesquisa Tópico, que consiste em um resumo produzido pela base a partir de elementos do texto completo da patente. Nas bases de acesso aberto Patentscope, Espacenet e Google Patents a busca foi aplicada no texto completo, por permitirem esta funcionalidade. O número de patentes recuperadas em cada base em 9 nov. 2021, sem restrição de período de dados, constam da Tabela 1.

A base Google Patents retornou o maior número de registros. As quantidades variadas de patentes entre as bases devem-se: - ao termo Mata Atlântica ser incomum nos campos título ou resumo das patentes; - a maioria das patentes brasileiras não possuir o texto integral disponível nos campos de busca das bases Patentscope e Espacenet; e, - a diferença na forma como operam os tradutores da Espacenet e da Google Patents na tradução do chinês para o inglês. Após a análise de conteúdo e remoção das duplicatas dos registros coletados nas quatro bases foram identificados 132 documentos relacionados ao bioma brasileiro Mata Atlântica.

Tabela 1 - Expressão de busca e número de patentes coletadas pela Estratégia 1, com a terminologia “Mata Atlântica”

Expressão de busca	Número de documentos de patentes			
	Derwent ¹	Patent scope ²	Espacenet ²	Google Patents ²
“atlantic forest” OR “atlantic rain forest” OR “atlantic tropical forest” OR “mata tropical atlantica” OR “floresta tropical atlantica” OR “floresta atlantica” OR “mata atlantica” OR “mata tropical atlântica” OR “floresta tropical atlântica” OR “floresta atlântica” OR “mata atlântica”	1	22	43	135
Total de patentes sobre “Mata Atlântica” sem duplicatas	132			

Fonte: Elaborada pelos autores (2021).

c) No conjunto das 132 patentes relacionadas à Mata Atlântica foram identificados 215 nomes científicos de espécies em latim na forma binomial (gênero + espécie). Em 73% dos casos o nome científico de uma ou mais espécies estava no título e/ou resumo da patente.

Entre estas 215 espécies, 8 eram de plantas nativas da Mata Atlântica, identificadas através da base SiBBR em 9 nov. 2021. Para estas 8 espécies foi realizada busca de patentes na base Derwent (em 19 nov. 2021, no campo Tópico) obtendo 542 patentes. A base Derwent foi escolhida em razão da comparação com a Estratégia 2.

Estratégia 2: Busca pelo nome das espécies do bioma

Para a coleta das patentes pelos nomes científicos das espécies que ocorrem em determinado bioma, é necessário identificar as espécies por tipologia e origem, em bases que reúnem informações sobre o patrimônio genético brasileiro. As fontes de informação sobre patrimônio genético brasileiro foram identificadas a partir de entrevistas a seis especialistas em botânica e fauna brasileiras, todos pesquisadores do INMA (Instituto Nacional da Mata Atlântica). Foram identificadas cinco bases: - SiBBR (Sistema de Informação sobre a Biodiversidade

Brasileira), - speciesLink, -Reflora/CNPQ, -Catálogo Taxonômico da Fauna do Brasil, e, - GBIF (Global Biodiversity Information Facility).

As cinco bases, todas de acesso aberto, variam: - na forma de seleção das espécies por local de ocorrência (região/bioma/estado brasileiro); - na forma de classificação dos grupos taxonômicos por tipo e origem (endêmica, nativa, naturalizada ou cultivada); e, - e na forma de exportação dos dados. As bases também possuem grande variação no número de espécies do catálogo, como mostrado na Tabela 2 para espécies que ocorrem na Mata Atlântica.

Tabela 2 - Número de espécies de animais e plantas da Mata Atlântica em bases de dados sobre a biodiversidade que possibilitam a busca por biomas brasileiros

Reino	Referência de número de espécies da Mata Atlântica (MMA, 2021)	Reflora*	SiBBr**	species Link
Animal	1361 - 2040	0	8.209	35.260
Vegetal	20.000	17.820	18.526	84.240

*Domínio fitogeográfico “Mata Atlântica”, endêmicas e não endêmicas do Brasil, todas as origens (cultivada, nativa ou naturalizada), busca até subespécie/variedade. **Nomes de espécies a partir da “lista de verificação das espécies” sem duplicatas. Acesso entre 25 e 26/10 de 2021.

Fonte: Elaborada pelos autores (2021).

Segundo os especialistas, a variação no número de espécies entre as bases é devida, principalmente:

- Ao fato de bases como SibBr, speciesLink e GBIF serem bases de espécimes (coleções científicas), acarretando problemas quando se geram listas de espécies por regiões/biomas, pois incluem diferentes nomes para as mesmas espécies, inflando as listas com redundâncias, além de erros de coordenadas;
- Constantes alterações taxonômicas devido à evolução do conhecimento científico, e consequente atraso na atualização das bases por especialistas. Atualização de informações relativas a, por exemplo, domínio geográfico de ocorrência de espécies e características como endemismo;
- Forma de compilação da informação sobre localização geográfica das espécies. No caso da Mata Atlântica pelos limites do mapa de biomas (IBGE, 2019) ou da Lei da Mata Atlântica (IGGE, 2008); e,
- Atraso na atualização entre bases de dados, como a alimentação de dados das bases REFLORA e SiBBr na GBIF.

Para dados sobre flora, utilizados como estudo de caso, a base Reflora/CNPQ foi indicada pelos especialistas como a mais atualizada, com listas de espécies que passaram por processo de curadoria, minimizando problemas de sinonímia. A base Reflora foi utilizada (em 11 de out. 2021) para recuperar uma lista de 9571 espécies de plantas nativas que ocorrem somente na Mata Atlântica.

d) A partir da lista de 9571 espécies foram recuperadas 1251 patentes a partir do campo Tópico da base Derwent Innovation Index (em 19 nov. 2021). Os dados de patentes foram exportados e processados no programa VantagePoint 5.0 (Search Technology, Inc.) para identificação de 138 espécies com patentes. A base Derwent foi escolhida por ser a única entre as analisadas a possibilitar o uso de até 16.000 termos no campo de busca.

3 RESULTADOS E DISCUSSÃO

A Estratégia 2 de busca de patentes baseada no nome científico das espécies permitiu identificar maior número de patentes (1251) e de espécies com patentes (138), se mostrando mais adequada em relação à Estratégia 1 para localizar patentes relativas a um bioma. Embora quase metade das patentes recuperadas pela Estratégia 2 terem sido também recuperadas pela Estratégia 1 (542 patentes) com apenas 8 espécies (Tabela 3).

Tabela 3 - Número de patentes para espécies de flora nativa da Mata Atlântica recuperadas na base Derwent pelas Estratégias de busca 1 e 2

Estratégia de busca	Espécies	Espécies de plantas nativas da Mata Atlântica	Espécies de plantas nativas da Mata Atlântica com patentes	Número de patentes
Estratégia 1	215 (vários reinos e biomas)	8	8	542
Estratégia 2	9571 (plantas nativas da Mata Atlântica)	9571	138	1251

Fonte: Elaborada pelos autores (2021).

Estando o uso da metodologia para os demais biomas brasileiros, a Estratégia 2 também pode ser mais adequada para localizar patentes, pois Amazônia, Caatinga, Cerrado, Pampa e Pantanal, são termos polissêmicos, compondo nomes de espécies, instituições e empresas. A aplicação da Estratégia 1 requereria maior refinamento na expressão de busca e análise de conteúdo pormenorizada.

Entretanto, entre as limitações identificadas no uso da Estratégia 2, destacam-se:

- A suposição de que o requerente da patente divulgará as origens de um composto ou preparação a partir de uma espécie particular utilizando termos latinos, desconsiderando variantes de nomes científicos de espécies, ou nomes populares que variam entre países/ regiões/ idiomas étnicos. Também a possibilidade de mencionar, na patente, somente o nível taxonômico mais alto da espécie, ou apenas o nome de compostos isolados originários de espécies (OLDHAM; BARNES; HALL, 2015);
- O caráter dinâmico do conhecimento científico sobre as espécies de um bioma, e seu respectivo registro em bases de dados em constante atualização;
- A atual restrição no número de termos nos campos de busca das bases de patentes de acesso aberto, de até 20, aproximadamente, quando podem ser necessários de 1.000 a 20.000 se deseja identificar o número de patentes relacionadas a um bioma na totalidade. E o fato de bases comerciais como a Derwent, que permite busca simultânea a grande número de termos, não possibilitar pesquisa no texto completo, ou em seções específicas da patente como no campo das reivindicações, onde se delimita o objeto da proteção e possibilitaria buscas mais precisas.

A principal limitação do uso da Estratégia 1 que implica na recuperação de menor número de patentes é que a declaração de origem do recurso natural, em geral, ocorre no texto descritivo da patente e não no título ou resumo. Além da possibilidade do depositante omitir, intencionalmente ou não, esta informação. A Estratégia 1, entretanto, pode ser útil no monitoramento de biopirataria, pois o fato do nome do país de origem/ bioma da espécie não constar no documento da patente pode ser avaliado como indício de acesso ilegal ao patrimônio genético naquelas patentes onde espécies da biodiversidade brasileira constem no campo reivindicações (SORIA-LÓPEZ; FUENTES-PÁRAMO, 2016).

4 CONSIDERAÇÕES FINAIS

A análise de patentes desenvolvidas a partir de recursos genéticos é um importante instrumento para avaliar como a biodiversidade de um país contribui para a agregação de valor econômico a produtos, e, ao mesmo tempo, para a conservação da biodiversidade e a preservação dos direitos de comunidades tradicionais. Ao identificar e analisar as características e limitações de duas estratégias voltadas à recuperação de informações de patentes para o conjunto da biodiversidade brasileira, o estudo fornece subsídios para a proposição de bases teórico-metodológicas de monitoramento de patentes em biotecnologia, e

também auxilia na construção de uma visão estratégica para orientar políticas públicas em ciência, tecnologia, inovação e conservação.

AGRADECIMENTOS

Aos pesquisadores do INMA que contribuíram com informações sobre bases de dados de biodiversidade brasileira. Ao CNPQ pelo financiamento (Processo n. 300916/2022-1).

REFERÊNCIAS

- YANAI, A. M. **Patentes de produtos naturais amazônicos: análise do impacto da inovação tecnológica mundial**. 2012. 154 f. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) – Universidade Federal de São Carlos, São Carlos. 2012.
- CORDEIRO, L.S. **Cenário das patentes relacionadas ao babaçu indexadas na base de dados Derwent Innovations Index**. 2021. 141 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal do Rio Grande do Sul, Porto Alegre. 2021.
- FAPESP - FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Biota -Fapesp. Program - Brazilian biodiversity research: a promising future**. 2021. Disponível em: https://fapesp.br/publicacoes/pasta_biota_2017.pdf. Acesso em: 01 mar. 2021.
- AMARAL, L.F.G.; FIERRO, I.M. Profile of medicinal plants utilization through patent documents: the andiroba example. **Revista Brasileira de Farmacognosia**, João Pessoa, v. 23, n. 4, p. 716-722, 2013.
- FREITAS, J. L.; SOBRAL, N. V.; SILVA, F. M. E. Indicadores de coautoria nas publicações sobre Mata Atlântica: análise bibliométrica na base Web of Science. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 21., 2021. **Anais [...]**. Rio de Janeiro: [s.n.], 2021.
- GUERRA, A. *et al.* Ecological restoration in Brazilian biomes: identifying advances and gaps. **Forest ecology and Management**, Orlando, v. 458, p. 117802, 2020.
- IBGE. **Mapa da Área de Aplicação da Lei nº 11.428, de 2006**. 2008. Disponível em: https://geoftp.ibge.gov.br/informacoes_ambientais/estudos_ambientais/biomas/mapas/lei11428_mata_atlantica.pdf. Acesso em: 22 fev. 2022.
- IBGE. **Mapa de Biomas e Sistema Costeiro-Marinheiro do Brasil**. 2019.. Disponível em: <https://geoftp.ibge.gov.br>. Acesso em: 22 fev. 2022.
- MMA - Ministério do Meio Ambiente. **Biomas**. Disponível em: <https://antigo.mma.gov.br/biomas.html>. Acesso em 22 out. 2021.
- OLDHAM, P.; BARNES, C.; HALL, S. **A review of UK patent activity for genetic resources and associated traditional knowledge**. [S.l.: s.n.], 2015.
- SIMMONDS, M. S. *et al.* Biodiversity and patents: Overview of plants and fungi covered by patents. **Plants, People, Planet**, Oxford, v. 2, n. 5, p. 546-556, 2020.
- SORIA-LÓPEZ, M.; FUENTES-PÁRAMO, I. The identification of biopiracy in patents. **World Patent Information**, Oxford, v. 47, p. 67-74, 2016.