

DEZEMBRO 2022

# V10/04

REVISTA DE CIÊNCIA ELEMENTAR. CASA DAS CIÊNCIAS



# REVISTA DE CIÊNCIA ELEMENTAR



## FICHA TÉCNICA

*Rev. Ciência Elem.*, V10(04)

**Publicação trimestral  
da Casa das Ciências**

ISSN 2183-9697 (versão impressa)

ISSN 2183-1270 (versão online)

[rce.casadasciencias.org](http://rce.casadasciencias.org)

DEPÓSITO LEGAL

425200/17

COORDENAÇÃO EDITORIAL

Alexandra Coelho

DESIGN

Rui Mendonça

PAGINAÇÃO

Raul Seabra

IMPRESSÃO E ACABAMENTO

Uniarte Gráfica S.A.

TIRAGEM

3150 exemplares

© Todo o material publicado nesta revista pode ser reutilizado para fins não comerciais, desde que a fonte seja citada.



## PROPRIETÁRIO

Casa das Ciências/ICETA

Faculdade de Ciências,

Universidade do Porto

Rua do Campo Alegre, 687

4169-007 Porto

[rce@asadasciencias.org](mailto:rce@asadasciencias.org)

## CORPO EDITORIAL DA REVISTA DE CIÊNCIA ELEMENTAR

### EDITOR

João Nuno Tavares (UNIVERSIDADE DO PORTO)

### EDITOR CONVIDADO

Alexandre Lopes Magalhães (UNIVERSIDADE DO PORTO)

### CONSELHO EDITORIAL

Alexandre Lopes Magalhães (UNIVERSIDADE DO PORTO)

Jorge Manuel Canhoto (UNIVERSIDADE DE COIMBRA)

Paulo Ribeiro-Claro (UNIVERSIDADE DE AVEIRO)

Paulo Fonseca (UNIVERSIDADE DE LISBOA)

José Cidade Mourão (INSTITUTO SUPERIOR TÉCNICO)

Luís Vítor Duarte (UNIVERSIDADE DE COIMBRA)

José Francisco Rodrigues (UNIVERSIDADE DE LISBOA)

### PRODUÇÃO E SECRETARIADO

Alexandra Coelho

Ana Santos

Guilherme Monteiro

Raul Seabra

### NORMAS DE PUBLICAÇÃO NA RCE

A Revista de Ciência Elementar dirige-se a um público alargado de professores do ensino básico e secundário, aos estudantes de todos os níveis de ensino e a todos aqueles que se interessam pela Ciência. Discutirá conceitos numa linguagem elementar, mas sempre com um rigor superior.

### INFORMAÇÃO PARA AUTORES E REVISORES

Convidam-se todos os professores e investigadores a apresentarem os conceitos básicos do seu labor diário numa linguagem que a generalidade da população possa ler e compreender.

Para mais informação sobre o processo de submissão de artigos, consulte a página da revista em [rce.casadasciencias.org](http://rce.casadasciencias.org)



**CASA DAS  
CIÊNCIAS**

EDULOG · FUNDAÇÃO BELMIRO DE AZEVEDO

DEZEMBRO 2022

# V10/04

## ÍNDICE

- 02 AGENDA
- 03 NOTÍCIAS
- 05 PROFESSOR DO ANO 2022
- EDITORIAL
- 07 **A Revolução no tratamento de dados em Ciência**  
Alexandre Lopes Magalhães
- ARTIGOS
- 09 **Classificação taxonómica em Biologia usando IA**  
Luís M. B. Lopes,  
Eduardo R. B. Marques
- 14 ***Big data***  
Sérgio P. Ávila, Carlos Melo,  
Patrícia Madeira
- 19 **Bases de dados de Proteínas**  
Ana Oliveira
- 24 **Imagens de satélite, geologia e IA**  
Joana Cardoso-Fernandes *et al.*
- 29 **Asterossismologia**  
Margarida S. Cunha,  
Mário J. P. F. G. Monteiro
- 35 **Determinismo biológico, genético e epigenético**  
Luca Ribeiro Mendes Nicola,  
Edson Pereira Silva
- 40 **Avanços da genómica**  
Agostinho Antunes
- AOS OLHOS DA CIÊNCIA
- 43 **Serendipidade na investigação científica**  
Alexandre Lopes Magalhães
- IMAGEM EM DESTAQUE
- 48 **Flor de Borragem**  
Rubim Almeida,  
José Pissarra

**20/12**<sup>(2022)</sup>

Uma viagem à origem da vida

“A origem e procura de vida no Universo são duas das grandes questões para as quais a Ciência ainda não tem resposta. A curiosidade humana não conhece fronteiras, e invariavelmente surge a pergunta: haverá vida noutros planetas? E como surgiu a vida na Terra?”

Na Conferência de Natal Ciência Viva 2022, a astrobióloga Zita Martins propõe-se a fazer uma viagem com o público pelo espaço, às nossas origens e aos locais do Universo onde existem condições para a vida se ter desenvolvido. Contando com a ajuda de asteroides, meteoritos, cometas e poeira cósmica vamos viajar pelo tempo e pelo espaço sideral para descobrir como tudo começou.

Evento grátis mas de inscrição obrigatória.

TEATRO ABERTO, LISBOA

[HTTPS://WWW.CIENCIAVIVA.PT/CONFERENCIA-DE-NATAL/2022](https://www.cienciaviva.pt/conferencia-de-natal/2022)

**17/07**  
**a 19/07**<sup>(2023)</sup>

VIII Encontro Internacional da Casa das Ciências

Subordinado ao tema da Energia, iremos rumar à cidade dos canais, Aveiro, para o VIII Encontro Internacional da Casa das Ciências.

Mantendo a estrutura de sempre, ofereceremos vários painéis, oficinas e plenárias temáticas para os grupos habituais.

No início de 2023 apresentaremos o programa e as inscrições serão abertas.

Esteja atento!

UNIVERSIDADE DE AVEIRO

[WWW.CASADASCIENCIAS.ORG/8ENCONTROINTERNACIONAL/](http://WWW.CASADASCIENCIAS.ORG/8ENCONTROINTERNACIONAL/)

**16/07**  
**a 20/07**<sup>(2023)</sup>

XI Congresso Nacional de Geologia Geociências e desafios globais

O XI Congresso Nacional de Geologia (XI CNG) terá lugar em Coimbra, no Edifício Central da FCTUC (Pólo 2 da Universidade de Coimbra), onde se localizam as instalações do Departamento de Ciências da Terra. No XI CNG procurar-se-á destacar o papel das Geociências na concretização dos objetivos do desenvolvimento sustentável segundo as Nações Unidas, contribuindo assim para a resolução de 8 Desafios Globais: (1) Minerais e inovação tecnológica, (2) Materiais e economia circular, (3) Riscos naturais e alterações climáticas, (4) Acesso e qualidade da água, (5) Descarbonização e fontes de energia, (6) Saúde e qualificação ambiental, (7) Segurança e sustentabilidade urbana e (8) Geodiversidade e geoconservação. Para além das sessões científicas (orais e posters), excursões e sessão da Sociedade Geológica de Portugal, o XI CNG contará ainda com *workshops* dirigidos a professores do ensino básico e secundário.

Para mais informações, consulte a página do congresso.

UNIVERSIDADE DE COIMBRA, PÓLO 2

[HTTPS://XICNG.NET/](https://xicng.net/)



---

## Ciência e Arte



FIGURA 1. *Théâtre D'opéra Spatial*. (Imagem de Jason M. Allen)

Um recente artigo do *The New York Times*, assinado por Kevin Rose, dá conta do 1.º lugar no concurso anual de pintura da Feira do Estado norte-americano do Colorado de 2022, atribuído ao artista Jason M. Allen. Ele concorreu na categoria digital mas a novidade é que usou o programa *Midjourney* baseado em Inteligência Artificial para criar a obra, que se intitula *Théâtre D'opéra Spatial* (FIGURA 1), passando-a finalmente à tela usando uma impressora própria.

Este resultado criou mal-estar no seio da comunidade artística por ultrapassar algumas linhas definidoras do *status quo*.

O programa *Midjourney* e outros, como o *Stable Diffusion* e o *DALL-E2*, funcionam com base em milhões de imagens acessíveis na internet, ensinando algoritmos de inteligência artificial a reconhecer padrões e relações entre eles para originarem novas imagens no estilo pretendido. O que significa que os artistas que disponibilizaram as suas obras estão, de certo modo, a

contribuir para o treino dos seus competidores algorítmicos. Mas este argumento nem é o mais forte, porque uma obra que é exposta numa galeria ou museu torna-se domínio público e tem o potencial de inspirar novos artistas. A diferença reside aqui no mecanismo de criação; nestes casos, puros amadores conseguem gerar obras sofisticadas introduzindo simplesmente algumas palavras numa caixa de texto do programa.

---

## Inteligência artificial na produção de hidrogénio verde



FIGURA 1. Estação de produção de hidrogénio na ilha de Eday-Escócia.

O hidrogénio é uma fonte de energia cada vez mais procurada por ser renovável e limpa; a sua combustão produz água. Muita investigação tem sido dedicada à produção e armazenamento deste gás utilizando fontes de energia renováveis e não poluentes, o que levou à criação da nova designação de hidrogénio verde, apesar da substância elementar ser sempre a mesma,  $H_2$ . Recentemente, a empresa *H<sub>2</sub>GO Power*

do Reino Unido, num consórcio com a *European Marine Energy Centre* (EMEC) e o *Imperial College* de Londres, instalou uma estação piloto para produção de hidrogénio verde na ilha escocesa de Eday que usa a Inteligência Artificial na sua gestão.

Este projeto foi recentemente reforçado com uma verba de 494 000 libras esterlinas do *Department of Business, Energy and Industrial Strategy* do governo britânico, o que mostra a importância deste tópico.

Os algoritmos de Inteligência Artificial são utilizados para integrar dados históricos e de previsão de meteorologia (ventos e marés), de preços da energia elétrica e de funcionamento da própria rede de distribuição, de modo a otimizar a produção e armazenamento de hidrogénio.

---

## Ilhas de plástico

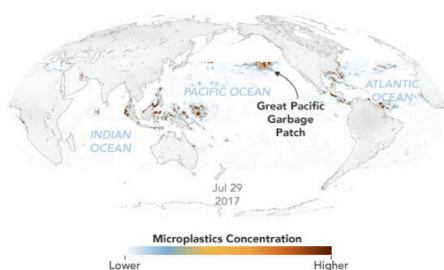


FIGURA 1. Distribuição de plásticos nos oceanos. (<https://www.earthdata.nasa.gov/learn/articles/ocean-plastic>)

Um grupo de investigadores da Universidade do Texas, em Austin-USA, apresentaram, em 2022, um estudo na revista *Nature* em que são usadas técnicas de *machine-learning* para conceber mutantes da enzi-

ma PETase que aumentem a sua atividade catalítica, com o objetivo de a aplicar à escala industrial. Esta enzima, descoberta em 2016 por investigadores da Universidade de Kyoto no Japão, é produzida naturalmente pela bactéria *Ideonella sakaiensis* e degrada o PET (Politereftalato de etileno), um tipo de plástico muito usado no fabrico de garrafas. A produção anual de PET no mundo é atualmente de 81 milhões toneladas, o que representa cerca de 20% da produção total de plásticos. A reduzida atividade química destes polímeros sintéticos torna a sua reciclagem difícil, criando-se assim um grave problema ambiental. As correntes marítimas vão acumulando os detritos lançados no ambiente em zonas específicas dos oceanos. Investigadores da NASA têm usado dados de satélites e técnicas de *machine-learning* para monitorizar essas ilhas nos oceanos da Terra; uma delas, localizada no nordeste do oceano Pacífico, e designada *Great Pacific Garbage Patch*, cobre uma área equivalente ao estado norte-americano do Texas ou da França (FIGURA 1).

Este é, portanto, um problema ambiental que tem ocupado equipas multidisciplinares de investigadores na procura urgente de soluções viáveis e em que técnicas de *Big Data* estão efetivamente a revelar-se bastante úteis.

---

# Professor do Ano 2022



**Francisco Félix**

Francisco António Fidalgo Félix Dias, de 63 anos, é natural de Peniche onde vive e trabalha como professor do grupo 520 (Biologia e Geologia) na Escola Secundária daquela cidade. Sempre ligado de forma umbilical à terra que o viu nascer, é Licenciado em Ensino de Biologia e Geologia, pela UA, desde 1984 e Mestre em Ciências da Terra e da Vida para o Ensino, pela FCUL, desde 2004. A sua dissertação de mestrado não podia deixar de estar associado a Peniche: “A Geologia do Concelho de Peniche: património natural, recursos pedagógicos e valor cultural”.

Exercendo funções docentes desde que se licenciou, na escola onde hoje está, desempenhou praticamente todos os cargos associados ao desempenho dessa atividade: diretor de turma, diretor de curso (Ensino Profissional), coordenador de ano, coordenador de departamento, representante de grupo de recrutamento, avaliador externo, membro da Assembleia de Escola e do Conselho Geral, e ainda Coordenador dos programas Eco-Escolas (2016—2022) e Escola Azul (2018—2022), sendo Fundador e responsável pelo Clube Ecológico *O Airinho* (1993—2004).

Um divulgador acérrimo do património da sua terra, participou em inúmeros congressos, seminários e colóquios, quer de índole educacional, quer de natureza mais abrangente, ligados às forças vivas e ao património geológico que sempre defendeu. Com mais de uma dezena de artigos científicos e livros publicados, colaborou em projetos de investigação envolvendo a Geologia de Peniche, integrando equipas do Departamento de Ciências da Terra da Universidade de Coimbra e do Departamento de Paleontologia da Universidad Complutense de Madrid.

Integra o Conselho Científico da AGEO — Associação Geoparque Oeste, sendo ainda sócio fundador e Presidente do Conselho Diretivo da Associação Arméria — Movimento Ambientalista de Peniche, do qual foi Vice-Presidente e Presidente da Assembleia Geral. Colabora também com a Câmara Municipal de Peniche na divulgação do património do Concelho. Em 2019, foi galardoado com o prémio *Profissional do Ano*, pelo *Rotary Clube de Peniche*, prémio associado à investigação, preservação e divulgação do Património Natural de Peniche, pelo trabalho desenvolvido em prol do património do Concelho e, nesse mesmo ano, a Câmara Municipal atribuiu-lhe a *Medalha de Mérito Municipal de Dedicção*. Em 2022, recebeu um *Louvor de Mérito* atribuído pelo Conselho Pedagógico da Escola Secundária de Peniche, pelo excelente trabalho de coordenação, colaboração e divulgação de atividades desenvolvidas no âmbito do ambiente, através do estabelecimento de diversas parcerias internas e externas à Escola.

Fotógrafo de excelência, possui centenas de trabalhos, dos quais mais de uma centena publicados no *Banco de Imagens* da Casa das Ciências, tendo recebido por diversas vezes distinções sobre essa atividade. Em 2018, recebeu o *Prémio de Distinção*, em 2019, o *Prémio de Mérito* e, em 2021, o *Prémio de Excelência*, sempre na categoria de Fotografia Educativa. Participou em várias exposições, sendo de destacar a exposição de fotografia *Superfícies, texturas e pavimentos na orla costeira de Peniche*, a exposição *Fragmentos da Geologia de Peniche — uma perspetiva estética*; colaborou na exposição coletiva *Arte [Des]Confinada* e ainda a exposição de fotografia *inflorescências*.

O reconhecimento da sociedade e dos seus pares, como provam os inúmeros testemunhos que recebemos e as distinções já referidas, é unânime na aceitação e satisfação pelo reconhecimento que a Casa das Ciências lhe presta este ano.

*O empenho e dedicação do Prof. Francisco Félix tem sido extremamente relevante na educação e formação dos jovens de Peniche, despertando vocações e conferindo competências para a formação de pessoas íntegras e profissionais de relevo, não só no contexto local como a nível nacional. Sinto-me honrado e com profundo sentimento de orgulho em ter sido seu aluno.*

Sérgio Leandro, Antigo aluno e Diretor da Escola Superior de Turismo e Tecnologia do Mar — IPLeiria

*Como colega e amiga, não teria aqui espaço para realçar todas as qualidades e as iniciativas que o Prof. Francisco Félix teve como professor da Escola Secundária de Peniche e como cidadão. O Prof. Francisco Félix é um verdadeiro homem de cultura, não só pelo que sabe, na sua área de formação, literatura, música, fotografia, etc., mas também pelo seu gosto em transmitir e envolver os outros. Como foi bom ter sido sua colega e poder participar em tantas atividades que ele idealizou, distribuindo tarefas de modo a integrar o maior número possível de membros da comunidade escolar e até da comunidade local.*

Maria Rosário Cavalheiro Silva, Professora da Escola Secundária de Peniche

*"Professor do Ano" é a distinção atribuída anualmente pela Comissão Editorial da Casa das Ciências a um professor em reconhecimento do seu mérito como docente do ensino básico ou secundário e da sua disponibilidade de partilhar a sua experiência com os colegas.*



---

# A Revolução no tratamento de dados em Ciência

As sociedades humanas atravessam neste momento alterações profundas na sua organização. O fácil acesso à informação e a sua rápida circulação estão a criar novas formas de relacionamento social entre os cidadãos e entre estes e os seus governos. As preocupações com a proteção de dados e a relação com as liberdades, direitos e garantias dos cidadãos entraram definitivamente na agenda política. Tudo isto se encontra intimamente ligado à revolução que se está a observar no campo da Ciência e que ilustra bem o impacto que ela tem nas nossas vidas.

De facto, a História mostra-nos que a evolução da Ciência passa por períodos de impulsos disruptivos que afetam o modo como interpretamos o Mundo e até a própria organização das sociedades. Há sensivelmente duas décadas, no dealbar deste novo século/milénio, iniciámos uma nova fase tecnológica, a era da globalização digital, que acelerou vertiginosamente a produção e troca de informação em todo o planeta. Assistimos à criação de novas formas de produzir, armazenar e analisar dados, que justifica a existência do novo campo interdisciplinar da Ciência de Dados, ou *Data Science*, que utiliza técnicas computacionais, algorítmicas, estatísticas e matemáticas para tratar eficientemente grandes quantidades de dados e permitir extrapolar conhecimento.

A crescente capacidade no armazenamento de dados e a diminuição consistente do seu preço, aliada ainda a uma maior facilidade de consulta, vai continuamente alimentando a obsessão com a acumulação de informação que, mesmo aparentemente inútil no momento, poderá ser utilizada posteriormente para um determinado fim. Esta nova era digital, além de abrir todo um novo mundo de possibilidades de bem-estar, ameaça-nos também com inesperados desafios ao *status quo* a que todos nós temos de responder. Como sempre aconteceu no passado, a humanidade enfrenta o perigo do aproveitamento de uma nova tecnologia para fins menos altruístas e, como reação, muitos mecanismos de proteção de dados são concebidos para manter a privacidade dos cidadãos e evitar o uso ilícito da informação.

Neste número da Revista quisemos dar realce a este fenómeno, reunindo um conjunto interessante de artigos que, de um modo direto ou indireto, abordam o uso de grandes quantidades de dados para gerar conhecimento científico. Assim, mostramos aqui exemplos de estudos na área da astrofísica para interpretação de dados sobre a dinâmica no interior das estrelas, na geologia, com o tratamento de dados imagiológicos de satélite para identificação de depósitos minerais na crosta terrestre, na bioquímica, com o recurso a grandes bases de dados de proteínas para a conceção de novos fármacos, na biologia, com o estudo da descodificação do genoma humano, com um outro sobre biogeografia marinha macaronésica, e ainda, com um artigo que exemplifica a utilização destas técnicas na conceção de uma aplicação para a classificação automática de imagens de espécies biológicas.

A dimensão da Revista não nos permite aumentar o número de exemplos de aplicação de técnicas de *Data Science* nas áreas tradicionais das ciências exatas, mas muitos outros poderiam ser apresentados. Esperamos, contudo, que a pequena seleção incluída nesta edição da Revista consiga despertar no leitor a curiosidade por esta temática e o estimule na procura de outras leituras.

Votos de um Bom Ano!

Alexandre Lopes Magalhães

Universidade do Porto

---

# Classificação taxonómica em Biologia usando IA

Luís M. B. Lopes, Eduardo R. B. Marques

DCC/ FC/ Universidade do Porto

A ubiquidade dos telemóveis, a sua capacidade computacional e a elevada qualidade dos sensores fotográficos que os equipam abrem imensas possibilidades à participação dos cidadãos em atividades científicas, um conceito designado por Ciência Cidadã (*Citizen Science*). É neste contexto que surge o *Biolens*, um projeto para classificação automática de imagens de espécies biológicas usando técnicas de inteligência artificial.

As aplicações *web* desenvolvidas pelo projeto sugerem uma classificação taxonómica para um animal ou planta cuja fotografia é carregada pelo utilizador. Assim, um utilizador pode não só recolher informação relevante sobre uma espécie (por exemplo, fotografia, data, local), mas tem também a possibilidade de saber, em geral com um grau de confiança elevado, qual o nome científico do animal ou planta que acabou de registar. Atualmente, a aplicação inclui modelos de classificação para diferentes grupos taxonómicos da fauna e flora de Portugal: libélulas e libelinhas (*Dragonlens*), borboletas diurnas (*Lepilens*) e noturnas (*Mothlens*) e plantas (*Floralens*) (FIGURA 1). Como foi referido, a classificação das imagens é feita usando ferramentas de inteligência artificial, nomeadamente de *machine learning*, cujo objetivo é “ensinar” conceitos aos computadores por forma a que estes possam reter esse conhecimento e aplicá-lo noutros contextos.

No caso concreto do projeto *Biolens*, utilizam-se redes neuronais profundas (*deep learning*) que tentam simular em computador a forma como o nosso cérebro aprende e sintetiza conhecimento através de redes de neurónios que processam e trocam informação entre si.

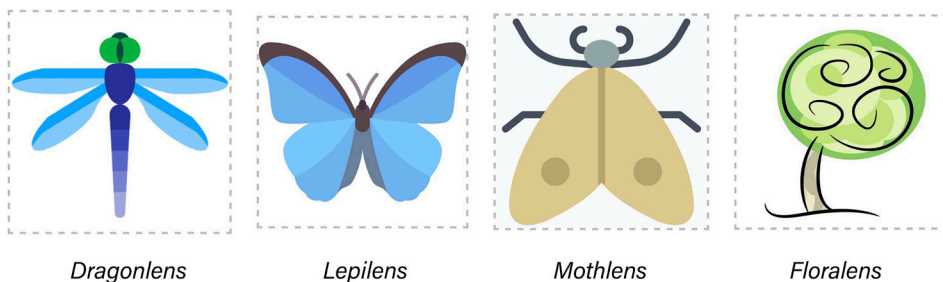


FIGURA 1. As classes taxonómicas atualmente suportadas pelo projeto *Biolens*.

No caso do *Mothlens*, por exemplo, o processo de aprendizagem passa por fornecer ao computador um conjunto de imagens de mariposas com a respetiva classificação taxonómica correta (indicada ou verificada por especialistas). Este conjunto funciona como uma referência para a rede neuronal aprender as diferenças e semelhanças entre as espécies representadas. Depois de um processo de treino, tipicamente bastante intensivo em termos computacionais e que pode demorar várias horas ou mesmo dias, a rede produz um programa que sintetiza o conhecimento adquirido. Este programa, designado "modelo", recebe imagens de mariposas e responde com sugestões de classificação (FIGURA 2).

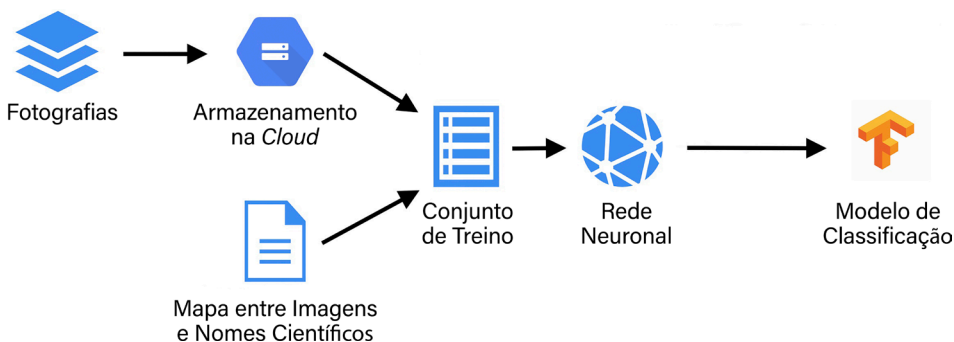


FIGURA 2. Criação do conjunto de treino, treino da rede neuronal profunda e geração do modelo de classificação.

Os modelos usados pelo *Biolens* foram treinados usando os serviços da *Google Cloud*, mais concretamente a plataforma *Google AutoML* e a biblioteca *open-source TensorFlow* para *machine learning*. As imagens para os conjuntos de treino foram extraídas principalmente do *iNaturalist Research-grade Observations* e do *Observation.org*, *Nature data from around the World datasets*, publicados no *Global Biodiversity Information Facility (GBIF)*, com imagens complementares extraídas de outros repositórios especializados, incluindo o grupo *Lepidoptera Portugal*, no *Facebook*, para as quais a classificação havia sido validada pelos especialistas, e o conjunto de imagens de excelente qualidade

identificadas por especialistas do projeto *FloraOn*, gentilmente disponibilizado pela Sociedade Portuguesa de Botânica.

A página do *Biolens* agrega as várias aplicações *web* que utilizam os modelos supracitados para classificar de forma automática imagens de animais ou plantas submetidas por utilizadores. Para cada imagem, a aplicação selecionada gera uma lista de sugestões de classificação ordenadas por ordem decrescente de confiança. Estes resultados são apresentados aos utilizadores numa página *web* gerada dinamicamente e guardados numa base de dados no servidor, juntamente com a imagem e a data de submissão (FIGURA 3).

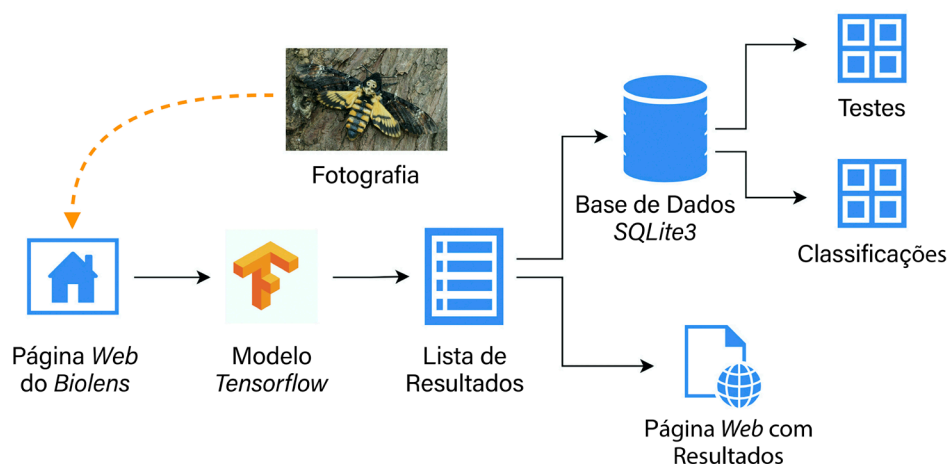


FIGURA 3. A arquitetura da aplicação *Web Biolens*.

A aplicação *Web* é albergada num computador servidor modesto com apenas 4 GB de RAM e 2 CPUs. Tal é possível graças ao uso dos modelos gerados pelo *Google AutoML* na variante *TensorFlow Lite*, um formato otimizado para a utilização em dispositivos com recursos computacionais limitados, não só computadores com recursos modestos, mas também dispositivos como telemóveis ou *tablets* da forma discutida mais abaixo.

Os modelos são, naturalmente, falíveis. Defeitos de construção no conjunto de treino, imagens mal identificadas no dito conjunto ou imagens de baixa qualidade submetidas pelos utilizadores finais, por exemplo, podem levar a que os modelos produzam sugestões incorretas. Os resultados devem por isso ser encarados como “sugestões informadas” e não como classificações definitivas, as quais devem sempre passar pelo crivo de um ou mais especialistas humanos (FIGURA 4).

No seguimento do desenvolvimento da página do *Biolens* e das várias aplicações *web* aí residentes, exploramos também a possibilidade de oferecer o *Biolens* através de uma aplicação móvel.

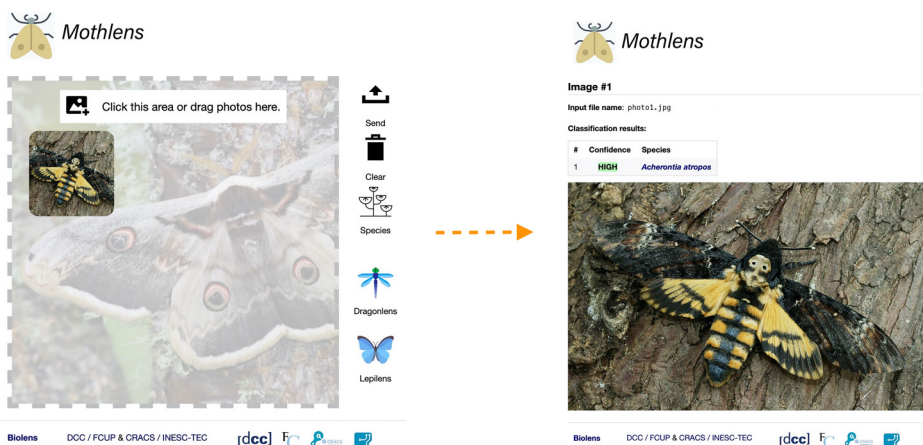


FIGURA 4. Interface da aplicação *Web* e exemplo de classificação de imagem.

Apesar das aplicações *web* poderem ser utilizadas em telemóveis ou *tablets* através de um *browser*, uma aplicação móvel poderia fornecer um conjunto extra de funcionalidades ao utilizador (por exemplo: utilização *online*, manutenção de um registo de avistamentos no dispositivo). O conceito foi concretizado para apenas um dos modelos do *Biolens*, no caso o *Lepilens*, com o desenvolvimento de uma aplicação para *iOS* (FIGURA 5).



FIGURA 5. Interface da aplicação *iOS*.

A aplicação permite que fotografias obtidas com o telemóvel ou carregadas de outros dispositivos ou da *web*, possam ser classificadas pelo modelo *TensorFlow Lite*. Um dos requisitos do desenvolvimento especificava que a aplicação deveria poder funcionar sem

estar ligada à *internet*, uma situação comum para quem realiza trabalho de campo. Assim sendo, todos os avistamentos de borboletas são documentados, com data e hora, coordenadas GPS do local e notas opcionais do utilizador, e guardados numa base de dados *SQLite3*. Uma vez ligado à *internet*, o utilizador poderá optar por descarregar esses dados para o seu próprio computador ou contribuir com eles para uma plataforma existente, por exemplo, o *iNaturalist* ou o *Observation.org*.

A arquitetura da aplicação móvel permite que seja facilmente estendida para incluir os restantes modelos suportados pelo *Biolens*. Este trabalho, ainda que menos criativo e mais sistemático, faz parte dos planos futuros de desenvolvimento do projeto.

---

# Big data

## coleções biológicas, museus e a biogeografia marinha da Macaronésia

Sérgio P. Ávila, Carlos Melo, Patrícia Madeira

CIBIO-Açores/ InBIO/ BIOPOLIS

Para quem, como os autores deste texto, trabalha em ilhas oceânicas e tenta decifrar os processos e padrões evolutivos que aqui decorrem atualmente, bem como os que decorreram ao longo do tempo geológico, assume particular relevo o conhecimento que é necessário ter acerca da ontogenia de cada ilha, ou seja, da história evolutiva de cada uma das ilhas estudadas, bem como da história de cada uma das espécies que, ao longo dos milhões de anos de tempo útil de uma ilha, foi capaz de lá chegar, aí viver e, pelo menos algumas, evoluir, originando espécies endémicas. A grande maioria das ilhas oceânicas passa por vários estados ontogenéticos, com uma história evolutiva que se pode resumir da seguinte forma: 1) fase de monte submarino; 2) fase de ilha emergente; 3) fase de construção de edifícios vulcânicos; 4) fase erosiva; 5) fase de atol (dependente da ocorrência de corais); 6) fase de *guyot*. Nas fases 1 a 3 predomina a construção da ilha, com altas taxas de emissão de produtos vulcânicos, ao passo que a partir do final da fase 3 predominam os processos erosivos que atuam a par dos usuais processos de subsidência do(s) edifício(s) vulcânico(s), fazendo com que, progressivamente, a ilha se transforme num monte submarino de topo aplanado, conhecido pelo nome de *guyot*.

Esta descrição é importante pois o impacto, sobre os seres vivos, da história evolutiva de uma ilha oceânica vulcânica é diferente, consoante trabalhemos com organismos marinhos ou com organismos terrestres. É fácil de entender que os organismos marinhos podem colonizar uma proto-ilha quando esta ainda está na fase de monte submarino, ao passo que, somente quando há área emersa, é que há substrato disponível para ser colonizado por organismos terrestres. Por outro lado, quando uma antiga ilha é arrasada por agentes erosivos e passa à fase de monte submarino, os organismos terrestres são extirpados, por vezes milhares de anos antes do desaparecimento dos organismos marinhos litorais, os quais só desaparecem quando, no processo de subsidência, os picos mais elevados do monte submarino descem abaixo de cerca de 50 metros.



Para além do conhecimento geológico detalhado, que é proporcionado pelas diferentes histórias evolutivas das ilhas, ilhéus e montes submarinos que formam um determinado arquipélago, é importante para os biólogos marinhos darem resposta às seguintes questões: 1) De onde vieram os colonizadores iniciais que povoaram a primeira ilha/monte submarino desse arquipélago? 2) Quando atingiram essa primeira ilha pela primeira vez? 3) O que lhes sucedeu ao longo do tempo evolutivo? Todas estas questões estão atualmente enquadradas por um ramo fascinante da Biologia que dá pelo nome de Biogeografia Insular. Tendo por base os estudos iniciais de Wallace, em 1880, a Biogeografia Insular sofreu um enorme incremento com a obra seminal de MacArthur & Wilson de 1967, e teve recentemente desenvolvimentos extraordinários para os investigadores que trabalham com organismos insulares terrestres, bem como para aqueles que estudam organismos insulares marinhos.

Enquanto que o estudo dos fósseis responde à segunda pergunta (o “Quando chegaram à ilha”), já a sistemática e taxonomia efetuada por reputados especialistas responde à terceira questão, a qual lida com os fenómenos de especiação. Falta a resposta à primeira questão, na qual se abordam as origens dos colonizadores. É aqui que entra, mais uma vez, a Biogeografia Insular e o estudo das relações biogeográficas dos organismos terrestres, bem como dos marinhos. Para responder cabalmente a esta importante questão, é necessário conhecer em detalhe as listas de espécies, validadas por especialistas, dos organismos selecionados (por exemplo, moluscos marinhos gastrópodes, equinodermes, algas) em cada local a comparar. É aqui que adquirem relevo as coleções curadas em Museus, as quais, após compiladas, acabam por preencher os requisitos para serem incluídas nos processos de *Big Data*.

Por definição, *Big Data* é uma área do conhecimento que se dedica ao estudo de um grande volume de dados. Nos últimos anos, temos assistido a uma cada vez maior utilização de *Big Data* pelos biólogos, seja através da construção de grandes bases de dados taxonómicas para organismos recentes (por exemplo: WoRMS — *World Register of Marine Species*) ou passados (isto é, de fósseis; ex: *PaleoBiology Database*), seja, por exemplo, através da construção de árvores filogenéticas utilizando genomas completos de vários organismos, o que requer capacidade computacional cada vez mais substancial. Consequência disto, é a procura crescente por parte do mercado científico e empresarial de cientistas (programadores, algoritmistas, engenheiros de sistemas) capazes de darem respostas aos problemas com que os investigadores se deparam na análise de volumes de dados cada vez maiores.

A par destes desenvolvimentos recentes nesta nova área do conhecimento que é o *Big Data*, assiste-se no nosso país a um recrudescimento do investimento em Museus, com o correspondente reconhecimento da importância e valorização da curadoria das coleções biológicas, paleontológicas e geológicas que são mantidas a cargo destas instituições. Dou aqui dois exemplos de sucesso: o Museu de História Natural e da Ciência da Universidade do Porto, instituído formalmente em 2015 e em funcionamento desde 2019 e a Casa dos Fósseis, inaugurada em 2016 em Vila do Porto, onde é divulgado ao público o extraordinário e

internacionalmente relevante património paleontológico existente na ilha de Santa Maria, no arquipélago dos Açores (FIGURA 1).



FIGURA 1. O Museu Casa dos Fósseis em Vila do Porto, na ilha de Santa Maria (Açores).

Estas coleções são fundamentais, não só como repositório da biodiversidade atual e passada — os grandes Museus são muito visitados por cientistas porque possuem enormes coleções e muito exemplares-tipo (os quais correspondem aos indivíduos que serviram para a descrição formal de uma dada espécie) — mas também pelo relevante papel educativo que desempenham ao promoverem a Ciência e o conhecimento.

Vamos exemplificar tudo o que atrás foi dito, através de um caso de estudo. A Macaronésia é uma ampla área geográfica localizada no Atlântico norte, onde estão cinco arquipélagos: Açores, Madeira, Selvagens, Canárias e Cabo Verde. No total, nesta região, existem 31 ilhas de maior dimensão e muitos ilhéus e montes submarinos (FIGURA 2).

A palavra “Macaronésia” foi cunhada pela primeira vez pelo botânico britânico Philip Barker-Webb (ca. 1845) para abranger os arquipélagos da Madeira, Selvagens e Canárias. Mais tarde, Engler, em 1879, incluiu os Açores na região da Macaronésia e Dansereau, em 1961, alargou ainda mais o conceito, ao incluir as ilhas de Cabo Verde. Alguns autores consideram que outras regiões apresentam também um número significativo de *taxa* comuns com as ilhas da Macaronésia, nomeadamente algumas áreas da Península Ibérica

e algumas zonas costeiras do noroeste africano adjacente. Embora o termo “Macaronésia” tenha sido usado com diferentes significados, a inclusão de Cabo Verde é um assunto particularmente controverso, em especial para os especialistas no meio marinho.

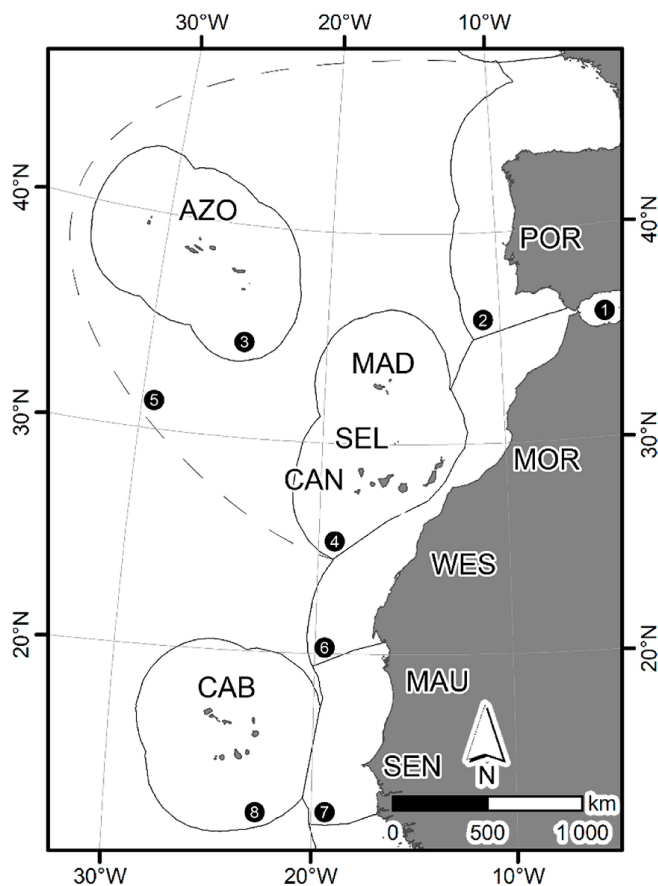


FIGURA 2. Enquadramento geográfico e biogeográfico da Macaronésia e das costas atlânticas de África e da Europa. As linhas pretas representam as ecorregiões biogeográficas de acordo com Spalding *et al.* (2007) e a linha pontilhada representa a província biogeográfica lusitana, segundo Freitas *et al.* (2019). 1) Mar de Alboran; 2) Ecorregião da Plataforma Atlântica Sul-Europeia; 3) Ecorregião dos Açores; 4) Ecorregião Webbnésia (que integra os arquipélagos da Madeira, Selvagens e Canárias); 5) Província biogeográfica Lusitana; 6) Ecorregião de ressurgência Saariana; 7) Ecorregião de ressurgência do Sahel; e 8) subprovíncia biogeográfica de Cabo Verde. A subprovíncia de Cabo Verde e a ecorregião de ressurgência do Sahel pertencem à Província biogeográfica de Transição da África Ocidental. AZO — Açores; MAD — Madeira; SEL — Selvagens; CAN — Canárias; CAB — Cabo Verde; POR — Portugal; MOR — Marrocos; WES — Saara Ocidental; MAU — Mauritânia e SEN — Senegal.

Vários estudos baseados em peixes marinhos costeiros e gastrópodes de Cabo Verde concluíram que a estrutura da comunidade e as relações biogeográficas do biota marinho deste arquipélago diferem significativamente dos outros arquipélagos da Macaronésia, mas, até recentemente, esta questão não tinha sido abordada de forma quantitativa, estruturada e através de um estudo multifilo.

Enquadrado pelas premissas do modelo dinâmico *Sea-Level Sensitive* (SLS) da biogeografia insular marinha e fundamentado nos processos e padrões evolutivos biogeográficos insulares, foi recentemente publicado o estudo mais abrangente, até à data, acerca da biogeografia marinha macaronésica. Coordenado por um dos autores (SPA), este trabalho permitiu efetivar, do ponto de vista marinho, a primeira análise comparativa multifilo para reavaliar este debate e, assim, testar se a Macaronésia era (ou não) uma unidade biogeográfica coerente. Para tal, efetuou-se uma compilação exaustiva de listas de espécies, avalizadas por especialistas, dos seis grupos nativos marinhos macaronésicos mais estudados: peixes costeiros, equinodermes, moluscos gastrópodes, crustáceos decápodos braquiúros, anelídeos poliquetas e macroalgas. Foram ainda incluídas outras regiões próximas (por exemplo, Portugal continental, o Mediterrâneo, ilhas Britânicas, entre outras), as quais serviram para melhor elucidar as relações biogeográficas entre os vários arquipélagos que formam a Macaronésia, e entre a Macaronésia e as outras regiões selecionadas. No total, foram incluídas 7 492 espécies marinhas, sendo que 3 737 destas ocorrem nos arquipélagos da Macaronésia. A principal conclusão deste trabalho levou à criação de uma nova unidade biogeográfica — a ecorregião da *Webbnesia* — assim homenageando Philip Barker-Webb e, pela primeira vez baseada em dados quantitativos sonantes, permitiu individualizar e separar os arquipélagos dos Açores e de Cabo Verde dos restantes.

Assim, Cabo Verde deverá ser considerado como uma subprovíncia biogeográfica isolada, os Açores como uma ecorregião e Madeira, Selvagens e Canárias formam uma nova ecorregião, denominada *Webbnesia*. É este o poder do uso de *Big Data*!

---

# Bases de dados de Proteínas

Ana Oliveira

LAQV/ REQUIMTE/ DQB/ FCUP

A biologia, a química e a bioquímica têm-se vindo a transformar em ciências ricas em dados. Atualmente, geram-se enormes quantidades de dados para estruturas tridimensionais, dados de atividade e funções, e particularmente sobre a estrutura primária de proteínas. Com o aumento de dados, cresceu também a necessidade de armazenar, integrar e comunicar estes grandes conjuntos de informação proteómica como um todo. Neste pequeno texto, focamo-nos na *UniProt*, a base de dados universal de proteínas, que reúne em si informação de várias outras bases de dados e, por isso, permite aos investigadores o acesso rápido e fácil a quantidades de informação massiva.

## A necessidade de uma Base de Dados em Proteínas

Não podemos falar em bases de dados sem primeiro introduzir, brevemente, a Bioinformática. Esta emergente área interdisciplinar combina o poder computacional e técnicas informáticas, matemáticas e de estatística com a biologia, a química, a farmácia e a medicina (FIGURA 1). Margaret O. Dayoff foi, sem dúvida, a mãe da bioinformática e a pioneira no desenvolvimento das suas técnicas. Foi ela quem pela primeira vez aplicou métodos matemáticos e computacionais à bioquímica, criando as primeiras bases de dados de proteínas e ácidos nucleicos. Dayoff deu origem ao atual código de uma letra dos aminoácidos, numa tentativa de reduzir o volume de dados, para que os (super) computadores da época pudessem armazenar mais informação. Desde então, a informática sofreu grandes avanços e a capacidade de armazenamento de dados aumentou exponencialmente. Em paralelo, ocorreram grandes avanços nas ciências genómicas e nas tecnologias de sequenciação de próxima geração, as quais permitiram, nas últimas décadas, descobrir informação genómica e proteómica de um variado número de organismos. Como consequência, o número de proteínas sequenciadas e arquivadas em bases de dados tem aumentado drasticamente nos últimos anos.

## Bioinformática

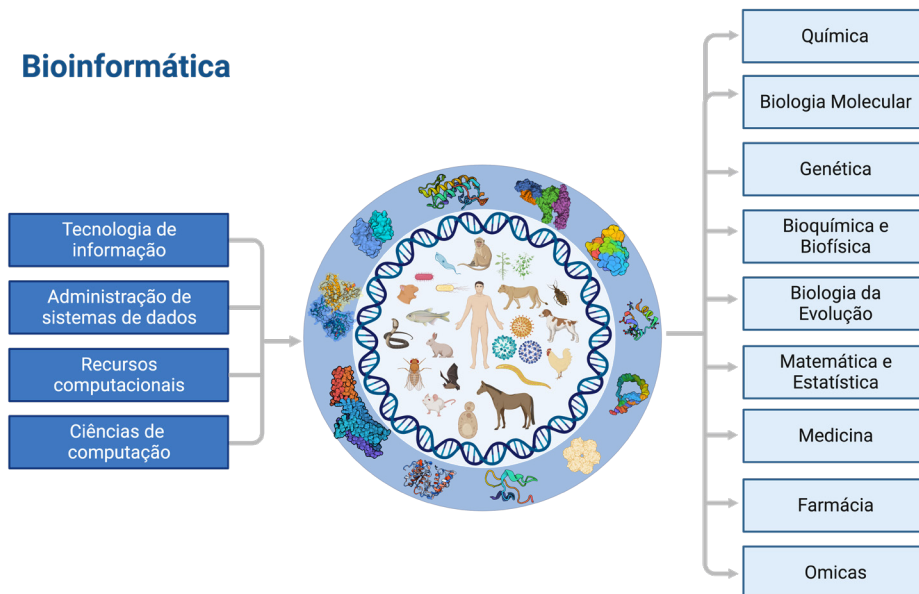


FIGURA 1. Integração da informação numa base de dados bioinformática. À esquerda encontram-se os sistemas técnicos que suportam a bioinformática, à direita as disciplinas científicas relacionadas com a bioinformática. No centro mostramos a informação vista como um todo, isto é, como o DNA dá origem a distintas proteínas e a sua diversidade taxonômica.

As proteínas são macromoléculas de particular interesse porque ocupam o campo molecular intermédio entre o gene e a transcrição, e desempenham um papel fundamental na estrutura e organização molecular e celular. Além disso, a maioria dos processos fisiológicos e patológicos manifesta-se a nível proteico, e, portanto, torna-se emergente a utilização de técnicas proteômicas e bioinformáticas de alto rendimento para alcançar uma melhor compreensão da biologia molecular básica e das suas alterações em casos de doenças. A comparação entre proteínas, ou entre famílias de proteínas, fornece informação sobre a sua relação no genoma ou com outras proteínas relativas em outras espécies, e também permite a identificação do porquê da sua implicação em doenças. Assim, a integração de dados sobre várias proteínas permite obter muito mais informação que aquela dada por uma simples proteína isolada. A riqueza dos novos dados proteómicos permite aos investigadores formular perguntas biológicas complexas e obter novos conhecimentos científicos que suportem estas novas hipóteses orientadas por dados. Nos últimos anos, foram desenvolvidas muitas bases de dados bioinformáticas relacionadas com proteínas, instalações de consulta e ferramentas de *software* de análise de dados para organizar e fornecer anotações biológicas de proteínas que apoiem análises sequenciais, estruturais, funcionais e evolutivas no contexto da bioquímica de redes e sistemas.

## A evolução de bases de dados de proteínas

A necessidade de reunir informação sobre a sequência de proteínas remonta a 1965, quando se publica o *Atlas da Sequência e Estrutura de Proteínas*, editado por Margaret O. Dayhoff. Este trabalho reunia apenas a informação sequencial de 66 proteínas diferentes, agrupadas em 10 categorias e continha aspetos muito básicos como a formação de pontes persulfureto ou os locais de interação de cofatores. Nesta altura procurava-se já relacionar a sequência da proteína com a sua estrutura e obter pequenas conclusões através da comparação com proteínas relativas.

Depois, em 1971, surge o *Banco de dados de Proteínas* (PDB), um repositório público e livre focado na biologia estrutural. O PDB contém toda a informação conhecida sobre a estrutura tridimensional de péptidos, proteínas, ácidos nucleicos e complexos formados entre macromoléculas biológicas ou entre macromoléculas e pequenos compostos orgânicos ou mesmo fármacos. Atualmente, contém mais de 198 mil estruturas determinadas experimentalmente por técnicas de raios-X, NMR e microscopia eletrónica, e cerca de 1 milhão de modelos obtidos computacionalmente por técnicas de bioinformática.

Em 1984, o atlas de sequência de proteínas deu lugar à primeira base de dados mundial de sequências de proteínas classificadas e funcionalmente anotadas — PIR-PSD. O PIR-PSD classifica as sequências de proteínas com base no conceito de superfamília. Estas sequências são também classificadas com base no domínio da homologia e nos motivos da sequência, ou seja, os domínios de homologia podem corresponder a blocos de construção evolutivos, enquanto os motivos de sequência representam sítios funcionais ou regiões conservadas. Esta abordagem na classificação permite uma compreensão mais completa da relação função-estrutura da sequência e da sua evolução.

Mais tarde, em 1986, nasce, na Universidade de Genebra, a primeira base de dados que visa reunir a informação detalhada sobre uma dada proteína — a SWISS-PROT DB. Para cada proteína incluída nesta base de dados estão disponíveis três tipos de informação: (i) a sequência da proteína; (ii) informação bibliográfica; e (iii) a sua informação taxonómica. Sempre que possível inclui-se também informação sobre a função da proteína, isoformas, modificações pós-tradução, informação detalhada sobre a estrutura, como os domínios e resíduos catalíticos, mutações, estrutura secundária e quaternária, similaridade com outras proteínas, doenças relacionadas e conflitos na sequência.

Finalmente, em 2002, forma-se o consórcio da *UniProt*. Este reúne equipas de investigação do Instituto Europeu de Bioinformática (EBI), do Instituto Suíço de Bioinformática (SIB) e do Recurso de Informação sobre Proteínas (PIR).

### **UniProt**

A *UniProt*, ou base universal de proteínas, é a maior base de dados mundial com informação sobre proteínas. Continua, em novembro de 2022, cerca de 289 milhões de sequências

únicas, que incluem sobretudo proteínas de bactérias e organismos eucariotas, e quase 283 mil proteomas distintos, compreendendo mais de 205 milhões de combinações de aminoácidos. Esta base de dados abrangente e não redundante de sequências de proteínas arquivadas, reúne a informação disponível em todos os principais recursos acessíveis ao público como a Swiss-Prot, a PIR e o PDB, referidos acima, e também a informação disponibilizada em outras bases de dados como o *TremBL*, *RefSeq*, *GenBank* e bases de dados médicas com informação sobre pacientes. A cada sequência de proteínas, ou entrada, em linguagem informática, é atribuído um código identificador único e um nome que identifica cada proteína, tendo em conta o gene que a codifica e o organismo de que provém. Por exemplo, a ciclooxigenase-2 (COX2) humana, uma enzima extremamente estudada pelas suas implicações em processos inflamatórios, tem o código identificador P35354 e o nome PGH2\_HUMAN. Se o leitor pretender procurar todas as sequências de ciclooxigenase-2 conhecidas, até à data, bastaria aceder à página *web* da base de dados e no local de busca (*Find your protein*), identificado por (A) na FIGURA 2, escrever *cyclooxygenase 2*. Em alguns segundos, a base de dados encontra 17 235 resultados, dos quais 17 200 são sequências consideradas não revistas e apenas 35 são entradas revistas. Mas então que diferença há entre a informação nas entradas? As sequências que tiveram a sua origem na base de dados de SWISS-PROT (cerca de 570 mil no total) são consideradas entradas revistas porque contêm informação validada por especialistas, que para além da informação da estrutura primária das proteínas, inclui também isoformas e variantes, informação detalhada sobre a sua função, mecanismo enzimático (quando aplicável), estrutura tridimensional, interações com outras proteínas ou DNA e a sua implicação em vias metabólicas e em doenças. Finalmente, e como já foi referido anteriormente, para uma melhor compreensão da informação, não devemos ver as proteínas isoladas, mas agrupadas com base na sua similaridade e identidade sequencial. Os bioinformáticos desenvolveram, por isso, técnicas de comparação da identidade estrutural de proteínas. Não poderíamos aqui focar-nos em todas as técnicas existentes, mas destacamos duas das mais importantes: *Basic Local Alignment Search Tool* (BLAST) e CLUSTAL, implementadas na base de dados. A primeira procura regiões de proteínas com sequência semelhante, enquanto a segunda compara as semelhanças entre 2 ou mais sequências. Assim, na *UniProt*, sempre que possível, as sequências das proteínas são agrupadas em três grupos distintos: 100% UniRef100 (se as proteínas compartilham 100% da sua identidade sequencial); UniRef90 (quando a identidade sequencial  $\geq 90\%$ ) ou UniRef50 (identidade  $\geq 50\%$ ). A cada um destes grupos é atribuído um representante, isto é, a sequência proteica do grupo melhor caracterizada, até à data, um grupo de proteínas relacionadas.



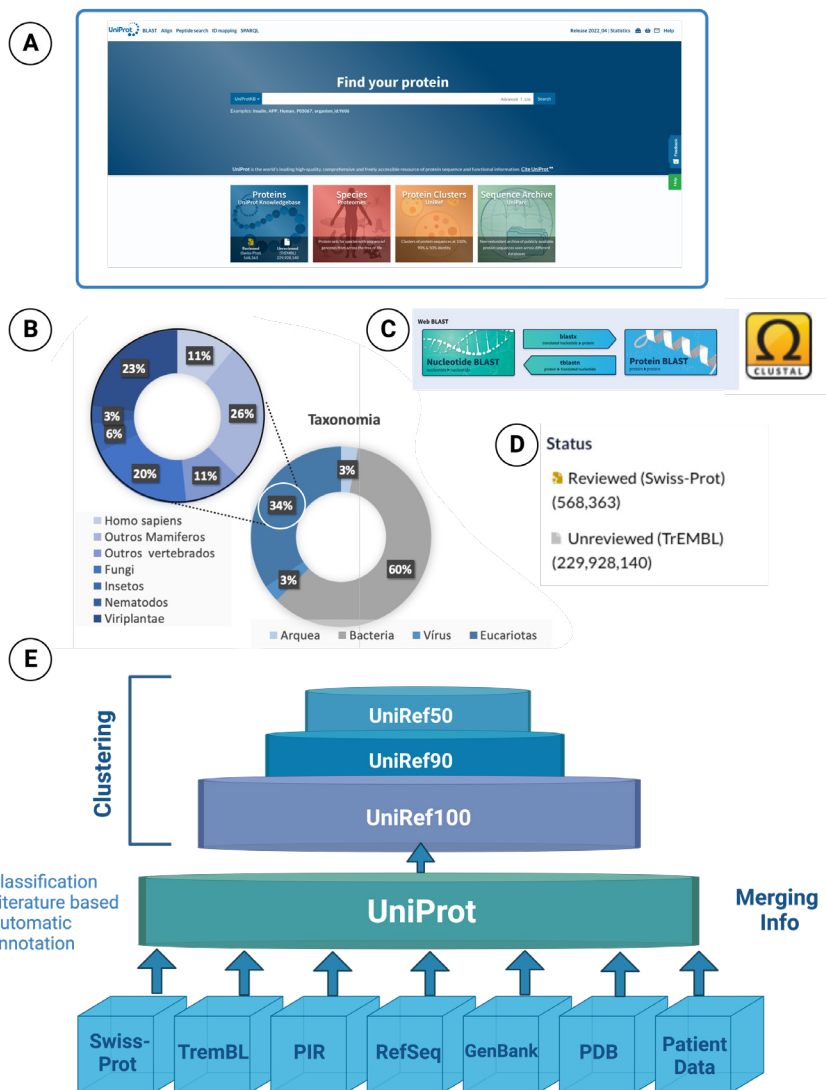


FIGURA 2. Integração de dados de seqüências de proteínas e a sua organização. (A) Vista geral do motor de busca da base de dados. (B) Distribuição da seqüência de proteínas de acordo com a sua classificação taxonômica. (C) Principais ferramentas bioinformáticas para análise de seqüências proteicas. (D) Número de entradas incluídas na *UniProt* em novembro de 2022. (E) Organização da informação disponível na base de dados.

Na versão *online* deste artigo, pode encontrar um guião para um trabalho prático que permite ao estudante realizar uma investigação autêntica usando a base de dados *UniProt*, de forma análoga como muitos cientistas levam a cabo, hoje em dia, na sua investigação.

---

# Imagens de satélite, geologia e IA

Joana Cardoso-Fernandes, Alexandre Lima, Ana Cláudia Teodoro  
FCUP/ICT/ Universidade do Porto

**A geologia, enquanto Ciência, procura fundamentalmente compreender os processos que levaram à formação do planeta Terra e à sua evolução contínua ao longo de milhões de anos, incluindo o ciclo de formação das rochas e a sua relação com a atmosfera, a biosfera e a hidrosfera, a colisão e rutura dos continentes como resultado da tectónica de placas, mas também compreender a formação de recursos geológicos e como explorá-los.**

A guerra na Ucrânia trouxe-nos uma nova perspetiva sobre as políticas energéticas, a transição energética e até sobre a soberania energética com os países que tentam encontrar soluções para a crise energética sem dar um passo atrás nos objetivos de desenvolvimento sustentável para alcançar a descarbonização global e combater as alterações climáticas.

Embora, como cientistas, possamos ter pouca influência sobre os decisores, podemos, no entanto, fazer a nossa parte no processo de transição energética. As energias renováveis, como a energia solar ou eólica, requerem uma multiplicidade de metais nos processos de fabrico e, para a maioria deles, a recuperação e reciclagem de peças antigas ainda não é possível, e pode não ser suficiente num futuro próximo. Portanto, a transição energética coloca uma grande pressão e cria uma grande procura de metais que ocorrem naturalmente na crosta terrestre, muitas vezes em conjuntos minerais estáveis em depósitos minerais. Estes depósitos minerais podem encontrar-se numa grande variedade de rochas e configurações. Um exemplo deste tipo de rocha são os pegmatitos, que podem ser explorados para minerais e elementos estratégicos para a digitalização e transição energética, como o quartzo de alta pureza para ser usado em ótica, fibras, semicondutores, energia solar e o lítio para mobilidade elétrica e armazenamento de energias renováveis. Se existirem na Europa, menor pegada carbónica terão ao serem explorados e convertidos no velho continente.

Durante as últimas quatro décadas, os dados de deteção remota, nomeadamente imagens de satélite, têm sido uma importante ferramenta na prospeção geológica para encontrar diferentes tipos de depósitos minerais, especialmente em áreas remotas e/ou de difícil acesso.

Recentemente, dados de detecção remota e, conseqüentemente, imagens de satélite, têm sido usados para desenvolver novas ferramentas de prospeção de baixo impacto, menos invasivas para as populações, flora e fauna. Estes instrumentos foram melhorados ao longo dos anos para a identificação de pegmatitos, uma importante fonte de recursos minerais.

Assim, é fundamental saber quais os dados (imagens de satélite) disponíveis para o desenvolvimento de tais abordagens e compreender as diferenças, vantagens e desvantagens de cada produto de satélite (FIGURA 1). Os sensores dos satélites medem a radiação eletromagnética em gamas específicas de comprimento de onda, que é quantificada e convertida numa imagem digital (*raster*), chamada banda, composta por uma matriz bidimensional de elementos de imagem (*pixels*). A intensidade de cada *pixel* corresponde a um valor numérico discreto armazenado como níveis cinzentos distintos, com níveis mais brilhantes para as mais altas intensidades. Cada banda tem uma determinada dimensão (*frame*), correspondente à área na superfície da Terra capturada por uma imagem, e cada *pixel* corresponderá a uma área menor (geralmente quadrada) dentro da *frame* da imagem. Com esta informação podemos comparar imagens de satélite de acordo com (i) a resolução espectral (quanto mais estreita for a faixa de comprimento de onda da banda e o maior número de bandas, melhor a resolução); (ii) a resolução radiométrica, quanto maior for o número de *bits*, maior o número de níveis cinzentos disponíveis e a resolução; e (iii) a resolução espacial (quanto menor a área representada por um *pixel*, maior é a resolução). Há também um outro tipo de resolução, específica aos dados de satélite, que é a resolução temporal, ou seja, o tempo necessário para revisitar a mesma região da Terra.

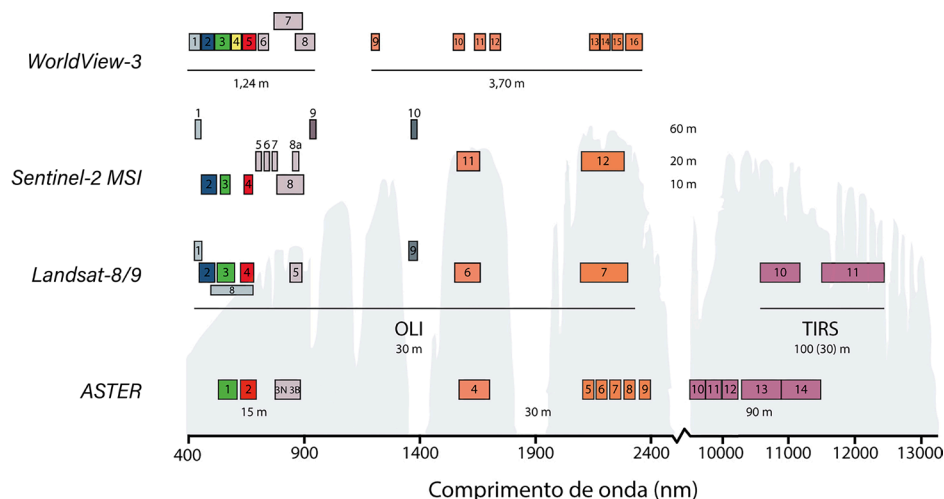


FIGURA 1. Comparação da resolução espectral e espacial de sensores de satélite distintos. O satélite *Landsat-8* contém dois sensores a bordo: *Operational Land Imager* (OLI) e *Thermal Infrared Sensor* (TIRS). O *Sentinel-2* transporta o sensor *Multispectral Instrument* (MSI).

O sensor *ASTER* a bordo do sistema de observação da Terra (EOS) da NASA, lançado em dezembro de 1999, fornece dados cruciais para os geólogos delinearem depósitos minerais ao longo dos anos devido à sua resolução espectral média-alta, com dados armazenados em 14 bandas cuja resolução espacial e radiométrica varia com o comprimento de onda (FIGURA 1). As bandas 1–3 têm uma resolução espacial de 15 m com quantização em 8 *bits* (256 níveis de cinza). As bandas 4–9, também com quantização em 8 *bits*, têm uma resolução de 30 m. As bandas 10–14 têm uma resolução espacial de 90 m com quantização em 12 *bits* (níveis cinzentos de 0 a 4 095). Infelizmente, houve uma falha num dos módulos do *ASTER* em 2008, afetando as bandas 4–9 para futuras utilizações.

O *Landsat-8*, lançado em 2013, e o seu sucessor *Landsat-9*, lançado recentemente, em 2021, apresentam nove bandas espectrais com uma resolução espacial de 30 m para bandas 1–7 e 9, 15 m para a banda 8 (pancromática), e duas bandas térmicas (10 e 11) que são recolhidas a 100 m mas reamostradas para 30 m de resolução espacial (FIGURA 1). Enquanto as bandas do *Landsat-8* apresentam os dados em 12 *bits* (ou seja, 4 096 tons de cinza), o *Landsat-9* tem uma resolução radiométrica melhorada de 14 *bits* (16 384 níveis de cinza distintos para um dado comprimento de onda).

A missão *Sentinel-2*, composta pelos satélites gémeos *Sentinel-2A* e *Sentinel-2B*, no âmbito do *Programa Copernicus da União Europeia*, fornece produtos de satélite com 13 bandas espectrais (FIGURA 1): quatro bandas a 10 m de resolução espacial, seis bandas a 20 m e três bandas a 60 m de resolução espacial, todas com uma resolução radiométrica de 12 *bits* (4 096 níveis de cinzento).

Todos os satélites mencionados fornecem dados abertos a pedido, uma vez que as missões correspondentes são lideradas por agências espaciais governamentais. No entanto, os satélites comerciais estão a ser cada vez mais utilizados devido às suas melhores especificações/resoluções. Um exemplo é o satélite *WorldView-3* da *Maxar*, lançado em 2014, um sensor multiespectral de muito alta resolução espacial que adquire dados em 11 *bits* (2 048 níveis de cinzento) em oito bandas recolhidas a 1,24 m de resolução espacial e dados de 14 *bits* (16 384 níveis de cinzento) em oito bandas a 3,70 m de resolução (FIGURA 1), embora os dados sejam reamostrados para uma resolução inferior para distribuição comercial. As imagens são armazenadas como inteiros de 16 *bits* (65 536 níveis de cinzento) ou são reduzidas a 8 *bits* (256 níveis de cinzento) para reduzir os tamanhos dos ficheiros e melhorar o desempenho do processamento de imagem.

Portanto, se considerarmos um produto *Sentinel-2* (290 km<sup>2</sup>) onde todas as bandas foram reamostradas para uma resolução de 10 m, cada banda será composta por 5 490 x 5 490 (ou seja, 30 140 100 *pixels* por banda) totalizando 391 821 300 *pixels* nas 13 bandas. Enquanto isso, se considerarmos uma imagem *Worldview-3* cobrindo 100 km<sup>2</sup>, com todas as bandas cobrindo a mesma gama de comprimentos de onda de *Sentinel-2* (FIGURA 1) reamostrada para uma resolução de 2 m, então cada banda será

composta por 26 403 836 *pixels*, totalizando 422 461 376 *pixels* em todas as 16 bandas. A FIGURA 2 apresenta uma comparação entre a resolução espacial e os respectivos cubos de dados para o campo pegmatítico de Tysfjord na Noruega, uma das áreas de estudo do projeto GREENPEG.

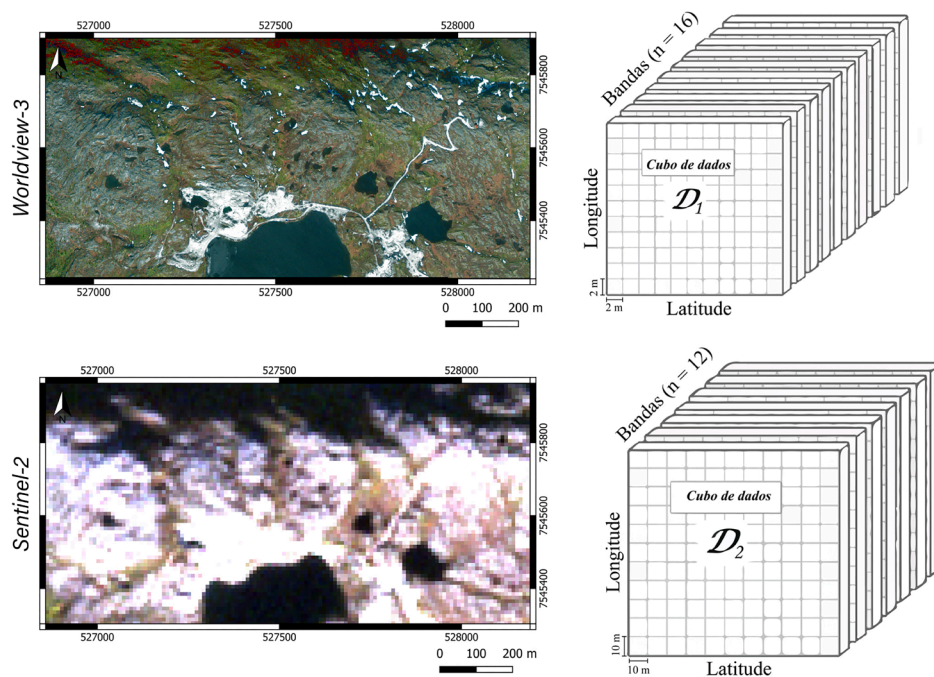


FIGURA 2. A) Comparação da resolução espacial dos satélites *Worldview-3* e *Sentinel-2* sobre um pegmatito conhecido em Tysfjord, Noruega. B) Comparação dos respectivos cubos de dados.

Assim, é claro como apenas dois produtos diferentes com resolução espacial média-alta e muito alta podem criar conjuntos de dados gigantescos com milhões de entradas de dados. No entanto, há outro tipo de sensor caracterizado por uma resolução espectral muito elevada: o sensor hiperespectral. Em 2019, a Agência Espacial Italiana lançou o satélite hiperespectral *PRISMA* que possui 238 bandas na região de comprimento de onda entre os 400 e os 2 500 nm com uma resolução espacial de 30 m. Com uma largura de 1 254 *pixels* por uma altura de 1 239 *pixels*, cada banda é composta por 1 553 706 *pixels*, e o total das 238 bandas resulta em 369 782 028 *pixels*. A FIGURA 3 apresenta uma pequena comparação entre o número de bandas na região de 400–1 000 nm para os satélites *PRISMA* e *Sentinel-2*, bem como uma comparação da resolução espacial de ambos os produtos de satélite na região de St. Austell na Cornualha (Reino Unido), muito semelhante geologicamente à maior mina de lítio a céu aberto a abrir ainda nesta década, que foi recentemente anunciada no Maciço Central Francês. A FIGURA 3B) apresenta o cubo de dados tridimensional para o produto *PRISMA* que cobre a área de St. Austell.

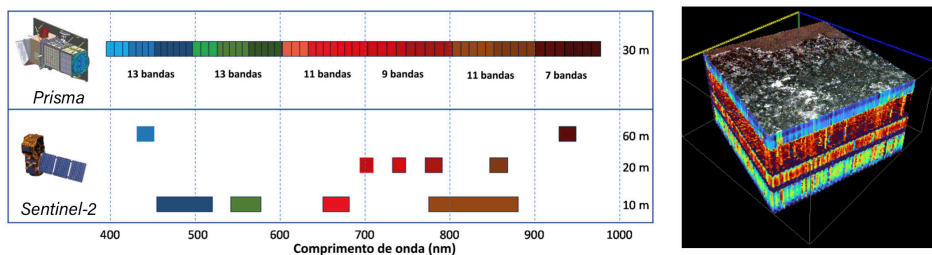


FIGURA 3. A) Comparação da resolução espectral dos satélites *PRISMA* e *Sentinel-2* na região de 400–1000 nm<sup>2</sup>. B) Cubo de dados *PRISMA* para uma área de estudo em St. Austell, Cornwall (Reino Unido). C) Comparação da resolução espacial dos satélites *PRISMA* (topo) e *Sentinel-2* (base) na região de St. Austell.

É evidente que não só é um desafio escolher o produto satélite certo, tendo em conta as características de uma área de estudo e os resultados esperados, mas também encontrar formas rápidas e precisas de lidar com esses grandes conjuntos de dados. Os algoritmos clássicos de processamento de imagem utilizados em aplicações geológicas são: (i) combinações RGB; (ii) rácios de bandas; (iii) análise principal dos componentes; (iv) *matched filtering* (MF); (vi) *linear spectral unmixing* (LSU), entre outros. No entanto, a maioria destes algoritmos não são adequados para trabalhar com conjuntos de dados muito grandes. Recentemente, novos desenvolvimentos foram feitos através do uso de inteligência artificial, isto é, métodos como algoritmos de *machine* e *deep learning*.

### Considerações finais

É evidente que nós, como cientistas, temos de sair da nossa “zona de conforto” na procura de formas e métodos inovadores para lidar com os desafios atuais que enfrentamos, explorando e gerindo conjuntos de dados cada vez maiores como resultado da evolução tecnológica e da digitalização das sociedades. Dessa forma, poderemos arranjar soluções que nos permitam combater as alterações climáticas, como por exemplo acelerar a descarbonização da economia, por exemplo através da aceleração da transição energética, que forçosamente se vai fazer à custa de encontrar mais metais, principalmente a menor distância de onde eles vão ser fabricados e utilizados, de forma a reduzirmos a sua pegada carbónica.

---

# Asterossismologia

## Como explorar os dados que nos mostram o interior das estrelas

Margarida S. Cunha\*, Mário J. P. F. G. Monteiro\*,<sup>1</sup>

\* IACE/ CAUP

<sup>1</sup> DFA/ FCUP

As estrelas são palco de inúmeros processos físicos que decorrem em condições muitas vezes impossíveis de reproduzir em laboratório. Neste contexto, o seu estudo permite-nos, não só procurar dar resposta à curiosidade intrínseca do ser humano acerca do universo que o rodeia, mas também estender o conhecimento da física, explorando uma região do espaço de parâmetros que não nos está acessível de outra forma.

É importante notar que, enquanto sistema físico, uma estrela comum, como o sol, goza de alguma simplicidade. Para isso contribui o facto das estrelas serem corpos aproximadamente esféricos, suportados essencialmente pelo equilíbrio entre a força da gravidade e a força que resulta do gradiente de pressão, compostos de gás maioritariamente ionizado. Não obstante, a complexidade aumenta quando se procura modelar as camadas mais superficiais da estrela, onde a escala de tempo para fenómenos térmicos diminui significativamente e onde a interação entre o campo magnético e o gás é particularmente relevante. Igualmente exigente é a modelação das regiões do interior onde o transporte de energia dominante muda entre o radiativo (transporte de energia assegurado por fotões) e a convecção (transporte de energia feito por movimentação do gás), ou onde a competição entre processos microscópicos e macroscópicos de transporte dos elementos químicos é considerável. De igual forma, o problema em mãos torna-se particularmente complexo quando procuramos compreender os fenómenos que estão na base da origem dos campos magnéticos observados direta ou indiretamente à superfície de uma estrela, a evolução do seu momento angular ou as instabilidades que precedem a fase final da sua vida.

O avanço na compreensão da estrutura e evolução estelar, bem como da interação entre as estrelas e os seus planetas, depende de forma crítica da correta modelação destes fenómenos mais complexos. Por sua vez, o acesso a uma grande quantidade de dados ultra precisos e diversos, hoje disponíveis graças ao investimento em instrumentos de observação a

partir da terra e do espaço, serve de alavanca para esse avanço. O objetivo de alcançar uma caracterização cada vez mais completa dos fenômenos estudados, exige que se recorra a várias técnicas de observação e a diferentes tipos de dados. É a combinação de todos esses dados e a quantidade e qualidade da informação que eles nos revelam, que tem sustentado os grandes desenvolvimentos da física estelar nas últimas décadas.

### Observando o interior das estrelas

O acesso a dados coletados por instrumentos a bordo de missões espaciais, pela sua qualidade e quantidade, permitiu uma verdadeira mudança de paradigma na abordagem do estudo da estrutura e evolução estelar. Essa mudança decorreu da detecção de pequeníssimas variações temporais no brilho de milhares de estrelas (FIGURA 1), associadas à propagação de ondas no seu interior. Este fenômeno, conhecido e estudado no contexto do sol desde a década de 60 do século XX, abriu as portas à inferência de informação direta acerca do interior das estrelas. O princípio subjacente à técnica utilizada para inferir a informação — conhecida por heliosismologia, no contexto do estudo do sol, e por asterossismologia, quando se considera o estudo de outras estrelas — é relativamente simples. De facto, as características das ondas detetadas (em particular, a sua frequência e amplitude) dependem do meio onde estas se propagam e da fonte de energia que as mantém.

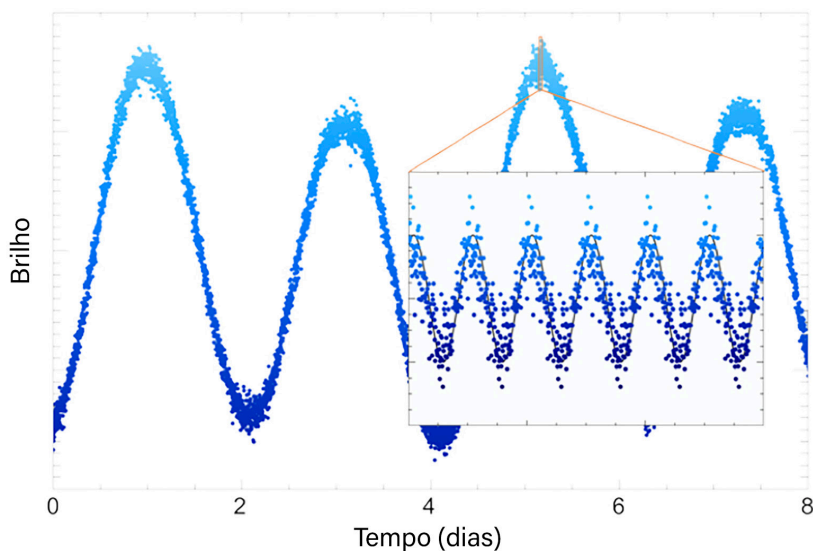


FIGURA 1. Variação do brilho da estrela TIC 237336864 observada pelo satélite TESS, da NASA. A variação de maior amplitude (de cerca de 15 mmag) resulta da presença de manchas (regiões de composição química diferente) à superfície da estrela. Estas aparecem e desaparecem do disco observado devido à rotação da estrela, que neste caso tem um período de cerca de 4,2 dias. As variações de brilho de menor amplitude (de cerca de 0,15 mmag) mostradas, de forma ampliada, no painel sobreposto, são manifestações de ondas acústicas que se propagam pelo interior da estrela com períodos de cerca de 7,4 minutos.



Na generalidade das estrelas, não tendo o benefício da proximidade que existe no caso do Sol, podemos detetar apenas fenómenos ondulatórios globais, resultantes da interferência construtiva dentro de uma cavidade correspondente a uma dada região da estrela.

Como cada modo de oscilação (correspondente a uma frequência própria) contém informação média acerca da cavidade onde se propaga, é possível combinar a informação contida em vários modos de oscilação para extrair informação localizada do interior de uma estrela.

A condição para a determinação dos valores próprios e a relação de dispersão evidenciam o facto de que a frequência de um modo acústico contém informação média acerca da estrutura da estrela (no caso presente, expressa na velocidade do som) entre os extremos da cavidade de propagação. Na prática, uma estrela normalmente exhibe também modos acústicos não radiais, cujas cavidades de propagação não se estendem até tão perto do centro. Este tipo de modos têm sido extensamente explorados no estudo do Sol. Dependendo da sua massa e fase de evolução, a estrela pode também exhibir modos de natureza gravítica, onde as oscilações são mantidas pela ação da gravidade. Podem ainda existir modos de natureza mista, caracterizados por duas cavidades de propagação, uma em que a onda tem uma natureza acústica e outra em que tem uma natureza gravítica, e entre as quais há trocas de energia — à semelhança de um sistema de molas acopladas. É a exploração desta diversidade de modos de oscilação, cada um contendo informação complementar acerca da estratificação da estrela, que está na base da técnica de asterossismologia. O desenvolvimento das técnicas matemáticas usadas inicialmente no estudo do Sol, e, mais tarde, no estudo de outras estrelas, teve como ponto de partida o conhecimento que tinha sido construído no contexto da sismologia da Terra. É de notar, ainda, que a rotação e os campos magnéticos alteram as frequências dos modos próprios, deixando assinaturas particulares que podem ser identificadas e estudadas, inferindo-se, assim, informação acerca da dinâmica e propriedades magnéticas do interior das estrelas.

### **A revolução dos dados**

No início do novo milénio, as chamadas *oscilações de tipo solar* — perturbações intrinsecamente estáveis produzidas pela turbulência associada à convecção — tinham sido observadas em apenas uma mão cheia de estrelas. O lançamento de satélites com capacidade para recolher séries temporais de fotometria (variações no brilho da estrela como função do tempo) de alta cadência e precisão, como o CoRoT (França/ESA), o Kepler (NASA) e o TESS (NASA) veio mudar esta realidade. Hoje, oscilações deste tipo já foram detetadas em mais de 150 mil estrelas. Tendo em conta que o satélite TESS continua em observação, varrendo sucessivamente diferentes regiões do céu e debitando cerca de 27 GB de dados por dia, é de esperar que este número continue a aumentar e que aumente ainda mais, com o lançamento do satélite PLATO (ESA), previsto para 2026.

Muito embora a recolha de séries temporais fotométricas seja essencial para a deteção das oscilações, o estudo do interior das estrelas requer igualmente o conhecimento das suas propriedades globais, tais como a luminosidade e a temperatura efetiva, bem como das suas propriedades químicas, em particular a metalicidade. Neste contexto, o lançamento do satélite Gaia (ESA) no fim de 2013, com a missão de mapear em 6D a nossa galáxia, determinando a posição e o movimento de cerca de 200 mil milhões de estrelas (cerca de 1% das estrelas da Via Láctea), foi determinante para o avanço do conhecimento nesta área de estudo. Um exemplo do resultado da exploração combinada de dados dos satélites GAIA e TESS é ilustrado na FIGURA 2. A partir destes dados, foi possível determinar a massa de mais de 100 mil estrelas gigantes vermelhas e verificar que as mais jovens (de maior massa) estão confinadas essencialmente ao plano do disco galáctico, enquanto que as mais velhas (de menor massa) ocupam regiões mais distantes do mesmo. Estes resultados são importantes porque permitiram testar empiricamente uma previsão fundamental da astronomia galáctica, segundo a qual se espera que estrelas mais jovens estejam mais próximas do plano da Via Láctea.

Os dados obtidos a partir do espaço, são complementados com dados de espectroscopia de alta resolução, obtidos com instrumentos no solo. Os espectros das estrelas (a distribuição da energia emitida em função do comprimento de onda) recolhidos a partir da Terra, alimentam os estudos que estabelecem a temperatura à superfície da estrela e a sua composição química. Nesta componente, têm sido fundamentais os desenvolvimentos tecnológicos que suportam a obtenção destes dados, como os espectrógrafos que se encontram no VLT do Observatório Europeu do Sul (ESO) e noutras infraestruturas, em diferentes partes do globo, que têm permitido realizar *surveys* com vista a coletar espectros de todas as estrelas observadas com o GAIA ou outras missões espaciais. Os dados de espectroscopia, tal como os dados do GAIA, são essenciais para que possamos extrair toda a informação que a asterossismologia nos fornece.

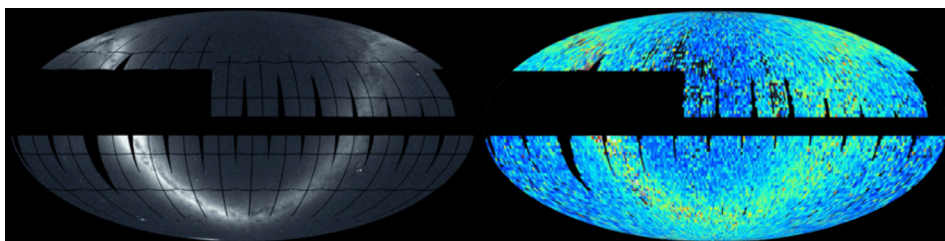


FIGURA 2. Mapas da galáxia em coordenadas eclípticas. A imagem da esquerda mostra o céu observado pelo TESS, onde a banda mais clara ilustra a região de maior densidade de estrelas, correspondente ao plano da galáxia. O mapa à direita é construído a partir de 158 mil estrelas gigantes vermelhas para as quais foi possível detetar oscilações, coloridas de acordo com a massa determinada por combinação de dados dos satélites Gaia e TESS. Os tons azulados mostram estrelas de massa mais pequena, entre 0,8 e 1,2 massas solares, enquanto que o verde, amarelo, laranja e vermelho, mostram estrelas de massa superior, até pouco mais de duas massas solares. A parte omissa nos mapas corresponde à região do céu que não foi observada pelo TESS durante os dois primeiros anos da missão.

De notar que no estudo ilustrado na FIGURA 2, em particular para a deteção da presença de oscilações nas estrelas do TESS, os autores utilizaram métodos de *machine learning*, que são cada vez mais importantes no contexto da análise de grandes volumes de dados. Estas técnicas são também cada vez mais utilizadas na comparação entre as previsões dos modelos de evolução estelar e as observações, que muitas vezes envolvem a criação de grelhas de modelos muito densas. Estes modelos são gerados a partir da simulação de diversos aspetos da física das estrelas, como, por exemplo, a equação de estado, a opacidade, as reações de fusão nuclear e a eficiência da segregação química. Desta forma, os modelos podem ser estendidos e aperfeiçoados, e, em última análise, validados (ou invalidados) com base na combinação dos dados que temos disponíveis para as estrelas.



FIGURA 3. A sonda espacial Gaia, operada pela Agência Espacial Europeia (ESA), lançada em dezembro de 2013, observa os céus a partir da sua órbita em torno da Terra com o objetivo de criar o maior e mais preciso mapa da nossa Galáxia. Esta missão está a produzir catálogos de dados que incluem medições de alta precisão — posições, distância e movimentos próprios — de cerca de 200 mil milhões de estrelas da nossa Via Láctea. Este catálogo fornece informação essencial para a realização de estudos revolucionários em muitas áreas da astronomia, incluindo a análise das populações de estrelas da Galáxia e a sua evolução. Esta imagem artística mostra a sonda Gaia com a Via Láctea ao fundo.

### Que futuro para a astrofísica estelar?

A física estelar é um exemplo paradigmático de como combinar dados muito distintos nos permite construir uma visão cada vez mais completa e precisa do nosso objeto de estudo — no caso, as estrelas. A asterossismologia, pela quantidade de informação que nos permite inferir, tem potenciado avanços significativos na compreensão de como a física opera em regimes e condições impossíveis de reproduzir em laboratório. Tal está na base da revolução que atravessamos, sustentada nos dados que temos e cuja exploração só agora começou. Sabemos hoje muito mais sobre como a energia é transportada no interior das estrelas, sobre como a sua composição química muda ao longo do tempo e sobre como a rotação e a convecção interagem para tornar as estrelas magneticamente ativas. Somos capazes de detetar assinaturas diretas de campos magnéticos escondidos

no núcleo das estrelas e de distinguir, de entre estrelas com características idênticas à superfície, aquelas em cujo núcleo decorre fusão nuclear. Já não é apenas o Sol o nosso laboratório de alta precisão: muitas das estrelas ao nosso redor estendem o nosso laboratório a condições e fenómenos que antes nos eram inacessíveis.

Esta revolução assenta, não só na coleção de dados observacionais (adquiridos por verdadeiras máquinas de coletar informação), mas também nos métodos de análise que estão a ser desenvolvidos. Esses métodos são o que permite transformar os dados em conhecimento, capaz de se concretizar numa nova geração de modelos e avanços científicos na compreensão da nossa estrela e das diferentes populações de estrelas que estão presentes na nossa galáxia. Este conhecimento será crítico para perceber a evolução de outras galáxias e a própria evolução do Universo.

Mas a recolha de dados não parou. Estão já previstas outras missões espaciais e detetores no solo que irão ampliar a quantidade e qualidade de dados, o que nos permite antecipar avanços que irão muito além dos alcançados até à data. Tal esforço tecnológico vai continuar a intensificar a necessidade de se otimizar a extração de informação dos dados disponíveis. O cientista tem de ser capaz de ultrapassar a quantidade e complexidade da informação que existe nos arquivos, para extrair o que necessita para responder a cada pergunta que formula sobre os objetos e fenómenos que estuda. Nunca, como hoje, este ponto foi tão crítico no avanço científico na astronomia em geral, e na astrofísica estelar em particular.

---

# Determinismo biológico, genético e epigenético

## As várias faces de uma fénix

Luca Ribeiro Mendes Nicola, Edson Pereira Silva

LGME/UFF

O determinismo biológico pode ser definido como a noção de que características biológicas determinam, nos seres humanos, o seu sucesso ou fracasso em determinadas atividades. Quando se trata de organismos não humanos, não há dúvida desse facto, visto que uma das bases da teoria evolutiva moderna é a mortalidade que ocorre dentro das populações naturais por conta das diferenças existentes entre os indivíduos. De modo que os indivíduos que sobrevivem são aqueles que passam adiante os seus caracteres.

Este fenómeno óbvio entre os outros seres vivos torna-se mais complexo nos seres humanos, uma vez que estes são caracterizados por apresentarem um desenvolvimento social que se sobrepõe aos seus limites biológicos. Por exemplo, a biologia do humano não lhe permite voar, mas por meio do desenvolvimento intelectual e técnico dos últimos séculos, a humanidade foi capaz de construir um veículo — o avião — que permite desafiar, mesmo que à custa de muita energia, a lei da gravidade. Do mesmo modo, para qualquer organismo individual que depende da sua visão, a miopia (um alongamento excessivo do globo ocular que promove a visão turva de longa distância) representaria uma séria dificuldade à sobrevivência, contudo, com o advento da oftalmologia e da ótica moderna, essa condição pode ter os seus efeitos reduzidos e até completamente remediados nos seres humanos pela utilização de óculos e lentes de contacto, possibilitando que adultos com miopia, cuja incidência varia entre 10% e 30% em diversos países, vivam normalmente. Dito isso, tem sido uma tentativa recor-

rente, fundamentalmente ao longo do último século, mostrar que nos seres humanos a sua biologia tem um papel determinante no sucesso ou no fracasso dos indivíduos e, especialmente, de determinados grupos, géneros, etnias, etc.. O conjunto dessas tentativas forjadas em explicações e teorias é o que compõe o chamado determinismo biológico.

### **Determinismo biológico**

Alguns cientistas do século XIX estiveram envolvidos com o programa de pesquisa anatómica que pretendia testar a hipótese de que a anatomia do cérebro humano estava correlacionada com supostas diferenças intelectuais e morais dos grupos humanos que eram identificados como “raças”. Uma das iniciativas deste programa de pesquisa era a craniometria, que media o tamanho de crânios de representantes de populações americanas, africanas, europeias e asiáticas na esperança que fossem encontradas correlações entre o tamanho dos crânios e a inteligência. Um dos objetivos subjacentes à craniometria era demonstrar que existia uma hierarquia racial biologicamente determinada que justificaria a dominação social imposta pela colonização, bem como as suas práticas, entre elas a escravidão.

### **Determinismo genético**

A redescoberta dos trabalhos de Gregor Mendel (1822—1884) no início do século XX por Carl Correns (1864—1933), Erich Tschermak (1871—1962) e Hugo de Vries (1848—1935) de maneira independente, permitiu o estudo dos padrões de herança observados para muitas características em vegetais e animais e a manipulação em larga escala de cruzamentos com o objetivo de aumentar a produtividade agropecuária. Além disso, muitos investigadores passaram a estudar a maneira pela qual os fatores hereditários propostos por Mendel (posteriormente chamados de genes) determinavam, também, o conjunto das características humanas. Nesse contexto, os adeptos do determinismo biológico criaram a expectativa de se reduzir as características sociais dos seres humanos ao determinado pela sua base genética. Essa crença exacerbada no poder dos genes designada determinismo genético e os trabalhos interessados nessa nova forma de determinismo pretendiam investigar a base genética de características complexas como agressividade, inteligência, homossexualidade, etc..

### **Projeto Genoma Humano**

A expectativa mais modesta do *Projeto Genoma Humano* (PGH) era que a descodificação da sequência de nucleotídeos que compõe o genoma humano trouxesse respostas sobre como tratar doenças complexas tais como o cancro. A expectativa mais ambiciosa era que fosse revelado o *Santo Graal* da “natureza humana”. Do ponto de vista técnico, o PGH

levou a grandes avanços, como o desenvolvimento de sequenciadores que permitiram reduzir drasticamente o tempo necessário para sequenciar cadeias de DNA. Em certa medida, também, o PGH influenciou decisivamente no estabelecimento do campo da genômica. Contudo, em relação à determinação genética das características, os resultados do PGH ficaram muito aquém do esperado. A ambição de encontrar relações diretas de causa e efeito entre sequências de DNA e características como inteligência, agressividade, preferência sexual, preferência política, etc., foi substituída pela conclusão de que nem a relação entre DNA e genes era bem conhecida. Um exemplo disso é que, no início do PGH, esperava-se encontrar algo em torno de 100 000 genes na sequência de DNA e, alguns anos após a sua conclusão, não só foram encontrados menos genes do que o esperado (em torno de 20 000 a 30 000), como, dependendo da definição de gene utilizada nas diversas bases de dados existentes, se encontraram diferentes quantidades totais de genes (TABELA 1). Nesse sentido, a grande conquista teórica do PGH foi — apesar das fantasias dos seus idealizadores — rechaçar a perspectiva grosseira do determinismo genético. Além da imensa frustração com os resultados do PGH, as pretensões do determinismo genético pareciam estar sendo sepultadas com o desenvolvimento das ideias da epigenética... Só que não.

TABELA 1. Diferentes estimativas de componentes do genoma de acordo com a base de dados utilizada. Nas colunas encontram-se as diferentes bases de dados e na linha a quantidade de genes que codificam proteínas.

	<i>Gencode</i>	<i>Ensembi</i>	<i>RefSeq</i>	<i>CHESS</i>
<b>Genes codificadores de proteínas</b>	19 901	20 376	20 345	21 306

### Determinismo epigenético

Conrad Waddington (1905—1975), no final da década de 1940, cunhou o termo “epigenética” para explicitar que o desenvolvimento de um organismo é o resultado da interação entre o seu conteúdo genético, as influências ambientais e o estado imediatamente anterior daquele mesmo organismo.

O determinismo epigenético é a noção de que os mecanismos epigenéticos contêm todas as possibilidades de variação fenotípica engendradas por variáveis ambientais. Assim, a noção de contingência dos fenômenos ambientais é subsumida por uma perspectiva de programa epigenético na qual toda variação fenotípica já está pré-determinada. O facto de que essa influência ambiental é incorporada no DNA, na forma de moléculas expressas pelo mesmo DNA, é usado pelo determinismo epigenético para afirmar que o fator ambiental age como mero gatilho para uma via de desenvolvimento contida no código epigenético. Ao afirmar isso, contudo, é negligenciada a noção de que cada etapa do desenvolvimento de um organismo depende da interação de diversos fatores com o próprio estado anterior

daquele organismo. Ou seja, não é possível tratar o desenvolvimento como um fenômeno linear cujo resultado já está pré-determinado no DNA. Um exemplo disso é a variação de cerdas entre os lados esquerdo e direito da mosca-do-vinagre (*Drosophila melanogaster*), que não é determinada por nenhuma variação genética ou ambiental, mas depende inteiramente de fenômenos moleculares aleatórios associados à formação daquelas células, sendo comumente referido como ruído de desenvolvimento (FIGURA 1). Ao submeter as eventualidades do desenvolvimento à noção de programa, o determinismo epigenético atualiza muitas das teses do determinismo genético, que por sua vez são tributárias do determinismo biológico.

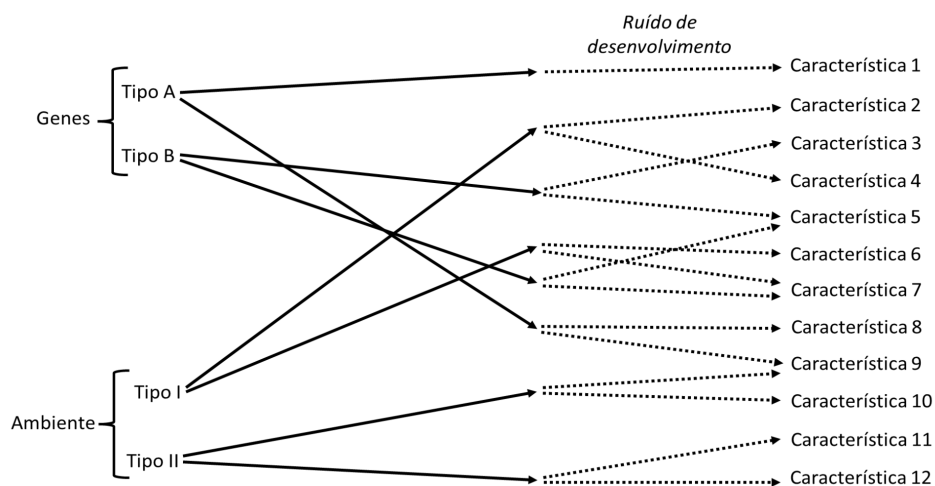


FIGURA 1. Interação entre genes, ambiente e diversos fenômenos caracterizados como ruído do desenvolvimento na constituição contingente dos organismos.

### Determinismo(s) na biologia

Da relação entre tamanho do crânio e inteligência até ao programa epigenético, o determinismo mantém o seu núcleo duro: a noção de que a biologia humana é responsável por determinar o sucesso ou fracasso de grupos humanos. Tentou-se demonstrar aqui que os diversos determinismos — biológico, genético e epigenético — têm como ponto comum a tentativa de deslocar a atenção da esfera social para a biológica, sonogando, assim, a responsabilidade política de agir para transformar realidades de desigualdade, exploração e opressão. Tentou-se demonstrar, também, que as ideias deterministas não estão ultrapassadas, pelo contrário, elas atualizam-se incessantemente de modo a adequarem-se aos avanços científicos. Neste sentido, este trabalho defendeu o ponto de vista de que a biologia humana, em articulação dialética com o desenvolvimento social, não nos aprisiona, mas oferece as ferramentas para a nossa libertação de algumas condicionantes da própria biologia.



---

# Avanços da genómica

## e os desafios na descodificação da vida

Agostinho Antunes

CIIMAR/FCUP

Os princípios básicos da genética foram lançados pela primeira vez por Gregor Mendel, em 1866, um monge da Ordem de Santo Agostinho que realizou experiências de cruzamentos controlados em ervilheiras. Mendel esclareceu a forma como as características (por exemplo: cor, forma) são transmitidas de geração em geração e introduziu os termos “recessivo” e “dominante” para explicar a relação desses traços na descendência. Descreveu a ação de fatores “invisíveis” no fornecimento de traços visíveis de maneiras previsíveis (esses traços “invisíveis”, correspondem na realidade aos genes).

### Como surgiu a revolução no conhecimento genómico?

A descodificação do DNA só viria a ser delineada com o desenvolvimento de técnicas rápidas de sequenciamento de DNA, desenvolvidas em 1977 por Frederick Sanger (Prémio Nobel em 1980). Esse avanço tecnológico iria possibilitar um desafio pioneiro — o Projeto Genoma Humano — oficialmente iniciado em 1990 com o Departamento de Energia (DOE) e os Institutos Nacionais de Saúde (NIH), nos EUA, e com previsão de 15 anos. Os objetivos incluíam: mapear o genoma humano e determinar as suas 3,2 giga bases (nucleótidos), bem como mapear e sequenciar os genomas de outros organismos úteis para o estudo da biologia, desenvolver tecnologia para a análise de DNA e estudar as implicações sociais, éticas e legais da investigação do genoma.

Em 2003, o Projeto Genoma Humano foi finalmente “concluído”, um projeto notável de esforços conjuntos do *International Human Genome Project* (HGP) Consortium e o *Celebra Genomics*, numa cerimónia apresentada na Casa Branca, com repercussões notáveis e equiparáveis ao feito de colocar o Homem na Lua. Em menos de 150 anos, desde os pioneiros como o Mendel, que demonstrou que características poderiam ser herdadas, é impressionante o avanço científico alcançado.

Esta forma final do genoma humano contém 2,85 giga bases (nucleótidos) e uma taxa de erro de apenas 1 evento em cada 100 000 bases sequenciadas (em 2001, o Projeto Genoma Humano publicou um “rascunho” cobrindo 90% do genoma humano, e subsequen-

temente foram preenchidas as lacunas em regiões ambíguas do DNA até se completar 99% do genoma humano). Surpreendentemente, foi identificado um número relativamente pequeno de genes que codificam proteínas (entre 20 000 e 25 000), havendo genes semelhantes com as mesmas funções presentes em diferentes espécies.

A sequenciação do genoma humano levantou também uma importante questão biológica: decifrar corretamente o genoma de um organismo requer a sua comparação com os genomas de outros organismos, partilhando ancestrais evolutivos comuns a diferentes escalas temporais. A genômica comparativa surge assim como um ramo de estudo prolífico, gerando-se imensos avanços científicos: percentagem de genes partilhados no Homem relativamente a diferentes espécies, por exemplo: chimpanzé (98%), ratinho (92%), mosca-da-fruta (44%), levedura (26%), planta (18%). Várias propostas são elaboradas para a sequenciação do genoma de diversos organismos, com base na relevância biológica, económica e saúde humana. São organizados vários consórcios internacionais para fomentar esse avanço científico, por exemplo: *Genome10K*, *i5K*, *Global Invertebrate Genomics Alliance*, *Fish10K*, *Bat10K*, *Vertebrate Genome Project*, *Earth BioGenome project*, *European Reference Genome Atlas*, etc.).

Os cientistas de todo o mundo continuaram a desenvolver a sua compreensão do DNA, incluindo o recurso a manipulações genéticas, criando com sucesso um organismo com um código genético artificial expandido, abrindo portas às edições de genomas. Isto permitiu a experimentação de mutações do DNA (relação genótipo-fenótipo), o tratamento de algumas doenças genéticas, a criação de organismos capazes de produzir remédios, ou produtos industriais de base orgânica.

Simultaneamente, vários avanços na área médica são alcançados, decifrando-se a base genética de várias doenças, com centenas de genes envolvidos, incluindo distúrbios complexos como a esquizofrenia, e impulsionando a produção de novos medicamentos.

### **Qual o estado da arte da compreensão do nosso genoma?**

Desde o seu lançamento, o genoma humano de referência cobriu apenas a fração eucromática do genoma, deixando importantes regiões heterocromáticas inacabadas. Produzir uma sequência completa e sem lacunas do genoma humano foi uma meta para os biólogos, um sonho que se realizou em 2022.

A sequenciação de leitura longa (*long reads*) apoiou a montagem completa de telómero a telómero (T2T) da linhagem de células humanas pseudo-haplóides CHM13. Abordando os 8% restantes do genoma, o Consórcio Telomere-to-Telomere (T2T) apresenta a sequência completa de 3,055 giga bases de um genoma humano, T2T-CHM13, que inclui montagens sem lacunas para todos os cromossomas, exceto o Y, corrigindo erros anteriores e apresentando quase 200 milhões de pares de bases de sequências contendo 1 956 previsões de genes, 99 dos quais são previstos como codificadores de proteínas. As regiões

completas incluem todas as matrizes de satélites centroméricos, duplicações segmentares recentes e os braços curtos de todos os cinco cromossomas acrocêntricos, revelando essas regiões complexas do genoma para estudos funcionais e variação genética.

As vastas melhorias na descoberta de variantes em amostras de diversos ancestrais posicionam o genoma de referência preciso e completo para revelar a diversidade entre as populações humanas. Adicionalmente, revelam a importância de aumentar os esforços para alcançar também montagens de nível T2T para primatas não humanos e outras espécies. Essa informação ajudaria a entender completamente a complexidade e o impacto das inovações genômicas observadas, anteriormente desconhecidas para algumas das regiões com evolução mais dinâmica no nosso genoma.

São alavancadas inovações em tecnologia, *design* de estudo e parcerias globais com o objetivo de construir a referência de pangenoma humano da mais alta qualidade possível (*Human Pangenome Reference Consortium*). Este avanço tecnológico irá permitir, futuramente, a montagem rotineira de genomas diploides completos em outras espécies. Com atenção às estruturas éticas, a referência do pangenoma humano conterà uma representação mais precisa e diversificada da variação genômica global, melhorará os estudos de associação gene-doença entre as populações, expandirá a investigação genômica para as regiões mais repetitivas e polimórficas do genoma, fomentando um recurso genético definitivo para o futuro da investigação biomédica e medicina de precisão.

### **O que o futuro reserva para a nossa compreensão da genômica?**

A publicação dos primeiros rascunhos do genoma humano lançou uma nova era na descoberta biológica, reduzindo o número de genes esperados, mas expandindo enormemente a compreensão da regulação genética. Atualmente, são produzidos genomas individuais e análises genômicas às dezenas de milhares, mas persistem alguns dos mesmos conflitos centrais em relação à disponibilidade de dados, equidade e privacidade.

Volvidos cerca de 20 anos da publicação do primeiro rascunho da sequência de todo o genoma humano, os investigadores mergulharam ansiosamente na era "pós-genômica", remodelando a investigação biológica e médica, assim como os desafios de recolha, curadoria e acesso de dados gerados pelos atuais projetos. Avanços das metodologias de sequenciação (notavelmente o uso de *long reads*) permitiram explorar, mais recentemente, genomas de grandes dimensões relativamente ao genoma humano (3 giga bases), como o genoma do grande tubarão branco (4,63 giga bases), o genoma do axolote (30 giga bases) ou o genoma do peixe pulmonado Africano (40 giga bases), e ainda a sequenciação de genomas de várias espécies extintas, como o periquito-da-Carolina ou o gato-de-cimitarra.

O progresso na sequenciação de genomas permite agora a geração em larga escala de genomas de referência. Ainda que a sequenciação de um genoma possa ser efetuada atual-

mente em poucos minutos, a sua montagem, anotação e decifração pode necessitar de inúmeros recursos humanos especializados e requerer um considerável tempo (FIGURA 1). O mundo pós-genoma esperado para os próximos anos requer o desenvolvimento e implementação de mais ferramentas bioinformáticas e o recurso a metodologias de inteligência artificial.



FIGURA 1. A montagem, anotação e decifração de um genoma pode requerer inúmeros recursos humanos especializados e um considerável tempo.

Consideráveis progressos têm sido ainda obtidos no campo inovador da epigenética (o estudo das mudanças nos organismos causadas pela alteração da expressão genética). Um elevado número de mecanismos moleculares pode afetar a atividade dos nossos genes. Surpreendentemente, as nossas experiências e escolhas de vida podem também alterar a atividade desses mecanismos, alterando a expressão génica. Ainda mais fascinante, é que essas modificações na expressão génica podem ser herdadas, o que significa que as experiências de vida dos nossos ancestrais podem influenciar fundamentalmente a nossa constituição biológica. Estas descobertas terão um impacto dramático no futuro do sistema de saúde, realçando a importância das escolhas de estilo de vida, o que trará impacto nas gerações futuras.

Outro grande desenvolvimento futuro espetável associado à investigação genómica é o desenvolvimento da medicina personalizada com o uso de medicamentos precisos, para eliminar os efeitos de muitas doenças genéticas causadas por genes mutantes, que diferem de uma pessoa para outra. Ao identificar essas combinações, através da sequenciação do genoma do indivíduo, os medicamentos podem ser adaptados de um modo preciso, proporcionando o melhor tratamento possível.

# Serendipidade na investigação científica

Alexandre Lopes Magalhães

DQB/ LAQV/ Universidade do Porto

Serendipidade é um termo frequentemente usado em contexto de investigação e descobertas científicas. Trata-se de um anglicismo recente no léxico português com o seguinte significado, segundo o dicionário da Porto Editora: 1 — Aptidão de atrair a si acontecimentos favoráveis de maneira fortuita; dom de fazer boas descobertas por acaso; 2 — Acontecimento favorável que se produz de maneira fortuita; acaso feliz; descoberta accidental. O termo *serendipity* foi introduzido pela primeira vez, em 1754, pelo escritor inglês Horace Walpole que se inspirou no conto popular persa *Os Três Príncipes de Serendip* (antigo nome do Sri Lanka) cujos heróis, em busca de um camelo perdido, vão fazendo descobertas, aliando o seu espírito sagaz à observação cuidada de acontecimentos fortuitos.

Mas como pode o acaso estar envolvido em investigação científica? O método científico consiste num conjunto de rigorosos procedimentos assentes em regras e princípios do pensamento crítico e raciocínio dedutivo. Este instrumento cognitivo é fundamental para o cientista poder garantir o avanço do conhecimento e validar grandes descobertas. Um dos pilares da ciência moderna é a observação crítica de fenómenos naturais ou de resultados de experiências planeadas. Contudo, há resultados que não encaixam no esquema preditivo do cientista e, apesar de estarem corretos, desafiam o edifício lógico que ele foi construindo ao longo da sua vivência. Por vezes a frustração assalta-o e pode até ser tentado a minimizar a importância desses resultados dissonantes, rejeitando-os definitivamente como erros. Quantas descobertas terão sido inibidas por não se prestar a devida atenção a raras observações fortuitas?

O acaso de uma grande descoberta não acontece com a afortunada simplicidade com que se encontra uma pepita de ouro a reluzir no solo durante uma despreocupada cami-

nhada pelo campo. É necessário possuir uma mente curiosa, observadora e crítica para identificar um acontecimento fortuito que tem potencial para vir a transformar-se numa descoberta. O cientista ou inventor com adequado treino científico é capaz de estabelecer relações entre o fenómeno observado e o conhecimento previamente adquirido. De facto, na original definição de *serendipity*, Walpole referia a sagacidade como característica fundamental. Assim, a partir de um acontecimento aparentemente desconexo, que para muitos não passaria de um contratempo ou lamentável insucesso, um espírito sagaz será capaz de identificar uma ideia com potencial e desencadear um plano de investigação, seguindo as regras do rigoroso método científico, que o leve à descoberta científica. Como disse Louis Pasteur "*O acaso só favorece a mente preparada*".

Muitos casos de descoberta por serendipidade surgem na história da Ciência, como por exemplo os raios-X, a radioatividade, o vidro inquebrável, o forno de micro-ondas, os polímeros condutores, os *post-its*, o *nylon*, o medicamento *sildenafil* (Viagra), só para mencionar alguns. A seguir, são descritos sucintamente três casos bastante populares de descobertas que se enquadram no fenómeno de serendipidade.

Um dos mais conhecidos é o da descoberta, em 1928, do primeiro antibiótico, a *penicilina*, pelo bacteriologista escocês Alexander Fleming (1881—1955). Na véspera de partir para duas semanas de férias, Fleming preparou várias placas com uma cultura de bactérias *Staphylococcus aureus*; mas, em vez de as colocar na estufa, deixou-as esquecidas em cima da bancada do seu laboratório no Hospital St. Mary em Londres. Num laboratório do piso imediatamente abaixo, trabalhava um perito em bolores que cultivava, entre outros, os esporos do fungo *Penicillium notatum*. Imagina-se que os esporos, muito leves, se tenham espalhado pela atmosfera do edifício e tenham entrado em grande quantidade no laboratório de Fleming, cuja porta estaria sempre aberta. Quando regressou de férias e resolveu arrumar o laboratório, ele reparou que havia placas que apresentavam largas zonas claras totalmente desprovidas de estafilococos, precisamente as que se encontravam cercadas pelo bolor. Fleming achou este fenómeno curioso e resolveu investigar a razão, realizando várias experiências que o levaram a concluir que os fungos do género *Penicillium* produzem uma substância responsável pelo efeito bactericida e que, compreensivelmente, ele designou de *penicilina*. Mais tarde, dois outros cientistas, Ernst Chain (1906—1979) e Howard Florey (1898—1968) purificaram e caracterizaram a penicilina, e mostraram que esta pode ser usada como um antibiótico não tóxico capaz de matar bactérias que causam infeções em animais e humanos. Estes três cientistas partilharam o prémio Nobel da Medicina em 1945. Na cerimónia de entrega do prestigiado prémio, Fleming declarou humildemente que o seu único mérito tinha sido, simplesmente, dar atenção à inesperada observação.

Outro caso não menos surpreendente é o da invenção do Velcro pelo engenheiro suíço George de Mestral (1907—1990). No verão de 1941, depois de regressar de uma das suas habituais saídas para a caça nos Alpes, George de Mestral reparou que a sua roupa e o

pêlo do seu cão estavam cobertos de sementes de *Arctium*, muito difíceis de arrancar. Este facto, apesar de o irritar, deixou-o também curioso e decidiu investigar o mecanismo da forte fixação das sementes. Usando um microscópio, ele constatou que as sementes eram revestidas por um grande número de minúsculos ganchos que se agarravam a filamentos vegetais ou animais. Imaginou então criar dois materiais que se pudessem ligar reversivelmente e que pudessem ter aplicação, por exemplo, na indústria do vestuário. Porém, o caminho que liga uma ideia genial a um produto de mercado com sucesso é muitas vezes tortuoso, e só é possível percorrê-lo com muita resiliência e engenho. George de Mestral tentou diferentes materiais e metodologias de fabrico durante 10 anos até conseguir submeter uma patente da sua ideia com o nome Velcro, que é um acrónimo das duas palavras francesas que significam veludo (*velours*) e gancho (*crochet*). O negócio foi crescendo, mas a penetração no mercado do vestuário foi mais difícil por razões estéticas; finalmente, a marca Velcro teve o derradeiro impulso quando a NASA, uma década mais tarde, usou os seus artigos em fechos e outras soluções de segurança nas missões espaciais Apollo. Hoje em dia será difícil encontrar uma pessoa que não conheça o popular sistema Velcro constituído pelas duas características peças que aderem reversivelmente uma à outra; quando observadas à lupa, verifica-se que uma delas é constituída por inúmeros ganchos minúsculos que engatam no emaranhado de fibras existente na outra, exatamente o mesmo mecanismo natural que intrigou Mestral num despreocupado passeio em 1941.

Finalmente, abordemos a importante descoberta do tratamento da borracha que lhe permite ter a ampla aplicação prática que conhecemos hoje em dia. A borracha é uma substância natural que já era utilizada pelos povos da Mesoamérica há muitos séculos, como por exemplo a civilização Olmeca, e que é produzida a partir do látex extraído da seringueira ou árvore-da-borracha. Este produto natural, que se apresentava maleável e pegajoso a temperaturas elevadas, e que se tornava duro e quebradiço a temperaturas baixas, pouco entusiasmo despertou nos europeus quando os primeiros exploradores do Novo Mundo o trouxeram a bordo das caravelas no século XVI. Uma das primeiras aplicações práticas desta substância natural foi como agente de remoção de marcas de lápis, isto é, como utensílio de escrita que ainda hoje se usa e é designado comumente por *borracha*. O termo *anglo-saxónico rubber* deriva precisamente de *rub* (fricção) e a palavra portuguesa borracha foi assimilada da equivalente espanhola que significa odre em couro cujo interior era impermeabilizado com essa substância para transporte de vinho. O químico escocês Charles Macintosh (1766—1843) inventou, em 1824, um processo de impermeabilização de vestuário que o tornou famoso, mas a variação das propriedades da borracha com a temperatura tornava as peças de vestuário e de calçado particularmente desconfortáveis.

Charles Goodyear (1800—1860), um químico autodidata americano, pretendeu inventar um método de tratamento da borracha que mantivesse as suas atrativas características invariáveis com a temperatura, de modo a melhorar e alargar o seu leque de aplicações.

Durante aproximadamente dez anos, Goodyear testou inúmeros métodos, entre os quais um que envolvia o tratamento com enxofre. Certa vez, inadvertidamente, deixou cair um pouco da mistura numa frigideira e, para sua surpresa, a borracha não derreteu e manteve-se flexível. Achando este acontecimento espantoso, resolveu colocar o material ao frio no exterior da sua casa. Na manhã seguinte, reparou que o material mantinha as propriedades e, mesmo quando o trouxe para dentro e o aproximou do fogo da sua lareira, o pedaço de borracha permanecia flexível e não pegajoso. Goodyear resolveu então otimizar o processo de estabilização da borracha, tendo registado a respetiva patente em 1844 com a designação de vulcanização, inspirada no deus romano do fogo *Vulcano*. A primeira fábrica Goodyear foi fundada em 1898 no estado americano de Ohio, fornecendo fundamentalmente pneus para bicicletas e carruagens. Com o advento do automóvel, a empresa cresceu muito e, atualmente, 75% da produção mundial de borracha vulcanizada é destinada à indústria de pneus de automóveis e de outros meios de transporte.



O fenómeno da serendipidade pode manifestar-se através de mecanismos de processamento da informação distintos, como foi estudado recentemente por Ohid Yaqub e que se pode inferir pelos três exemplos acima descritos. Segundo este autor, existe ainda um quarto tipo de descoberta por serendipidade que engloba as fortuitas observações que, não tendo aplicação imediata, conduzem a soluções de problemas a *posteriori*. Um interessante exemplo é o da formulação do vidro inquebrável (ou de segurança) criado por Edouard Benedictus (1878–1930) e que teve origem numa inesperada observação feita no seu laboratório, em 1903, ao deixar cair um frasco ao chão que, curiosamente, não se despedaçou em inúmeros estilhaços de vidro. Contudo, somente seis anos mais tarde é que Benedictus resolveu registar a patente do vidro de segurança ao ler a notícia de um grave acidente de automóvel em que duas mulheres ficaram gravemente feridas pela quebra do para-brisas.



A descoberta e conceção de novos fármacos, como o caso da penicilina, é uma área onde abundam exemplos de serendipidade. Um caso, deveras curioso, é o da nitroglicerina sintetizada em 1846 e usada como potente explosivo. Anos mais tarde, verificou-se que um elevado número de trabalhadores que manuseavam este explosivo diariamente apresentavam queixas de cefaleias. Resolveu-se então efetuar um estudo sistemático epidemiológico e concluiu-se que a substância era de facto a causa deste efeito e, posteriormente, foi possível concluir também que as dores resultavam da dilatação dos vasos sanguíneos do cérebro causada pela ação da nitroglicerina. A comunidade médica usou este estudo para propor que a dilatação de vasos sanguíneos do coração poderia ser muito útil no tratamento de doenças cardiovasculares. Desde 1879, a nitroglicerina, também conhecida nos meios médicos por trinitrato de glicerilo, tem vindo a ser usada sob a forma de aerossóis ou drageias de administração sublingual na profilaxia e tratamento de insuficiência cardíaca e hipertensão arterial.

Estudos epidemiológicos são suportados por análises de dados obtidos em amostras de uma dada população e podem revelar correlações inesperadas entre fármacos e os seus efeitos. Tradicionalmente, estes estudos usam amostras devidamente preparadas para um dado objetivo e baseiam-se em técnicas básicas de estatística. A análise de grandes volumes de dados aumenta a probabilidade de se encontrarem correlações inesperadas que podem configurar descobertas serendipitosas. Porém, a aplicação massiva de métodos de estatística clássica pode encontrar, por coincidência numérica, correlações espúrias entre variáveis; por exemplo, Tyger Vigen constatou que o número de doutoramentos em engenharia civil nos Estados Unidos da América entre os anos de 2000 e 2009 está fortemente correlacionado (95,9%) com o consumo *per capita* de queijo *mozzarella*. Ninguém acredita que estas duas variáveis estejam correlacionadas! Ou estarão?

Na última década, a sociedade global tem assistido a um aumento extraordinário e ininterrupto de informação depositada em enormes bases de dados, o que estimulou o desenvolvimento de novos métodos de tratamento de grandes quantidades de dados. Estas inovadoras técnicas de análise estão a criar uma autêntica revolução em quase todos os campos da Ciência porque se têm revelado extremamente eficazes na previsão de fenómenos. De facto, como já foi vaticinado por vários autores, as modernas técnicas de Inteligência Artificial e outros algoritmos de *Big Data* podem representar uma perfeita Máquina de Serendipidade, no sentido em que permitem observações de causalidade inesperada que, obviamente, necessitarão de tratamento científico posterior adequado para se consolidar uma verdadeira descoberta científica.

A serendipidade, por definição, é um fenómeno aleatório. Se fosse possível controlá-lo, deixaria de se classificar como serendipidade. Não podemos, portanto, prever quando estamos perante um potencial caso de serendipidade, mas devemos manter sempre a nossa capacidade de observação aguda e o nosso espírito curioso, inquisitivo e crítico perante fenómenos que nos vão surgindo pela frente, mesmo quando aparentam ser erróneos.

# Flor de BorrAGEM

in casadasciencias.org/banco-imagens

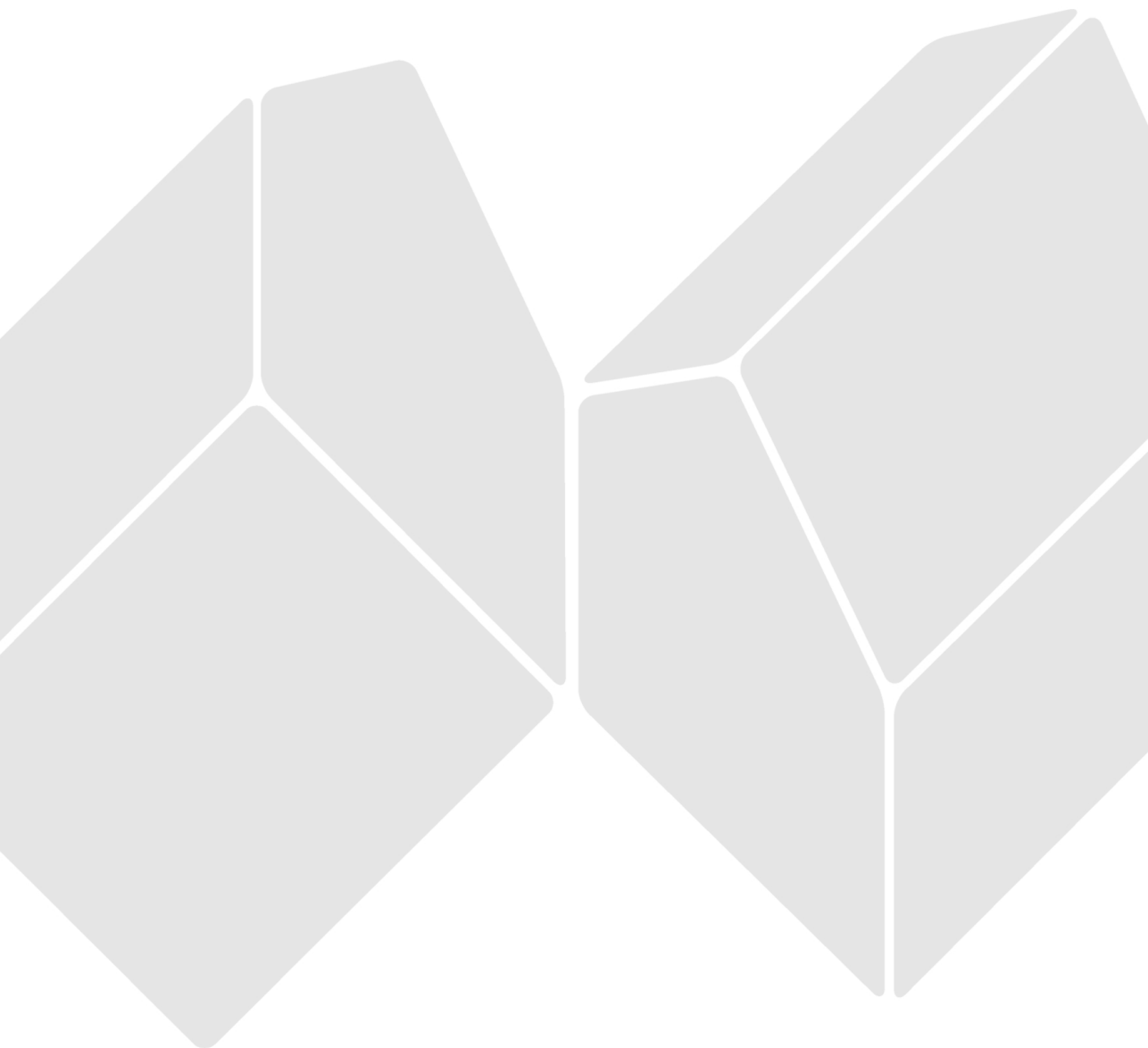
A propósito da belíssima flor da borragem (*Borago officinalis*), este comentário remete para uma descrição anatómica da planta e suas utilizações. A borragem é uma herbácea anual que pode atingir 70 cm de altura. Nativa da região Mediterrânea, encontra-se naturalizada em outras regiões. Os caules e as folhas são recobertos por indumento hirsuto-áspero. As folhas são alternas, simples; as basilares são ovadas, elípticas ou lanceoladas; as caulinares são ovado-oblongas a lanceolado-oblongas, as superiores, decorrentes e rentes. As flores, em inflorescências cimeiras ramificadas, têm pedicelos longos avermelhados, corola em forma de estrela com cinco pétalas azul-violeta ou rosadas, nectários a rodear o tubo floral e estames com anteras purpúreas e compridas. Ainda que prefira solos calcários, a borragem encontra-se em todo o país em prados, pousios, incultos e escombros. É cultivada para alimentar apiários e consociada em culturas variadas para atrair insetos auxiliares no controlo de pragas. Das sementes extrai-se um óleo rico em ácido  $\gamma$ -linoleico utilizado em cosmética.

Ao longo da história muitas são as referências a Borago ou BorrAGEM, planta cheia de propriedades medicinais e fonte de uma grande variedade de nutrientes únicos. Nativa do mediterrâneo, *Burrach* foi o nome que os celtas deram a esta planta, significando "alegre coragem". É conhecida desde a antiguidade e Plínio "o Velho" acreditava que esta planta era a *Nepenthe* da *Odisseia de Homero* que induzia o esquecimento quando infundida em vinho. Dioscórides, médico romano, precursor da Botânica moderna, na sua obra *De materia medica*, sugeria infusões vínicas de borago para "levantar os espíritos deprimidos" e Gerald, na sua obra *Herbal* celebrou as qualidades da borragem afirmando que "*conforta o coração, afasta a tristeza e aumenta a alegria*". Curiosamente, entre outras novidades, a investigação mais recente demonstra que a borragem pode estimular a produção de adrenalina verificando-se ter importante papel no tratamento da depressão. Sendo a fonte vegetal mais rica conhecida de ácido gama-linoleico (GLA), parece que também poderá trazer alguma esperança na luta contra o cancro.

José Pissarra  
Ciências/Universidade do Porto

Rubim Almeida  
Ciências/Universidade do Porto





**CASA DAS CIÊNCIAS**  
EDULOG · FUNDAÇÃO BELMIRO DE AZEVEDO