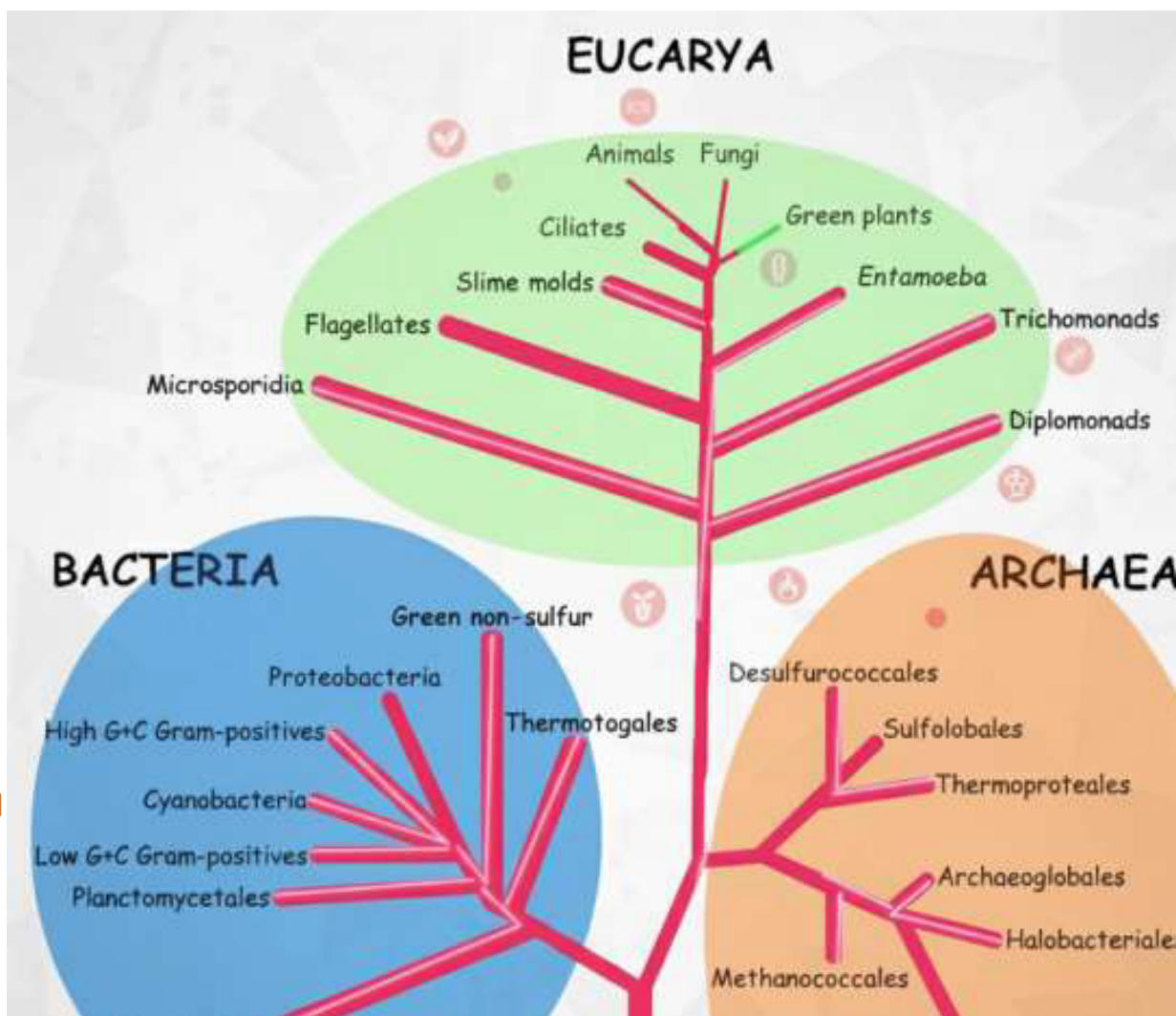


Dep. de Biologia

Faculdade de Ciências

CADERNOS DE
BIOLOGIA
VOL. VII

SISTEMÁTICA FILOGENÉTICA



Rubim Almeida

2025

Conteúdo

Capítulo 1 — O que é a Sistemática Filogenética e por que ela não é só “fazer árvores”	3
Capítulo 2 — Como se escolhem e codificam os caracteres: a arte (e a ciência) por trás da matriz.....	4
Capítulo 3 — Como se constrói a matriz de dados — e por que ela é o coração da análise.....	6
Capítulo 4 — Que programas usar, que métodos aplicar e como interpretar os resultados	7
Capítulo 5 — Como se mede a confiança numa árvore: bootstrap, probabilidades e consenso	8
Capítulo 6 — Como se transforma uma árvore numa classificação — e por que as velhas categorias não servem	10
Capítulo 7 — O que fazer depois: revisão, publicação, comunicação e próximos passos	11
Bibliografia	12
Referências Bibliográficas.....	12
UM CURTO GLOSSÁRIO	13

NOTA DO AUTOR

Este trabalho sobre Sistemática Filogenética tem como principal objetivo introduzir os alunos ao fascinante universo da classificação biológica baseada em relações evolutivas. Não se pretende, de forma alguma, esgotar o tema, que exige estudo aprofundado, leitura crítica e dedicação continuada. A sistemática filogenética é uma área em constante evolução, que integra dados morfológicos, genéticos e moleculares para reconstruir a história da vida na Terra.

Ao apresentar os conceitos fundamentais e a relevância desta abordagem para a compreensão da biodiversidade, espera-se despertar o interesse dos alunos e proporcionar uma base para estudos futuros. A importância da sistemática filogenética ultrapassa os limites da taxonomia tradicional, oferecendo ferramentas poderosas para investigar padrões evolutivos, relações entre grupos biológicos e processos de especiação.

Este é, portanto, um convite à curiosidade científica e ao aprofundamento autônomo, com a consciência de que o verdadeiro domínio deste campo requer muito mais do que uma leitura introdutória.

Capítulo 1 — O que é a Sistemática Filogenética e por que ela não é só “fazer árvores”

A Sistemática Filogenética é a disciplina científica que tem por objectivo reconstruir as relações evolutivas de parentesco entre os organismos, expressando essas relações numa classificação que seja, simultaneamente, um sistema de comunicação e uma hipótese testável sobre a história da vida. Ao contrário do que por vezes se supõe, não se trata de um mero exercício de ordenação ou de aplicação de algoritmos estatísticos. Trata-se de uma ciência interpretativa, profundamente ancorada na teoria da evolução, cujo método central — o **método cladístico** — se baseia na identificação de **sinapomorfias**, isto é, caracteres derivados partilhados por dois ou mais taxa e herdados de um ancestral comum imediato.



Figura 1. Capítulos de *Scabiosa atropurpurea* (Flora on, 2015) e de *Centaurea cyanus* (Jardim Botânico UTAD), exibindo as suas flores epigínicas.

Estas sinapomorfias constituem a única evidência válida para a delimitação de grupos **monofiléticos**, ou **clados** — conjuntos que incluem um ancestral e todos os seus descendentes. Qualquer outro tipo de agrupamento, seja **parafilético** (que exclui alguns descendentes) ou **polifilético** (que reúne linhagens sem ancestral comum recente), é rejeitado por distorcer a história evolutiva real. Um exemplo botânico simples, retirado da flora portuguesa, ilustra este ponto: a presença de um ovário ínfero — em que as peças florais se inserem acima do ovário — é uma sinapomorfia que une todas as espécies do clado das asterídeas, como a *Scabiosa atropurpurea* (escabiosa) e a *Centaurea cyanus* (uma espécie exótica).

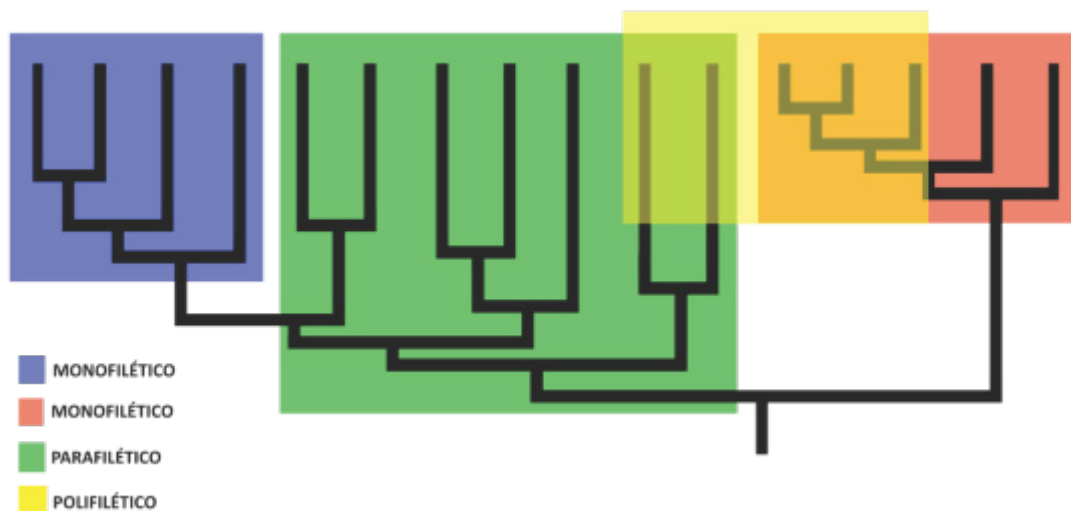


Figura 3. Árvore hipotética mostrando dois **clados** (azul e salmão) e 2 **grados** (verde e amarelo).

Já a semelhança na forma das folhas entre a *Cistus ladanifer* (esteva) e a *Halimium halimifolium* (sargaça-maior) — ambas com folhas coriáceas e tomentosas — não indica necessariamente parentesco próximo, mas pode ser uma adaptação convergente ao mesmo tipo de ambiente mediterrânico, ou seja, uma **homoplasia**. A tarefa do sistemata é, portanto, distinguir o que é fruto de herança comum do que é

fruto de coincidência adaptativa — e essa distinção é feita não por intuição, mas por inferência rigorosa, baseada em dados comparáveis e num método explícito.

Capítulo 2 — Como se escolhem e codificam os caracteres: a arte (e a ciência) por trás da matriz

A seleção e codificação de caracteres é a etapa mais crítica — e mais frequentemente negligenciada — de toda a análise filogenética. Não se trata de listar todas as diferenças observáveis entre as plantas, mas de identificar caracteres que sejam informativos para inferir relações

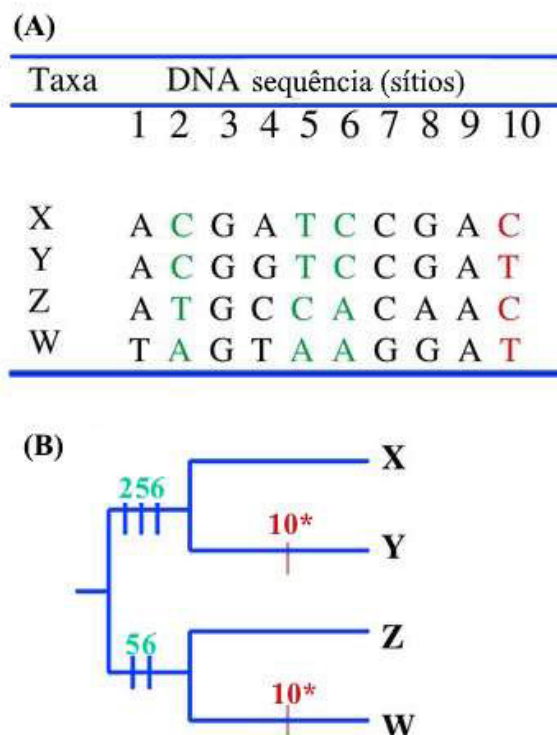


Figura 2. (A) Esquema simplificado de uma matriz de caracteres de dados de nucleótidos e (B) árvore filogenética montada a partir dos dados obtidos. Os sítios destacados a verde na matriz e na árvore correspondem aos sítios informativos, e aqueles a vermelho representam um suposto sítio de homoplasia.

de parentesco. Esses caracteres podem ser morfológicos (ex: tipo de inflorescência, presença de estípulas, número de lóculos no ovário), anatômicos (ex: tipo de estoma, padrão de nervação), palinológicos (ex: ornamentação do grão de pólen) ou moleculares

(ex: sequências de DNA dos genes *rbcl*, *matK* ou ITS).



Figura 4. Árvore desenraizada (unrooted), isto é, em que não foi usado um

definir os seus estados — as variações discretas que em um exemplo, o caracter “tipo de inflorescência” pode ter os estados 1 = umbela. A codificação de variações contínuas — exige “*outgroup*”. cuidado: deve-se evitar criar estados “longo”) sem um critério biológico claro; em muitos casos, a variação contínua em caracteres discretos com base em estados éticos.

Um passo crucial é a determinação da polaridade dos caracteres — ou seja, decidir qual estado é ancestral (**plesiomórfico**) e qual é derivado (**apomórfico**). Isso é feito geralmente por comparação com um grupo externo (**outgroup**), um taxon que se sabe estar próximo, mas fora do grupo em estudo. Por exemplo, ao estudar as relações entre espécies de *Lavandula* (alfazema), pode-se usar uma espécie de *Rosmarinus* (alecrim) como grupo externo. Se o alecrim possui inflorescência em espiga e todas as alfazemas também, então esse é provavelmente o estado ancestral; se apenas algumas alfazemas apresentam inflorescência em capítulo, esse é um estado derivado — e, se for compartilhado por um subconjunto de alfazemas, pode ser uma sinapomorfia que define um clado interno.



Figura 5. Diferença entre analogia e homologia. Em (A) está representada a asa de um morcego, em (B) a asa de uma águia e em (C) a asa de uma mosca. As três asas não surgiram por ancestralidade comum, mas por convergência ou paralelismo. Os mamíferos (D) (vaca) e (E) (mula) apresentam homologia quanto às glândulas mamárias e pelos porque têm uma ancestralidade comum recente.

A homologia primária — o reconhecimento de que uma estrutura em diferentes taxa é “a mesma coisa” evolutivamente — é uma inferência, não um dado. Decidir que o cálice gamossépalo da *Cistus salviifolius* é homólogo ao cálice gamossépalo da *Tuberaria lignosa* exige conhecimento de desenvolvimento, posição e conexão dos órgãos — e essa hipótese só será testada posteriormente, pela análise filogenética.

Capítulo 3 — Como se constrói a matriz de dados — e por que ela é o coração da análise

A matriz de caracteres é a representação formal das hipóteses de homologia e polaridade estabelecidas na etapa anterior. É uma tabela em que as linhas correspondem aos taxa (espécies, géneros, etc.) e as colunas correspondem aos caracteres codificados. Cada célula da matriz contém o estado do carácter para aquele taxon — por exemplo, “0”, “1”, “2”, ou “?” (para dados desconhecidos ou polimórficos).

Tabela 1 Tabela referente à codificação de um novo gene fotorreceptor – neochrome – em fetos. (Fay-Wei *et al.*, 2015; **CCO 1.0 Universal**)

Gene	GSK3 Group	root	leaf	shoot tip	seedling	stem	premeiotic flower/cone	postmeiotic flower/cone	open flower/cone	fruit/seed
AtSk12	I.1	1171,173	1369,389	1255,737	1496,503	1740,1	1543,194	1741,491	775,7536	
AtSK13	I.2	297,6193	150,7719	508,1273	296,4092	204,4016	895,07	501,1058	548,8987	564,2821
AtSK21	II.1	911,0573	891,8329	702,8005	661,8951	1109,992	556,86	520,8506	700,0856	527,013
AtSK22	II.1	489,35963798	288,9424	255,7171	387,1946	486,7		263,4807	264,6767	
AtSK23	II.1	465,6963	510,250995	497,6194	512,9411	701,73	531,9455	342,9738	09	
AtSK31	III	26341,5628,670	20,60309	24,04264	165,5989	21,37	176,3584	2194,571	664	
AtSK32	III	846,3768	562,59	433,9607	554,9998	636,2549	545,76	379,9331	493,2054	234
AtSK41	IV	579,9056	303,8114	276,9257	297,354	484,1589	419,21	302,283	262,2211	230,7889
AtSK42	IV	235,3798	275,1919	163,8532	215,1822	705,1833	200,84	240,5647	309,7219	294,9018
Nuphar_b4_c15797	I.1	80,66734	46,61361	49,07423	33,93296	38,44317	66,83579	61,54762	73,88134	
Nuphar_b4_c15797	I.2	16,382553204	757,20256	3,084815	13,45511		40,39062	9,80098	16,88716	
Nuphar_b4_c25205	II.1	35,376734237	2914	54,52692	15,42407	49,97612	38,46726	21,56216	54,88328	
Nuphar_b4_c24467	III	14,0291	34,96021	23,62833	6,16963	9,610793	35,17673	23,08036	3,920392	33,77433
Nuphar_b4_c10985	IV	129,7692	100,9962	142,67	234,4459	9284	110,8067	159,6391	111,7312	116,0993
Amborella_b5_c7151	I.1	228,2766	60,02807	71,184	NA	6754	94,83749	NA	92,21947	NA
Amborella_b5_c18184	I.1	5626,3901	123,14,0562	69,30973		56,4	144,688	NA	68,50589	
Amborella_b5_c18182	I.1	1058,6106	125,6568	55,478,7	NA	85,8177	69,30432	NA	102,7588	
Amborella_b5_c18004	I.2	51311,2	54,4	50,79299	44,10619	NA	56,45901	29,18076	NA	68,50589
Amborella_b5_c18507	III	52,35702	48,48421	114,9911	NA	88,07606	75,38364	NA	142,2815	
Amborella_b5_c5868	IV	69,11127	90,04211	55,13274	NA	81,30098	92,40576	NA	42,15747	NA
Aristolochia_b4_c4273	I.1	101,6888	69,27416	73,27287	66,70579	NA	237,4976	70,4435	5,818512	48,24409

É fundamental compreender que esta matriz não é um repositório neutro de observações, mas sim um modelo hipotético da evolução das estruturas em estudo. Cada escolha feita — desde a inclusão ou exclusão de um carácter até à forma como os seus estados foram definidos — influencia directamente o resultado da análise. Um erro

comum é tratar a matriz como um dado bruto, algo que “fala por si”. Na realidade, a matriz é o momento em que a biologia se transforma em dados — e essa transformação é profundamente interpretativa. Por exemplo, ao codificar a presença do indumento glandular em *Cistus ladanifer* como “1” e a sua ausência nas restantes Cistáceas como “0”, está-se a assumir que a ausência é o estado ancestral e que a presença é uma apomorfia. Se, numa análise posterior, se descobrir que o gene responsável pela produção das glândulas está ausente no genoma de algumas populações de *C. ladanifer*, isso poderá indicar que a característica é mais volátil do que se pensava — e que a codificação inicial foi inadequada.

A matriz deve, portanto, ser vista como um documento vivo, sempre sujeito a revisão. A sua construção exige honestidade intelectual: incluir caracteres que geram conflito não é um erro — é uma oportunidade para compreender melhor os processos evolutivos, como a convergência (quando estruturas semelhantes evoluem independentemente) ou a reversão (quando um carácter derivado é perdido e o estado ancestral reaparece). A matriz é, em última análise, o testamento do julgamento biológico do sistemata — e é nela, muito mais do que nos algoritmos, que reside o verdadeiro rigor da Sistemática Filogenética.

Capítulo 4 — Que programas usar, que métodos aplicar e como interpretar os resultados

Uma vez construída a matriz, esta é submetida a análise computacional utilizando programas especializados. Os mais utilizados na botânica são o Mesquite (ideal para construção e edição de matrizes, além de análises básicas), o TNT (rápido e eficiente para análises de parcimónia, especialmente com grandes matrizes morfológicas), o PAUP* (versátil, permite análises de parcimónia e máxima verosimilhança), o MrBayes (para inferência bayesiana) e o BEAST (para análises bayesianas que incorporam dados de tempo, como fósseis ou taxas moleculares).



Figura 6. Logos dos diferentes softwares referidos neste texto e cuja obtenção pode ser feita online e de forma gratuita..

Cada um destes programas aplica um método diferente para encontrar a árvore filogenética que melhor explica os dados. O **método da parcimónia** procura a árvore que exige o menor número total de mudanças evolutivas (passos) nos caracteres. É intuitivo

e eficaz para muitos conjuntos de dados morfológicos, mas pode falhar quando a homoplasia é abundante — como acontece frequentemente em caracteres adaptativos, como a forma das folhas em ambientes semelhantes.

A **máxima verosimilhança** utiliza modelos matemáticos explícitos de evolução — por exemplo, modelos que consideram que certas substituições de nucleótidos são mais prováveis do que outras (como o modelo GTR — General Time Reversible) — para calcular a árvore mais provável dada esses modelos. É particularmente adequada para dados moleculares. A **inferência bayesiana** vai mais longe: também usa modelos de evolução, mas incorpora “probabilidades a priori” (conhecimento prévio sobre taxas de evolução, idades de fósseis, etc.) para calcular a probabilidade de cada árvore ser a correcta, gerando uma distribuição de árvores plausíveis.

Nenhum destes métodos é “o melhor” em absoluto; a escolha deve ser guiada pela natureza dos dados e pela biologia do grupo. Por exemplo, para analisar sequências de DNA do gene *rbcL* em espécies de *Quercus* (carvalhos), um modelo de substituição como o GTR poderá ser mais adequado do que a simples parcimónia. O resultado da análise é uma ou mais árvores, geralmente representadas como **cladogramas** (sem escala de tempo) ou como **filogramas** (com ramos proporcionais ao número de mudanças). A interpretação biológica começa aqui: quais são os clados bem suportados? Quais as sinapomorfias que os definem? Onde ocorrem as homoplasias — e o que elas nos dizem sobre a evolução do grupo?

Capítulo 5 — Como se mede a confiança numa árvore: bootstrap, probabilidades e consenso

Obter uma árvore não é o fim, mas o início da análise. A questão crucial é: quão confiável é essa hipótese? Para responder, recorre-se a métodos de mensuração do apoio estatístico para cada ramo (nó) da árvore. O mais comum é o **bootstrap**, um método de reamostragem com reposição. Funciona assim: o programa gera, por exemplo, 1000 novas matrizes, cada uma criada ao seleccionar aleatoriamente caracteres da matriz original (alguns caracteres podem ser escolhidos várias vezes, outros nenhuma). Para cada uma dessas 1000 matrizes, é calculada uma árvore. Depois, verifica-se em quantas dessas árvores cada nó da árvore original aparece. Se o nó que une *Cistus ladanifer* e *Cistus salviifolius* aparecer em 950 das 1000 árvores, diz-se que tem 95% de apoio bootstrap. Isto não significa que há 95% de probabilidade de esse nó ser verdadeiro, mas sim que é robusto — persiste mesmo quando os dados são perturbados.

Nas análises bayesianas, obtém-se directamente uma “probabilidade a posteriori” para cada nó — por exemplo, 0.95. Este valor é mais intuitivo (representa a probabilidade de o nó ser correcto, dadas as premissas do modelo e os dados), mas depende fortemente da qualidade do modelo e das probabilidades a priori utilizadas.

Valores de bootstrap acima de 70% ou probabilidades bayesianas acima de 0.95 são geralmente considerados fortes.

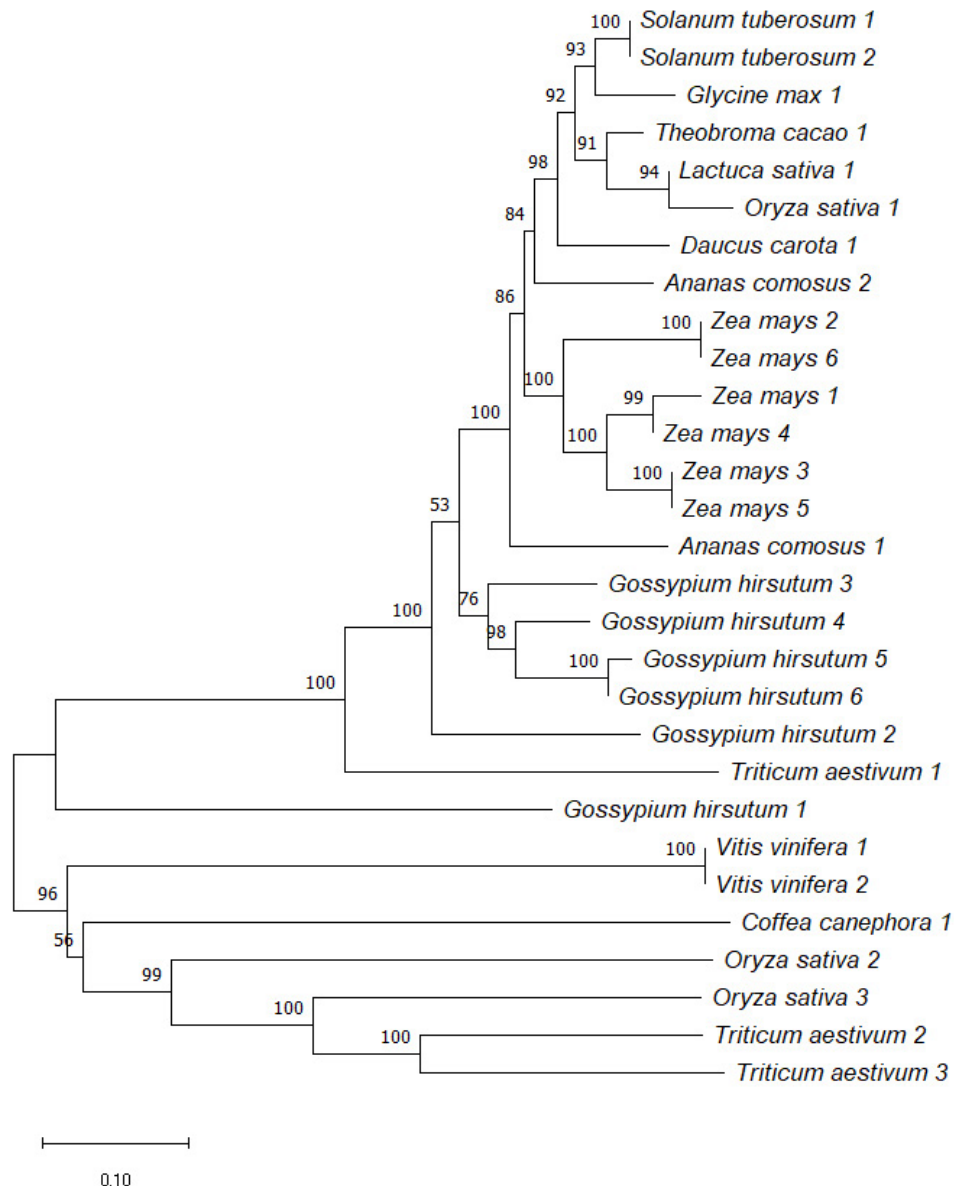


Figura 7. Árvore gerada com bootstrap, método de agrupamento de vizinhos (NJ) demonstrando relações filogenéticas, para sequências candidatas ao gene SERK, entre diversas espécies agronômicas: algodão (*Gossypium hirsutum*), videira (*Vitis vinifera*), abacaxi (*Ananas comosus*), arroz (*Oryza sativa*), cafeeiro (*Coffea canephora*), alface (*Lactuca sativa*), trigo (*Triticum aestivum*), milho (*Zea mays*), soja (*Glycine max*), cacau (*Theobroma cacao*), cenoura (*Daucus carota*) e batata-inglesa (*Solanum tuberosum*). (Santos et al. (2023) [CC BY 4.0](#))

Quando diferentes métodos (parcimônia, verossimilhança, bayesiana) ou diferentes conjuntos de dados (morfológico vs. molecular) produzem topologias conflitantes, não se deve forçar um consenso mecânico. Deve-se, isso sim, investigar a causa do conflito: pode ser devida a homoplasia não detectada, a um modelo evolutivo inadequado, a amostragem insuficiente, ou mesmo a processos biológicos reais como hibridização (comum em grupos como *Quercus* ou *Armeria*). A incongruência não é um erro — é uma pista.

Capítulo 6 — Como se transforma uma árvore numa classificação — e por que as velhas categorias não servem

A árvore filogenética obtida serve de base para a classificação científica. O princípio fundamental é o da monofilia: cada grupo taxonómico deve incluir um ancestral comum e todos os seus descendentes. Isto implica abandonar grupos parafiléticos — como o antigo conceito de “gimnospérmicas”, que exclui as angiospérmicas — ou polifiléticos — como o conceito de “plantas suculentas”, que reúne cactos (não nativos) e eufórbias como a *Euphorbia characias* (da flora portuguesa) apenas por causa de uma adaptação convergente.

Na prática, isto significa que a classificação deve reflectir directamente a topologia da

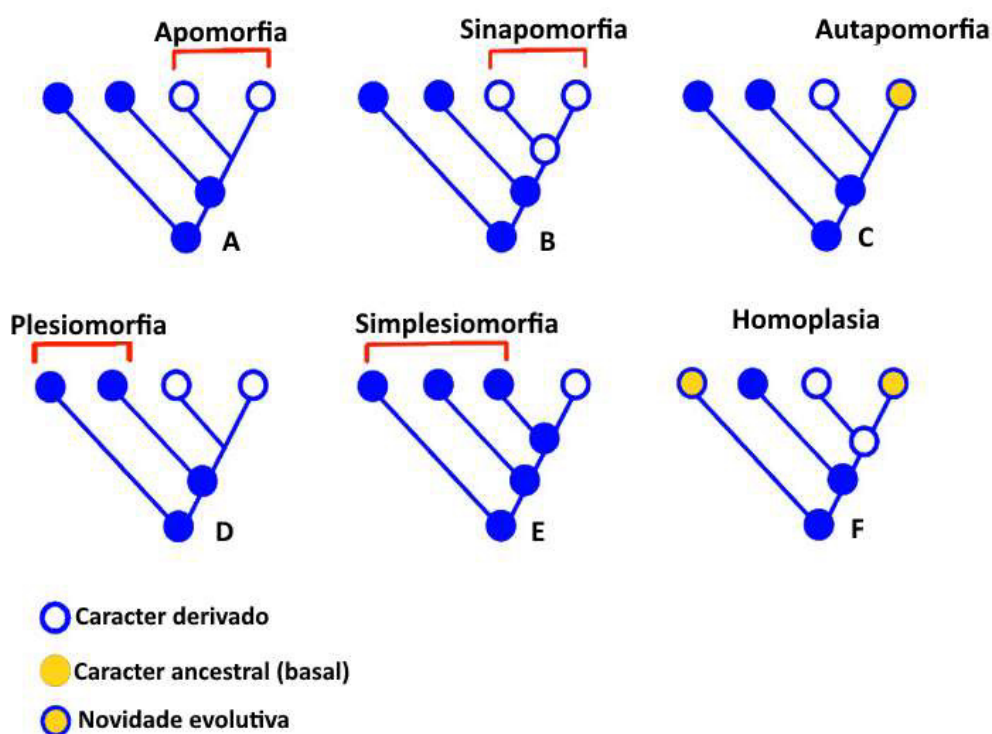


Figura 8. Tipos de caracteres utilizados nas árvores filogenéticas. Os diferentes caracteres derivados (apomorfia, sinapomorfia e autapomorfia) são mostrados nas árvores (A), (B) e (C). Quanto aos caracteres derivados (plesiomorfia e simplesiomorfia), são apresentados nas árvores (D) e (E). A árvore (F) apresenta a homoplasia, uma convergência.

árvore. As velhas categorias de Lineu — género, família, ordem — ainda são utilizadas por convenção, mas carecem de qualquer significado evolutivo intrínseco. Um “género” de orquídeas pode ser mais antigo do que uma “família” de gramíneas. Por isso, a tendência moderna é simplificar: manter “género” e “família” como rótulos úteis para comunicação, mas abandonar as categorias superiores (ordem, classe, filo) ou, pelo menos, deixar de as tratar como se tivessem um valor comparativo universal.

O importante é que cada nome corresponda inequivocamente a um clado — um e apenas um —, definido por uma ou mais sinapomorfias. Por exemplo, em vez de

dizer que a *Digitalis purpurea* (dedaleira) pertence à “família **Scrophulariaceae**” (numa circunscrição tradicional e parafilética), a classificação moderna coloca-a nas **Plantaginaceae**, num clado monofilético bem suportado por dados moleculares. A classificação filogenética não é um catálogo estático; é um sistema dinâmico de hipóteses testáveis, sempre sujeito a revisão à luz de novas evidências.

Capítulo 7 — O que fazer depois: revisão, publicação, comunicação e próximos passos

Concluída a análise, o trabalho do sistemata está longe de terminar. A árvore filogenética obtida é uma hipótese — não uma verdade absoluta. Deve ser apresentada de forma clara, com figuras que mostrem os clados e as sinapomorfias que os sustentam, e acompanhada de uma descrição detalhada dos caracteres utilizados, dos métodos aplicados e das decisões tomadas (especialmente sobre homologia e codificação).

As matrizes de dados e as árvores devem ser disponibilizadas em repositórios públicos, como o *MorphoBank* (para dados morfológicos) ou o *TreeBASE* (para dados moleculares e árvores), para que outros investigadores as possam testar, replicar ou expandir. A publicação deve incluir uma discussão crítica: quais as limitações do estudo? Quais os caracteres que geraram mais conflito? Quais as implicações biológicas das homoplasias detectadas?

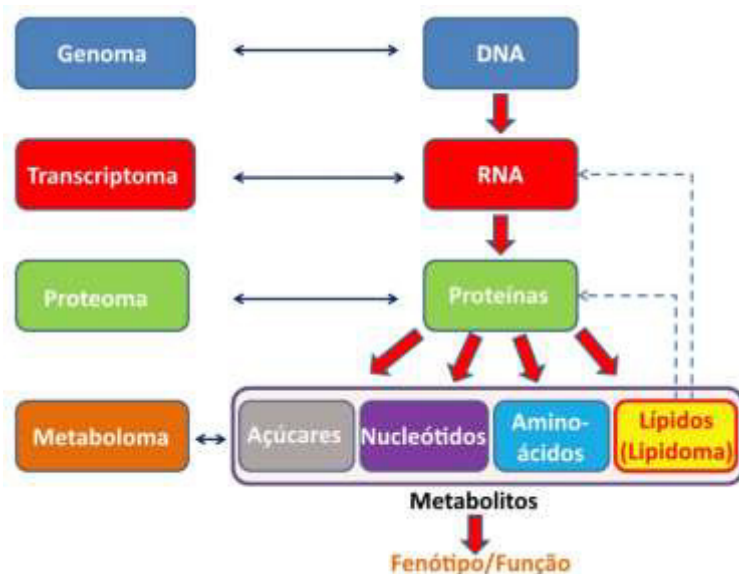


Figura 9. Esquema geral que mostra as relações do genoma, transcriptoma, proteoma e metaboloma (lipidoma). (CC BY-SA 3.0, Lmaps, 2009)

A Sistemática Filogenética não é uma ciência de respostas fechadas, mas de perguntas cada vez melhores. Um bom trabalho abre caminho para os próximos: a inclusão de mais taxa, a adição de novos tipos de dados (ex: transcriptomas, dados de desenvolvimento), a integração com outras disciplinas como a biogeografia (para

estudar como as plantas se dispersaram, por exemplo, pela Península Ibérica) ou a ecologia (para entender como as adaptações evoluíram em resposta ao clima mediterrânico). É esta transparência, este rigor, esta coragem de assumir que toda a classificação é uma hipótese — e não uma verdade revelada — que faz da Sistemática Filogenética uma das disciplinas mais exigentes, mas também mais gratificantes, da biologia contemporânea.

Ao transformar a aparente diversidade caótica da flora— desde os fetos húmidos do Norte até aos arbustos aromáticos do Sul— numa narrativa coerente e testável, esta ciência cumpre, talvez, a sua função mais nobre: dar sentido à vida, não por decreto, mas por investigação.

Bibliografia

Fay-Wei, L. *et al.* (2014). Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *PNAS*, 111 (18): 6672-6677 <https://doi.org/10.5061/dryad.fn2rg>

Santos, L. *et al.* (2023). Desenvolvimento de marcadores estádios-específicos para efeitos morfogenéticos in vitro por genes candidatos *in silico* em espécies frutíferas do cerrado. *Bionorte*. 12: 333-345. <https://doi.org/10.47822/bn.v12i1.502>

Referências Bibliográficas

APG IV. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. *Botanical Journal of the Linnean Society*, 181(1), 1–20. <https://doi.org/10.1111/boj.12385>

Darwin, C. (1859). *On the Origin of Species*. John Murray.

Futuyma, D. J. (2013). *Evolution* (3rd ed.). Sinauer Associates.

Hennig, W. (1966). *Phylogenetic Systematics*. University of Illinois Press.

Hillis, D. M., Moritz, C., & Mable, B. K. (Eds.). (1996). *Molecular Systematics* (2nd ed.). Sinauer Associates.

Judd, W. S., Campbell, C. S., Kellogg, E. A., Stevens, P. F., & Donoghue, M. J. (2016). *Plant Systematics: A Phylogenetic Approach* (4th ed.). Sinauer Associates.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772), 853– 858.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12), 1572–1574. <https://doi.org/10.1093/bioinformatics/btg180>

Sanderson, M. J., & Hufford, L. (Eds.). (1996). Homoplasy: The Recurrence of Similarity in Evolution. Academic Press.

Stevens, P. F. (2001 em diante). Angiosperm Phylogeny Website. Versão 14; continuamente atualizado. Missouri Botanical Garden.
<http://www.mobot.org/MOBOT/research/APweb/>

Swofford, D. L., Olsen, G. J., Waddell, P. J., & Hillis, D. M. (1996). Phylogenetic inference. In D. M. Hillis, C. Moritz, & B. K. Mable (Eds.), Molecular Systematics (pp. 407–514). Sinauer Associates.

Wiens, J. J. (2003). Incomplete taxa, incomplete characters, and phylogenetic accuracy: Is there a missing data problem? *Journal of Systematics and Evolution*, 41(3), 205–214.

Yang, Z. (2014). Molecular Evolution: A Statistical Approach. Oxford University Press.

Flora-On: Scabiosa — <https://flora-on.pt/?q=Scabiosa>
 Jardim Botânico UTAD - https://jb.utad.pt/especie/Centaurea_cyanus

UM CURTO GLOSSÁRIO

- Ancestral comum mais recente: O ancestral mais próximo que une todos os últimos terminais, nós e ramos dele derivados.
- Autoapomorfia: Um estado de carácter derivado exclusivo de um terminal (clado final).
- Apomorfia: Um estado de carácter derivado numa reconstrução filogenética.
- Árvore filogenética: Uma representação gráfica das relações filogenéticas entre diferentes terminais.
- Caracter: Qualquer característica ou qualidade de um taxon.
- Clado: Um conjunto de organismos relacionados numa reconstrução filogenética que partilham um antepassado comum.
- Estado de Caracter: Cada um dos estados que um carácter pode apresentar.
- Evidência total: Uma análise filogenética baseada na combinação de todos os caracteres disponíveis, mesmo que sejam de natureza diferente. É frequentemente utilizada para se referir à combinação de caracteres morfológicos e moleculares ou, em geral, a várias partições.
- Filogenia: A história evolutiva de um grupo de taxa.
- Filogenómica: A disciplina que estuda as relações filogenéticas numa fracção significativa do genoma de um organismo.
- Genómica comparativa: O estudo dos genomas através de comparações entre espécies ou populações.

- Grupo basal: O grupo localizado na base de uma árvore filogenética.
- Grupo derivado: Um grupo resultante de uma divergência interna da árvore filogenética. 14
- Grupo externo: Um grupo considerado a priori próximo do grupo de estudo e que serve para polarizar uma filogenia (**outgroup**).
- Grupo irmão: Um grupo que partilha o antepassado comum mais recente com o grupo de estudo.
- Grupo interno: Um grupo de estudo que é considerado a priori com base em informação taxonómica e/ou filogenética.
- Homologia: Uma característica ou estado de carácter que aparece em dois organismos ou linhagens diferentes e que tem a mesma origem evolutiva, ou seja, tem origem no mesmo antepassado comum mais recente.
- Homoplasia: Uma característica ou estado de carácter adquirido independentemente em dois ou mais organismos ou linhagens.
- Índice de consistência: Um valor que indica os níveis de homoplasia num conjunto de dados após a realização de uma reconstrução filogenética.
- Nó: Um ponto de divergência entre dois clados ou ramos, representando o antepassado comum.
- Plesiomorfia: Um estado de carácter presente no grupo externo e nos antepassados.
- Polaridade: A direção de transição entre estados de carácter, de um nó ancestral para um derivado.
- Raiz: Um nó que dá origem a todos os nós terminais.
- Relógio molecular: Uma metodologia utilizada para datar nós numa árvore filogenética com base na divergência entre marcadores moleculares. Ou seja, deduz o tempo decorrido a partir do número de diferenças entre sequências de ADN (nucleótido) ou de proteína (aminoácido).
- Simplesiomorfia: Uma plesiomorfia partilhada entre os terminais de um clado. Sinapomorfia: Uma apomorfia partilhada por todos os terminais de um clado.
- Taxon: Uma unidade taxonómica de qualquer hierarquia.
- Topologia: Um padrão de relações espaciais entre terminais numa árvore filogenética.