

Detectando al flujo génico en la era genómica

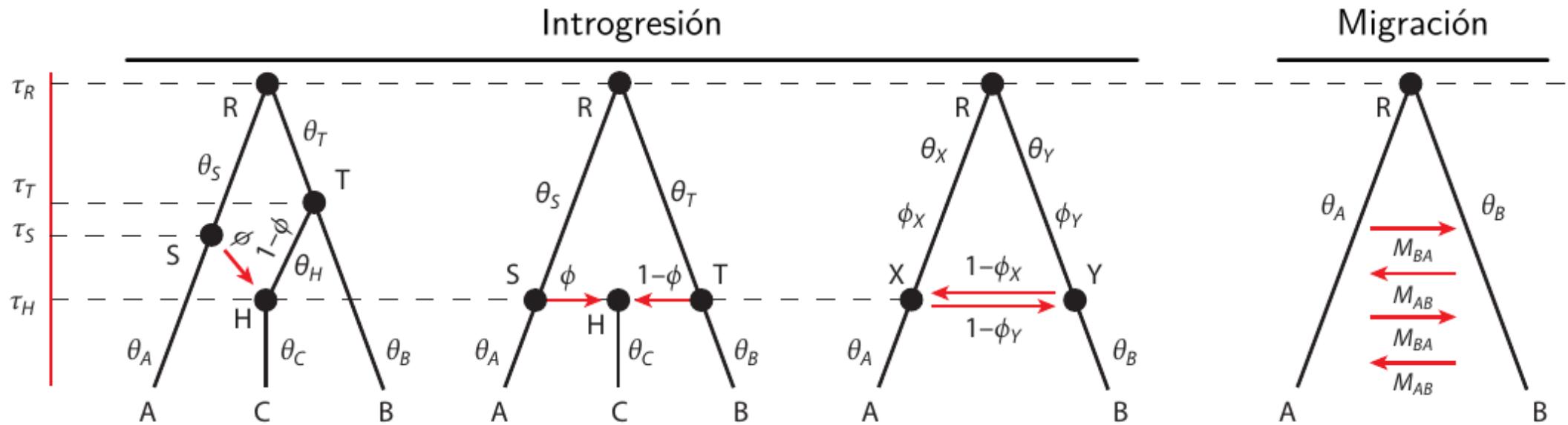
Kevin I. Sánchez; Mariana Morando

IPEEC-CONICET (Pto. Madryn, Chubut)

5 Oct. 2023

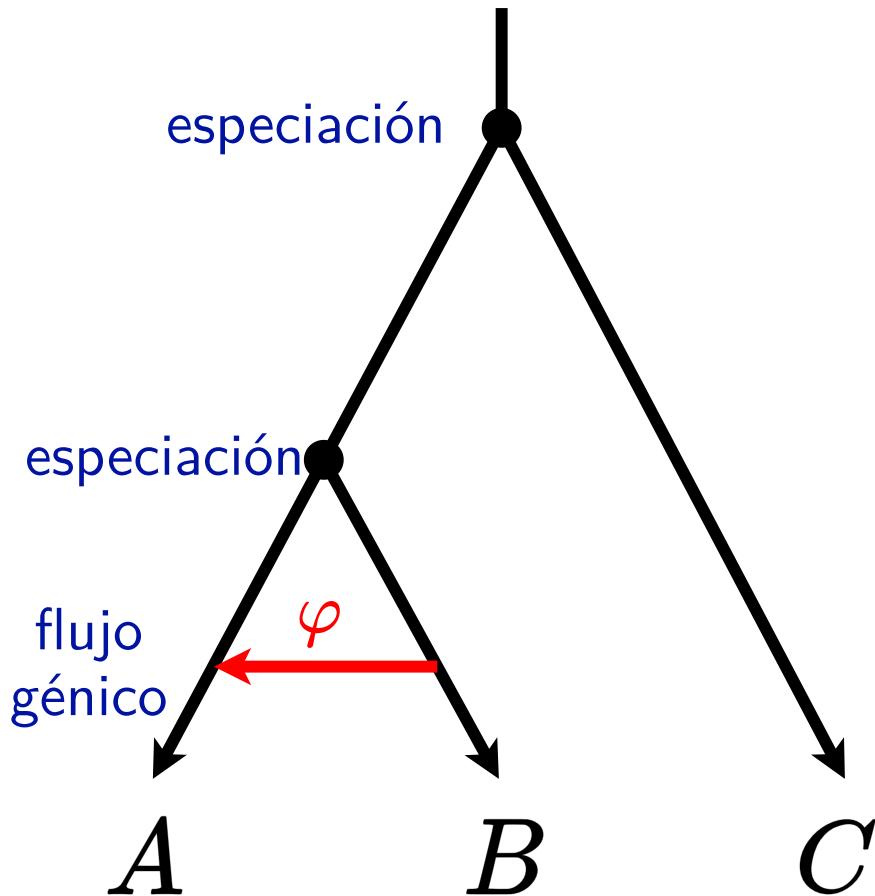
Definiciones

Flujo génico → reproducción e/linajes divergentes (poblaciones/especies)



Modif. de Edelman y Mallet (2021)

Redes filogenéticas → nodos de especiación + nodos de hibridación

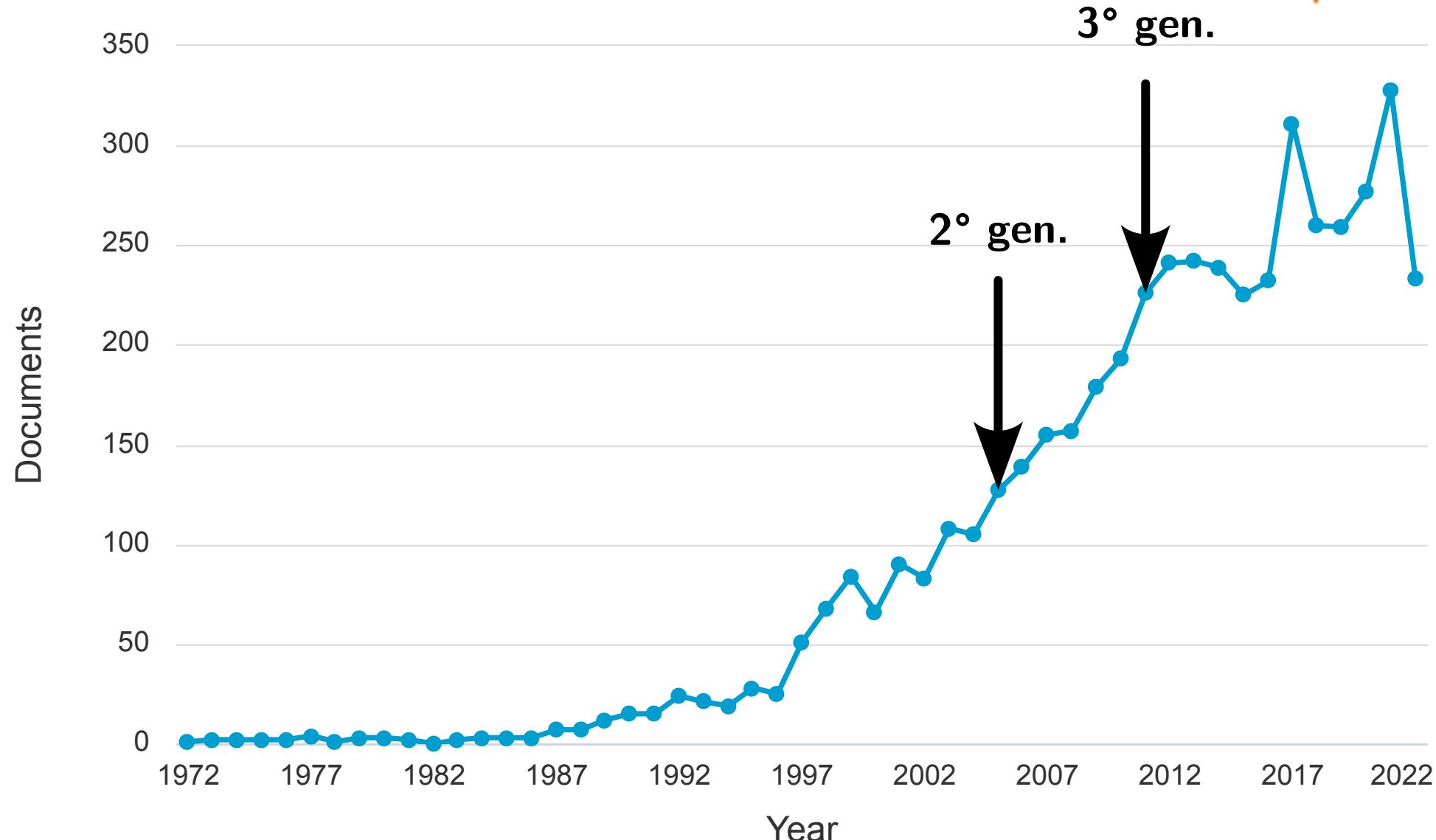


φ (pl. Φ) → probabilidad de introgresión

Importancia de considerar al flujo génico

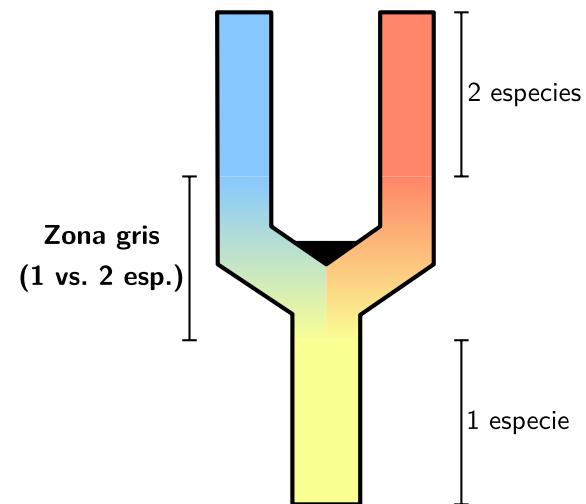
introgression OR “gene flow”

Scopus



Flujo génico influye en...

- Parámetros → Ψ (topología), θ (tam. pob.), τ (tiempo diverg.)
- **Variabilidad** genética (p. ej. introgresión diferencial) y fenotípica (p. ej. evolución transgresiva)
- Radiaciones **adaptativas** (Seehausen 2004, Edelman y Mallet 2021)
- Cambio de paradigma en **concepto de especie** (etapas tempranas)

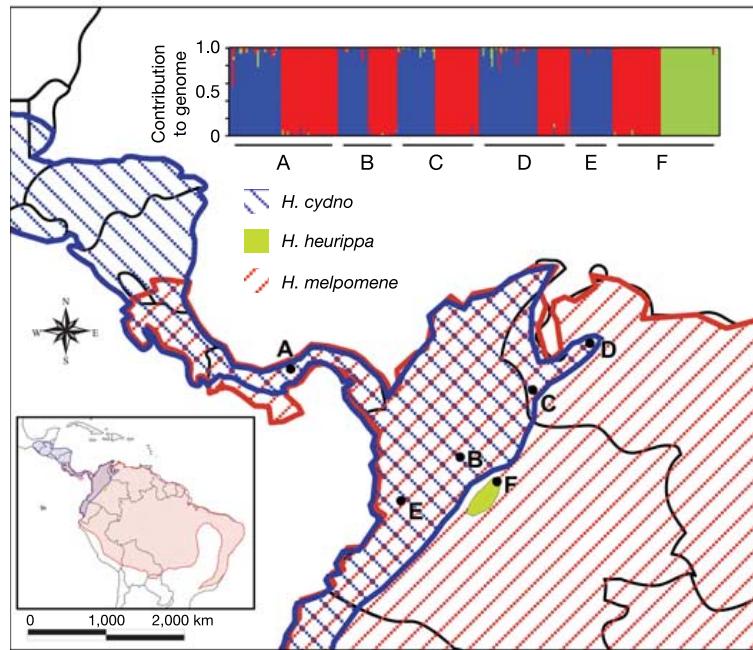


Cómo detectar y caracterizar al flujo génico?

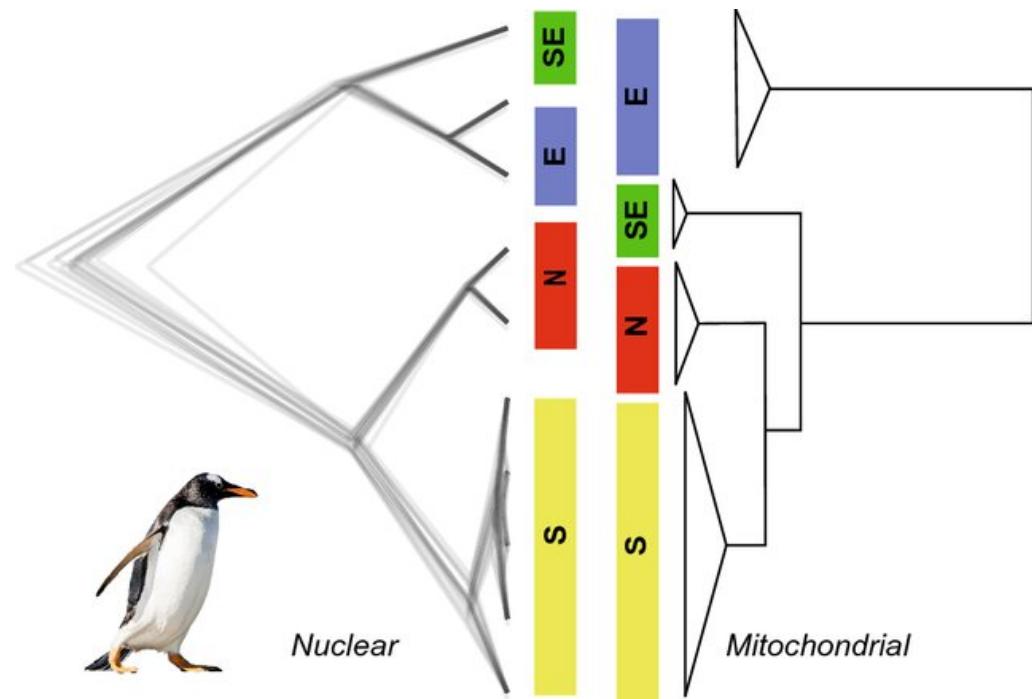
1. Métodos indirectos

1. Métodos indirectos

- Fenotipos intermedios



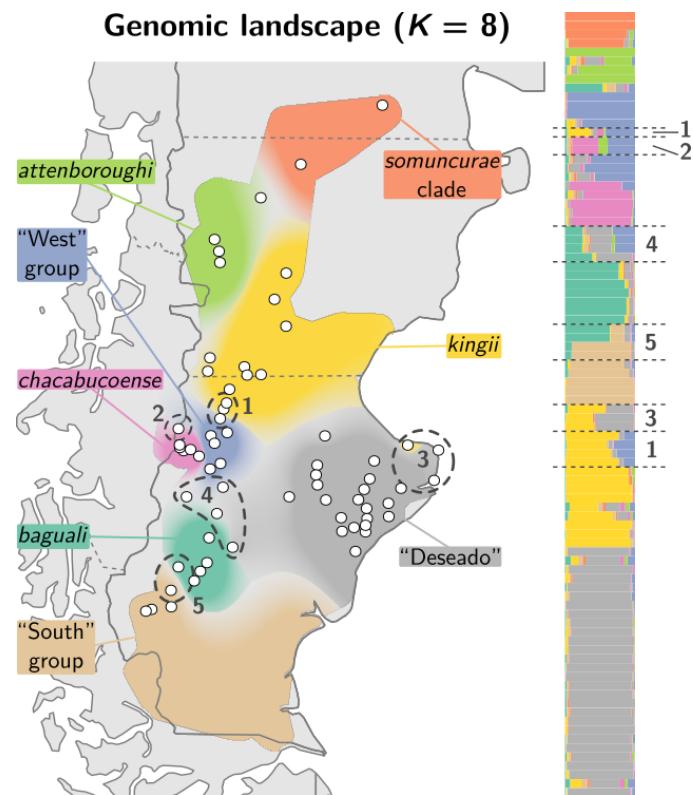
- Discordancias mt-nc



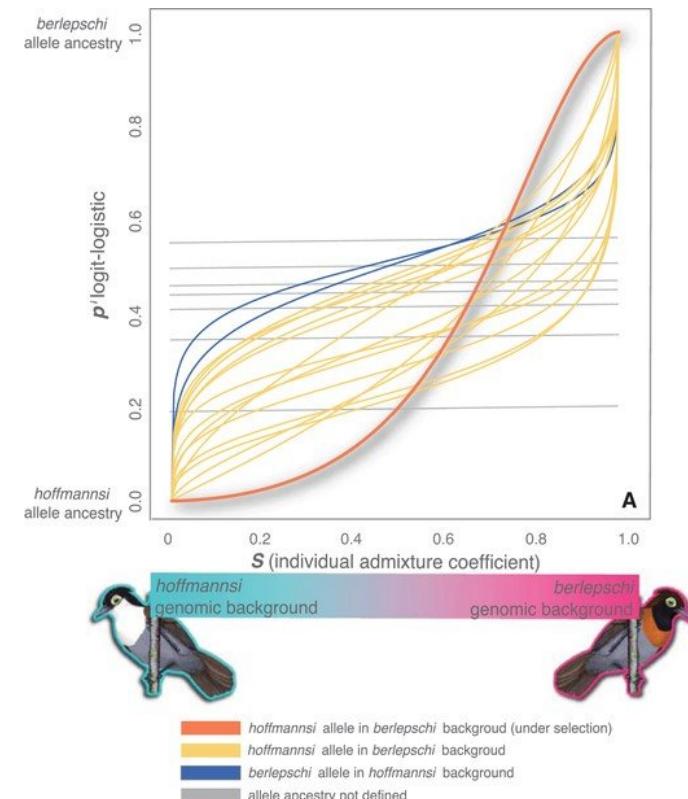
Noll et al. (2022)

1. Métodos indirectos

- Asignación poblacional



- Clinas genómicas

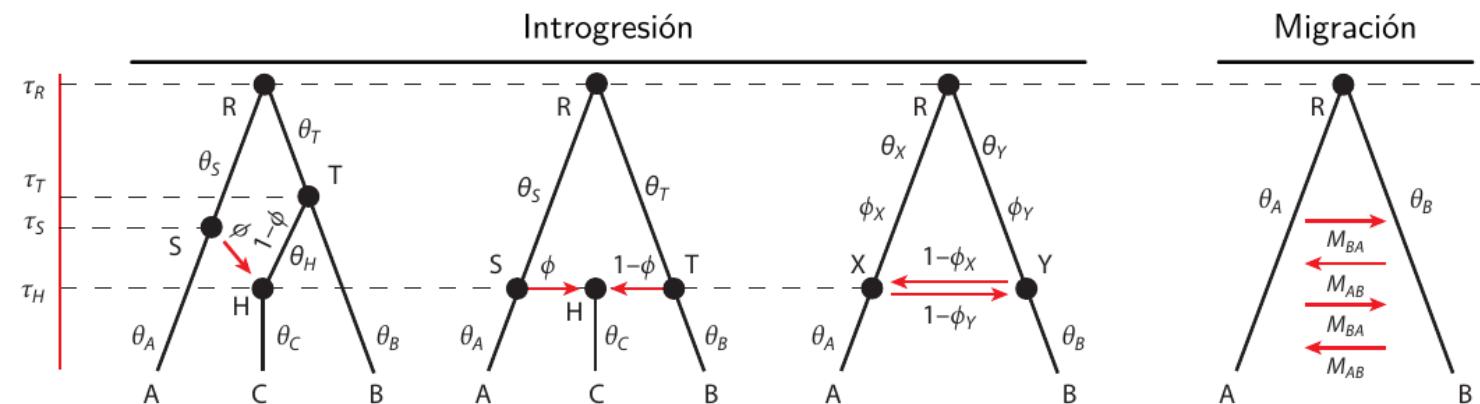


2. Métodos directos

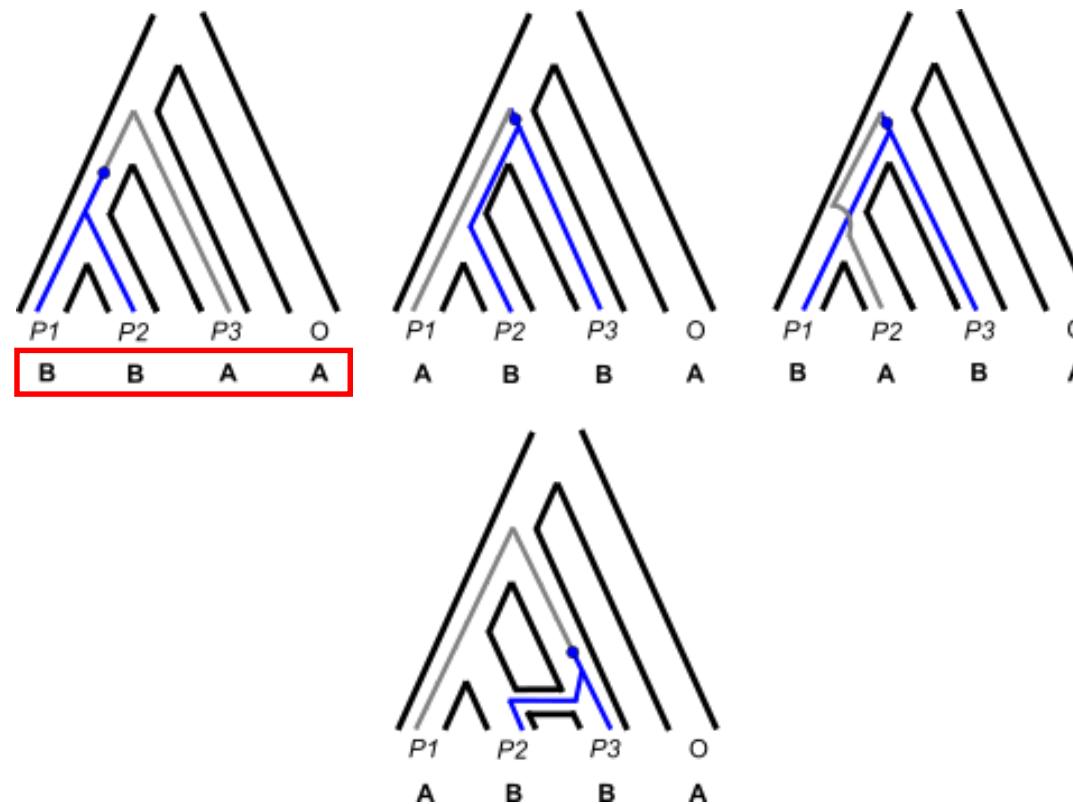
Heurísticos → resumen de secuencias

- Patrones de sustituciones (SNPs bialélicos)
- Árboles de genes (basados en MCEM + i)

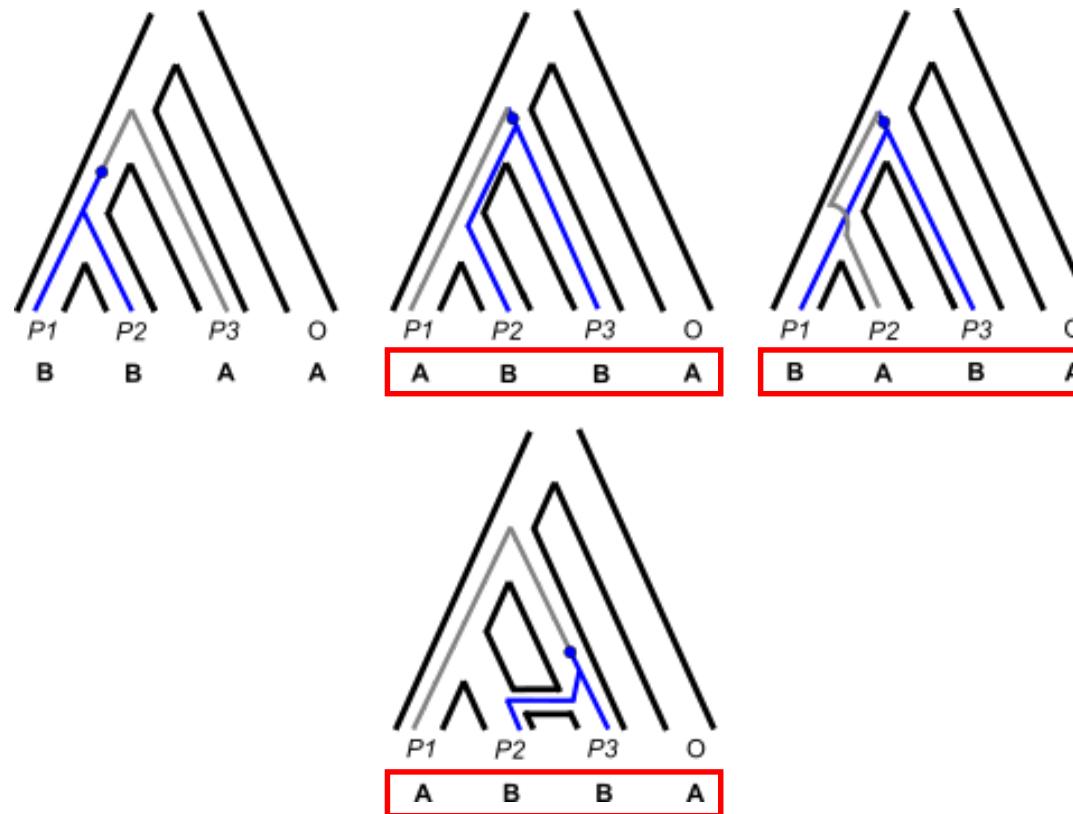
Verosimilitud completa → uso de secuencias



Estadístico D o prueba ABBA-BABA (Green et al. 2010)



Estadístico D o prueba ABBA-BABA (Green et al. 2010)



Heurísticos → patrones de sustitución

$$\Psi \quad P\left(\begin{array}{c} \diagup \quad \diagdown \\ \dots & \dots \\ P_1 & P_2 & P_3 & O \end{array}\right) \approx P\left(\begin{array}{c} \diagup \quad \diagdown \\ \dots & \dots \\ P_2 & P_1 & P_3 & O \end{array}\right)$$

$$\Psi_1 \quad P\left(\begin{array}{c} \diagup \quad \diagdown \\ \dots & \dots \\ P_1 & P_2 & P_3 & O \end{array}\right) < P\left(\begin{array}{c} \diagup \quad \diagdown \\ \dots & \dots \\ P_2 & P_1 & P_3 & O \end{array}\right)$$

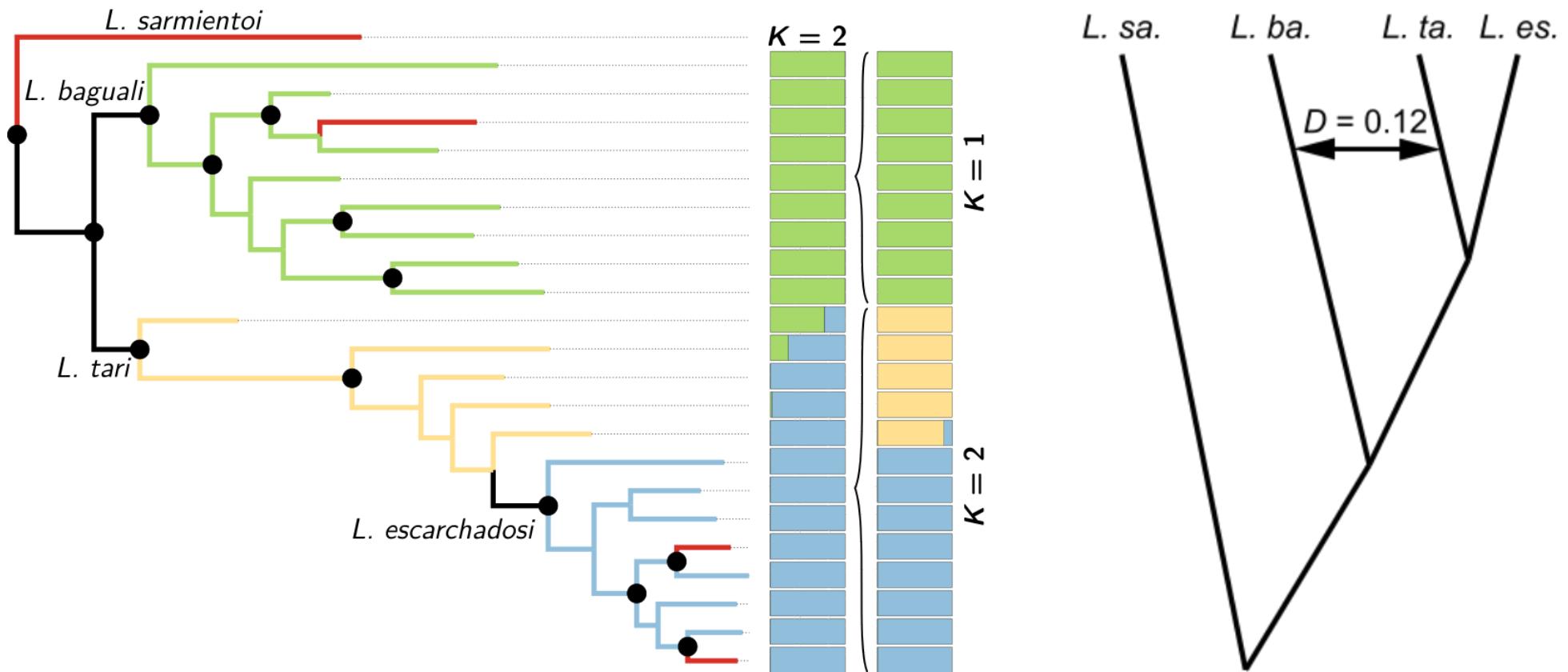
$$\Psi_2 \quad P\left(\begin{array}{c} \diagup \quad \diagdown \\ \dots & \dots \\ P_1 & P_2 & P_3 & O \end{array}\right) > P\left(\begin{array}{c} \diagup \quad \diagdown \\ \dots & \dots \\ P_2 & P_1 & P_3 & O \end{array}\right)$$

$$D = \frac{n_{ABBA} - n_{BABA}}{n_{ABBA} + n_{BABA}}$$

Heurísticos → patrones de sustitución

- Detecta **presencia** de flujo génico
- Medida a nivel de genoma o en ventanas
- Requiere corrección por pseudo-replicación (cuartetos)
- No detecta flujo génico entre taxa hermanos
- Rápido (1000s SNPs)

Heurísticos → patrones de sustitución

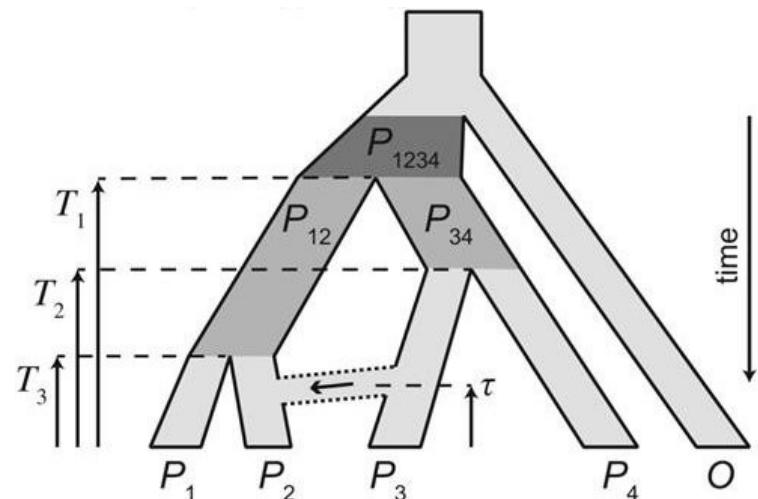


Sánchez et al. (2021)

Heurísticos → patrones de sustitución

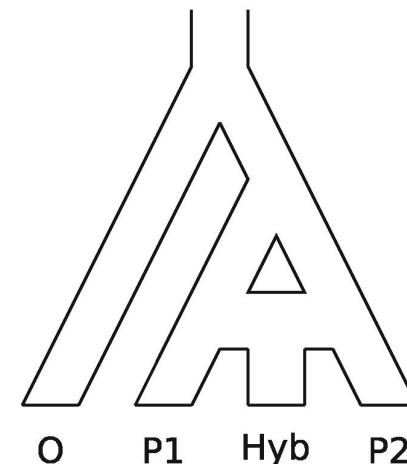
D_{FOIL} (Pease y Hahn 2015)

- Detecta presencia y dirección
- Árbol simétrico de 5 taxa



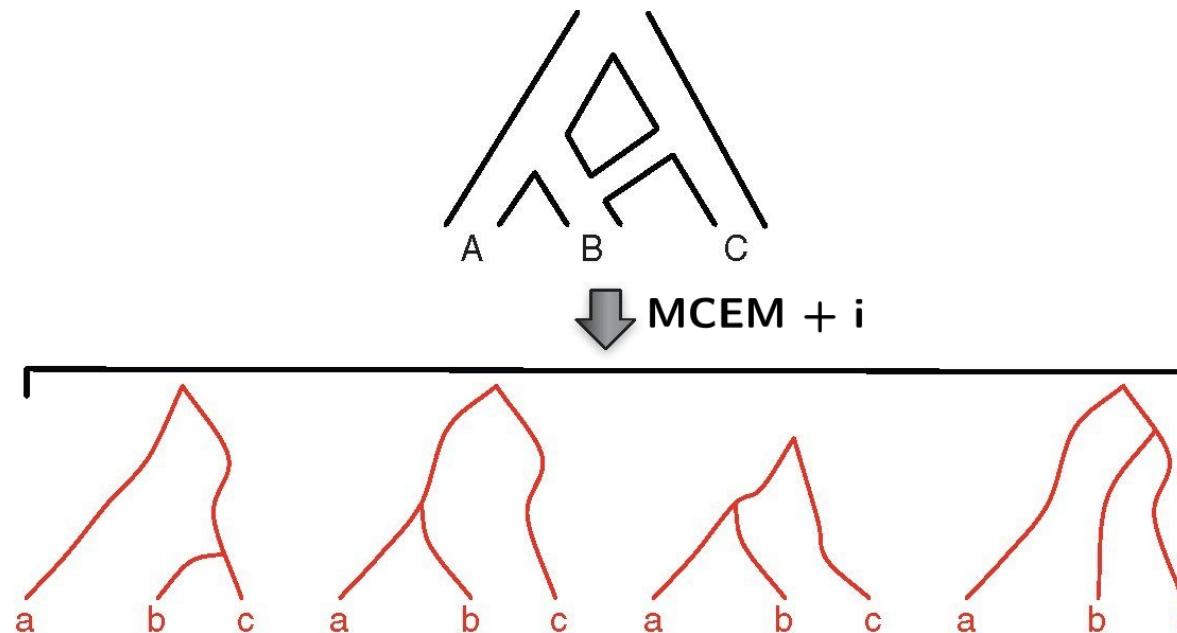
HyDe (Blischak, 2018)

- Detecta presencia, dirección y magnitud



$$\hat{\varphi} = \frac{n_{BBAA} - n_{BABA}}{n_{BBAA} - 2n_{BABA} + n_{ABBA}}$$

Modelo Coalescente de Especies Múltiples + introgresión (Yu et. al. 2014)

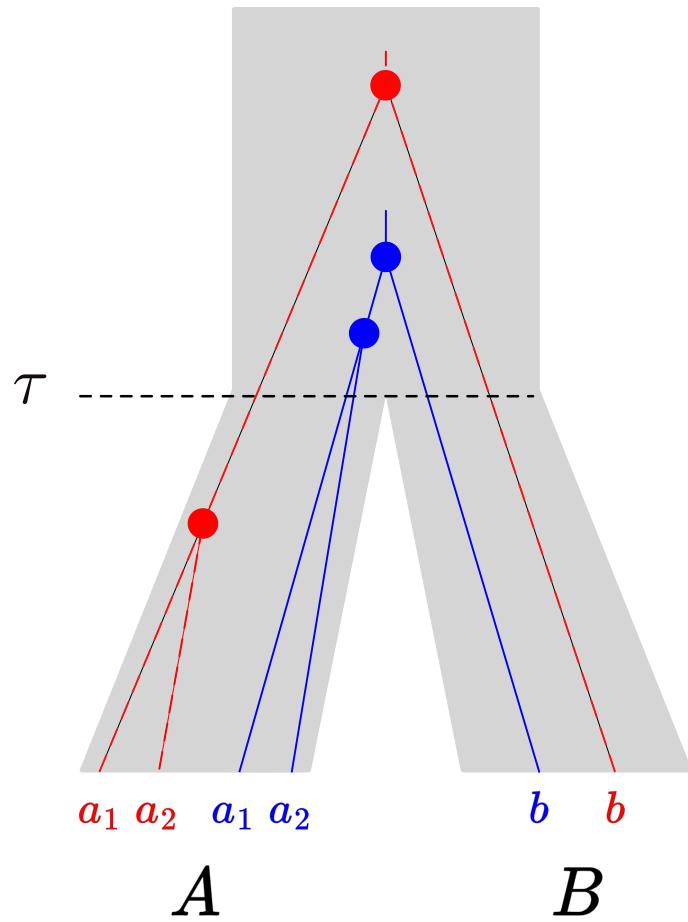


Modif. de Wen y Nakhleh (2018)

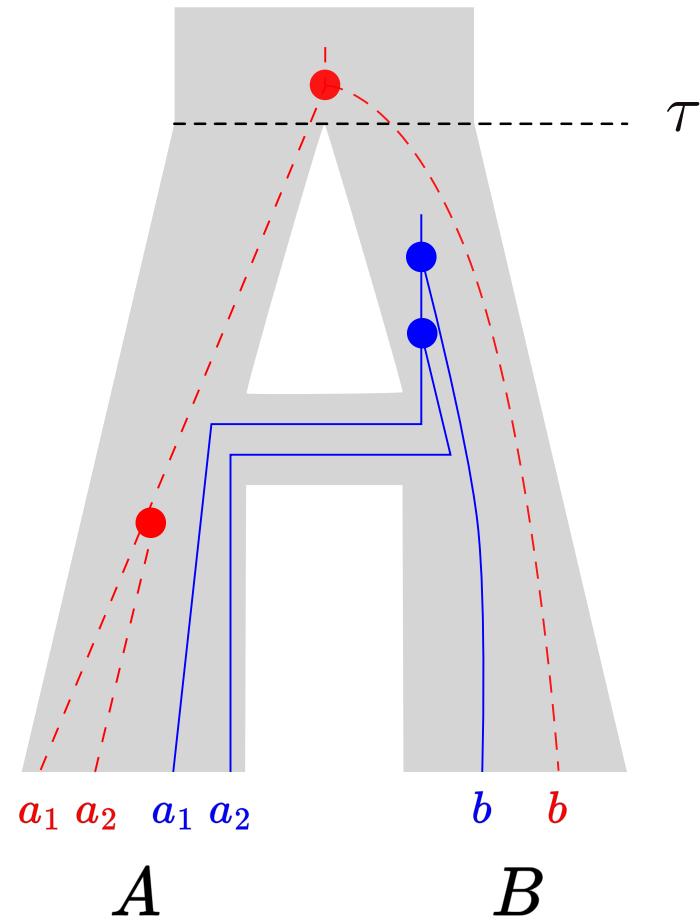
Parámetros: $\Psi, \varphi, \theta, \tau$

Heurísticos → árboles de genes

MCEM

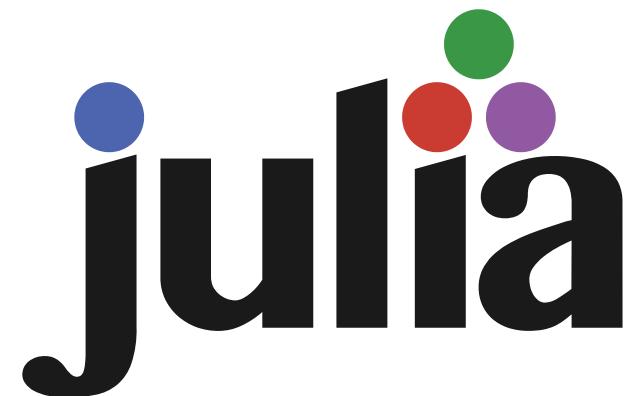
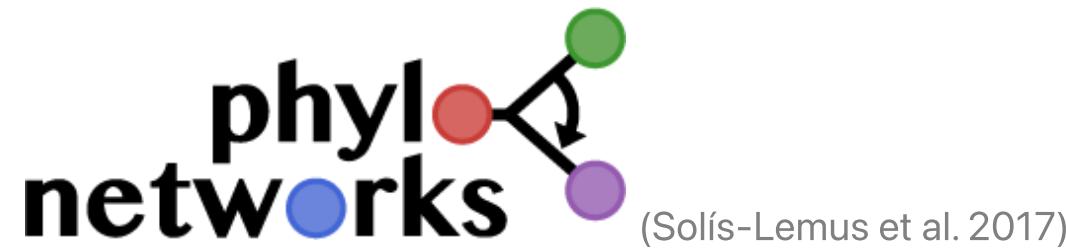


MCEM + i



SNaQ (Solís-Lemus y Ané 2016)

Species Networks applying Quartets



Heurísticos → árboles de genes

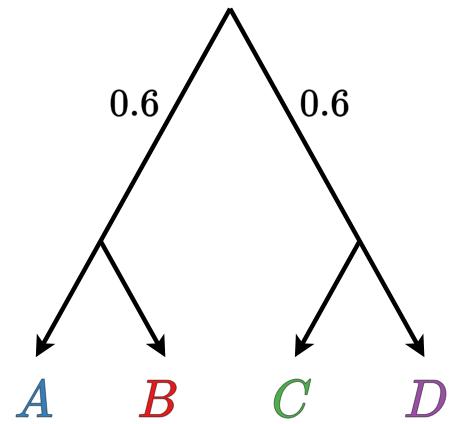
- **Input** → árboles de genes → Factores de Concordancia (FC)

Taxon 1	Taxon 2	Taxon 3	Taxon 4	FC _{12_34}	FC _{13_24}	FC _{14_23}
A	B	C	D	0.2	0.6	0.2
A	B	C	E	0.3	0.5	0.2
...

$$\binom{n}{4}$$

- Detecta presencia, dirección, magnitud y topología

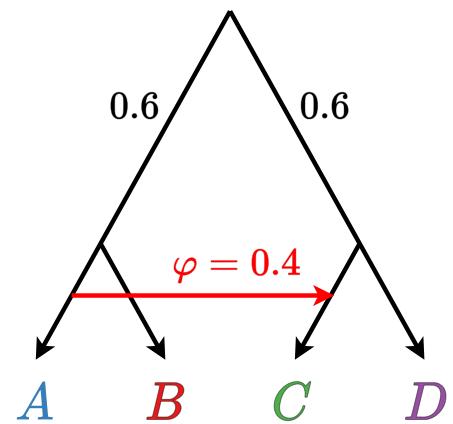
Heurísticos → árboles de genes



$$P\left(\begin{array}{c} A \\ | \\ B \end{array} \middle| \begin{array}{c} C \\ | \\ D \end{array}\right) \approx 0.8$$

$$P\left(\begin{array}{c} A \\ | \\ C \end{array} \middle| \begin{array}{c} B \\ | \\ D \end{array}\right) \approx 0.1$$

$$P\left(\begin{array}{c} A \\ | \\ D \end{array} \middle| \begin{array}{c} B \\ | \\ C \end{array}\right) \approx 0.1$$



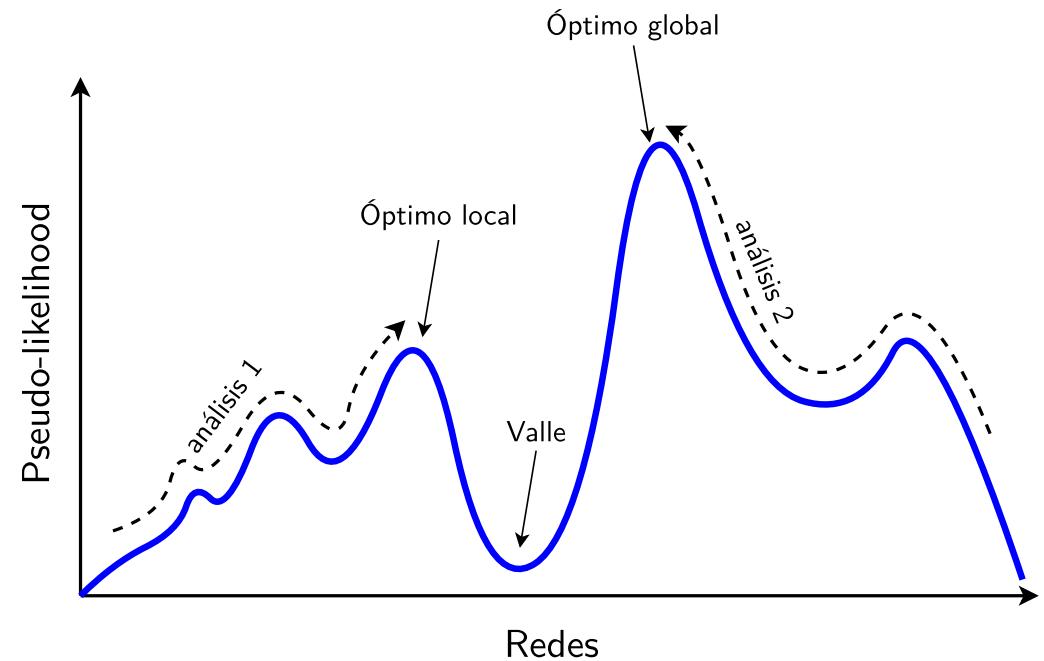
$$P\left(\begin{array}{c} A \\ | \\ B \end{array} \middle| \begin{array}{c} C \\ | \\ D \end{array}\right) \approx 0.2$$

$$P\left(\begin{array}{c} A \\ | \\ C \end{array} \middle| \begin{array}{c} B \\ | \\ D \end{array}\right) \approx 0.6$$

$$P\left(\begin{array}{c} A \\ | \\ D \end{array} \middle| \begin{array}{c} B \\ | \\ C \end{array}\right) \approx 0.2$$

Heurísticos → árboles de genes

- “hill-climbing” → optimización de Ψ , φ y largos de ramas
- Afectado por errores en estimación de árboles de genes (tabla FC)
- No todos los parámetros son identificables (θ)

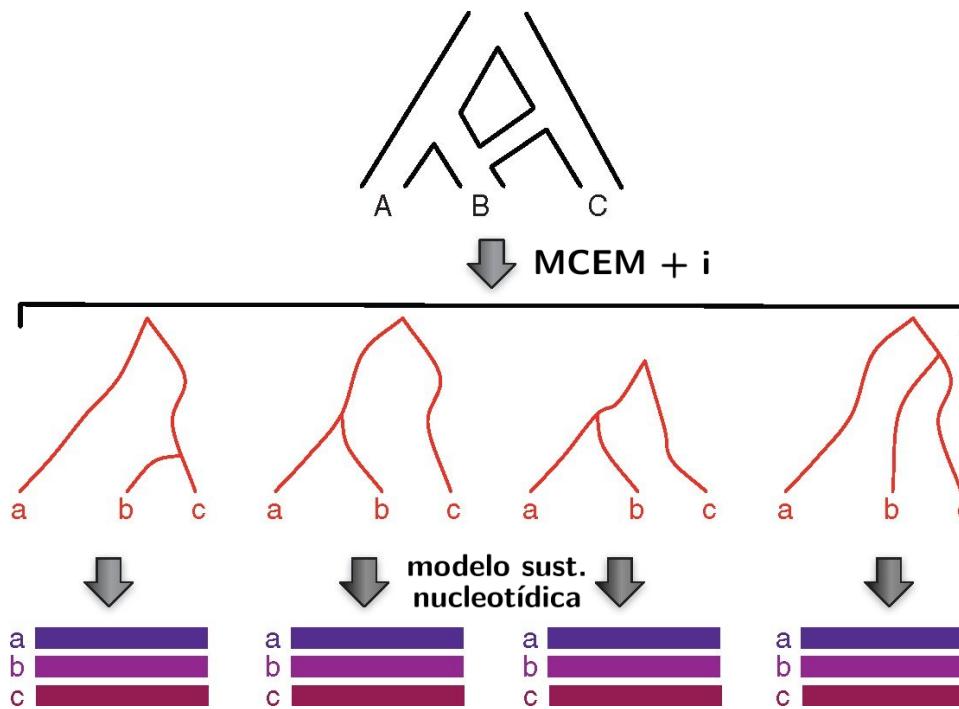


$$\max \mathcal{L} = \max \left\{ \prod_{s \in S} (\text{FC}_{q_1})^{X_{q_1}} \times (\text{FC}_{q_2})^{X_{q_2}} \times (\text{FC}_{q_3})^{X_{q_3}} \right\}$$

Variantes

- **PHYLONET** (Wen et al. 2018b):
 - **InferNetwork_MPL** (Yu y Nakhleh 2015) → uso de tripletes
 - **InferNetwork_ML** (Yu et al. 2014) → estima todos los parámetros
 - **MCMC_GT** (Wen et al. 2016)
 - **MLE_BiMarkers** (Zhu y Nakhleh 2018)
- **NANUQ** (Allman et al. 2019)

Verosimilitud completa

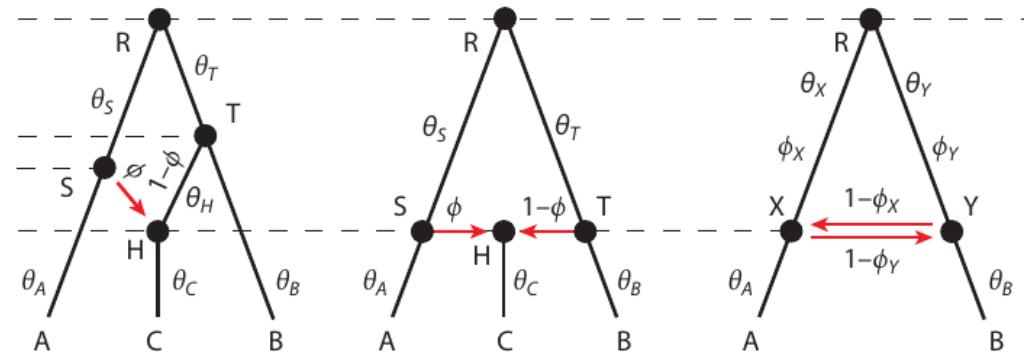


Modif. de Wen y Nakhleh (2018)

$$P(\Psi, \Phi, G | S) \propto P(\Psi) P(\Phi) \prod_{i=1}^m P(s_i | g_i) P(g_i | \Psi, \Phi)$$

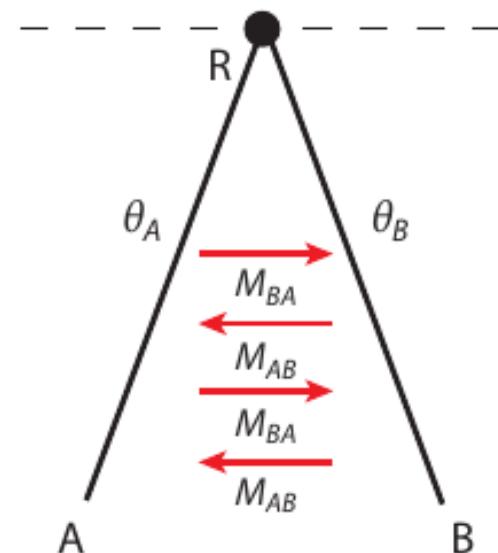
Modelo MCEM + i

- PHYLONET: MCMC_SEQ (Wen y Nakhleh 2018a), MCMC_BiMarkers (Zhu et al. 2018)
- SPECIESNETWORKS (Zhang et al. 2018) → *birth-hybridization* (λ, ν)
- BPP (Flouri et al. 2020) → modelos especificados
- PHYNEST (Kong et al. 2022)

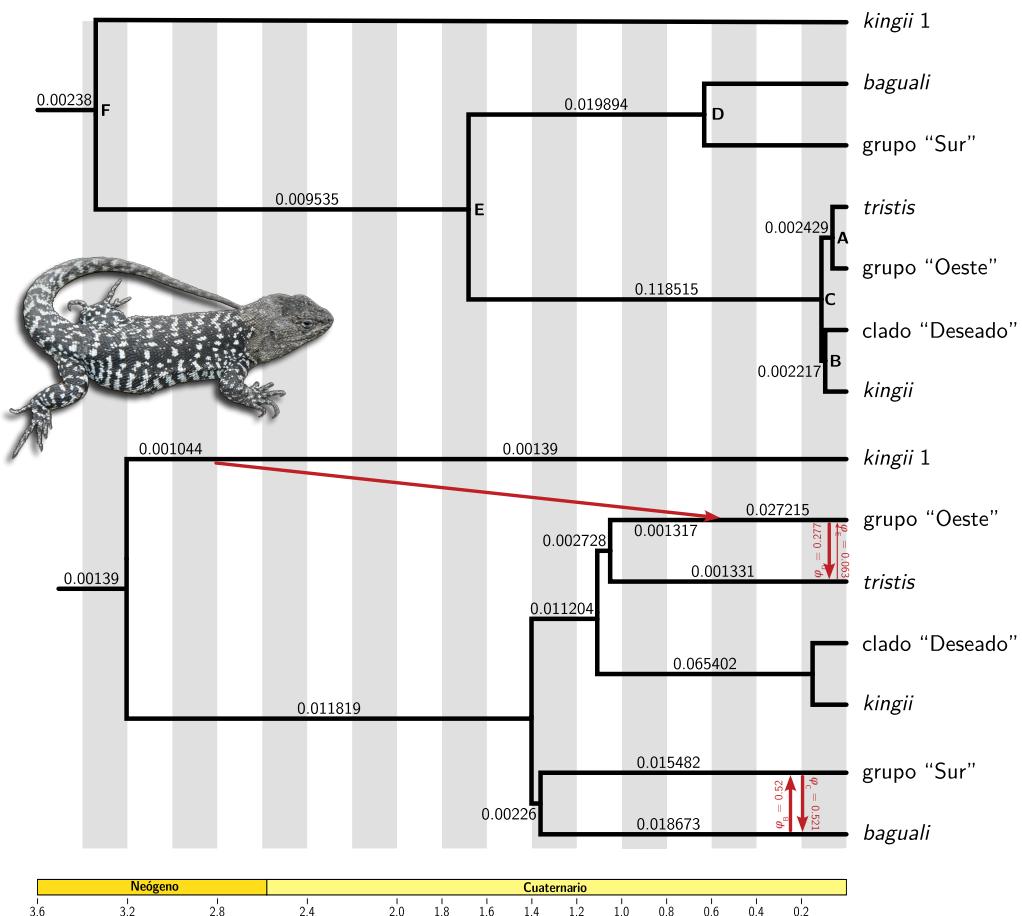


Modelo MCEM + m

- IMA (Hey 2018)
- G-PHoCS (Gronau et al. 2011)
- 3s (Dalquen et al. 2017) → 3 spp.



Lagartijas del clado *Lioleamus kingii*



1. Filogenia bifurcada (MSC)
2. Ajuste de datos al MSC (pruebas predictivas *a posteriori*)
3. Flujo génico (métodos heurísticos: PHYLONETWORKS y PIPEMASTER)
4. Parámetros del MCEM + i (*full-likelihood*: BPP)
5. Calibración

Sánchez et al. (2023)

Conclusiones

Combinar múltiples enfoques:

- Métodos heurísticos son + rápidos → **exploración**
- Métodos paramétricos + robustos (uso directo de secuencias) → **parámetros**

Evaluar adecuación a modelos con/sin flujo

Referencias

- Allman ES et al. 2019. DOI: 10.1186/s13015-019-0159-2
- Blischak PD et al. 2018. DOI: 10.1093/sysbio/syy023
- Dalquen DA et al. 2017. DOI: 10.1093/sysbio/syw063
- Del-Río G et al. 2021. DOI: 10.1111/evo.14377
- Edelman NB y Mallet J. 2021. DOI: 10.1146/annurev-genet-021821-020805
- Flouri T et al. 2020. DOI: 10.1093/molbev/msz296
- Green RE et al. 2010. DOI: 10.1126/science.1188021
- Gronau I et al. 2011. DOI: 10.1038/ng.937
- Hey J et al. 2018. DOI: 10.1093/molbev/msy162
- Hibbins MS y Hahn M. 2021. DOI: 10.1093/genetics/iyab173
- Jiao J y Yang Z. 2021. DOI: 10.1093/sysbio/syaa052
- Kong et al. 2022. DOI: 10.1101/2022.11.14.516468
- Mavárez et al. 2006. DOI: 10.1038/nature04738
- Noll et al. 2021. DOI: 10.1038/s41598-022-07562-0
- Pease y Hahn 2015. DOI: 10.1093/sysbio/syv023
- Sánchez K.I. et al. 2021. DOI: 10.1016/j.ympev.2020.107046
- Sánchez K.I. et al. 2023. DOI: 10.1093/sysbio/syad019
- Seehausen 2004. DOI: 10.1016/j.tree.2004.01.003
- Solís-Lemus C., Ané C. 2016. DOI: 10.1371/journal.pgen.1005896
- Solís-Lemus C. et al. 2017. DOI: 10.1093/molbev/msx235
- Yu Y. et al. 2014. DOI: 10.1073/pnas.1407950111
- Yu y Nakhleh 2015. DOI: 10.1186/1471-2164-16-S10-S10
- Wen y Nakhleh 2016. DOI: Wen y Nakhleh 2016. PLOS Genetics
- Wen y Nakhleh 2018a. DOI: 10.1093/sysbio/syx085
- Wen y Nakhleh 2018b. DOI: 10.1093/sysbio/syy015
- Zhang et al. 2018. DOI: 10.1093/molbev/msx307
- Zhu y Nakhleh 2018. DOI: 10.1093/bioinformatics/bty295

