

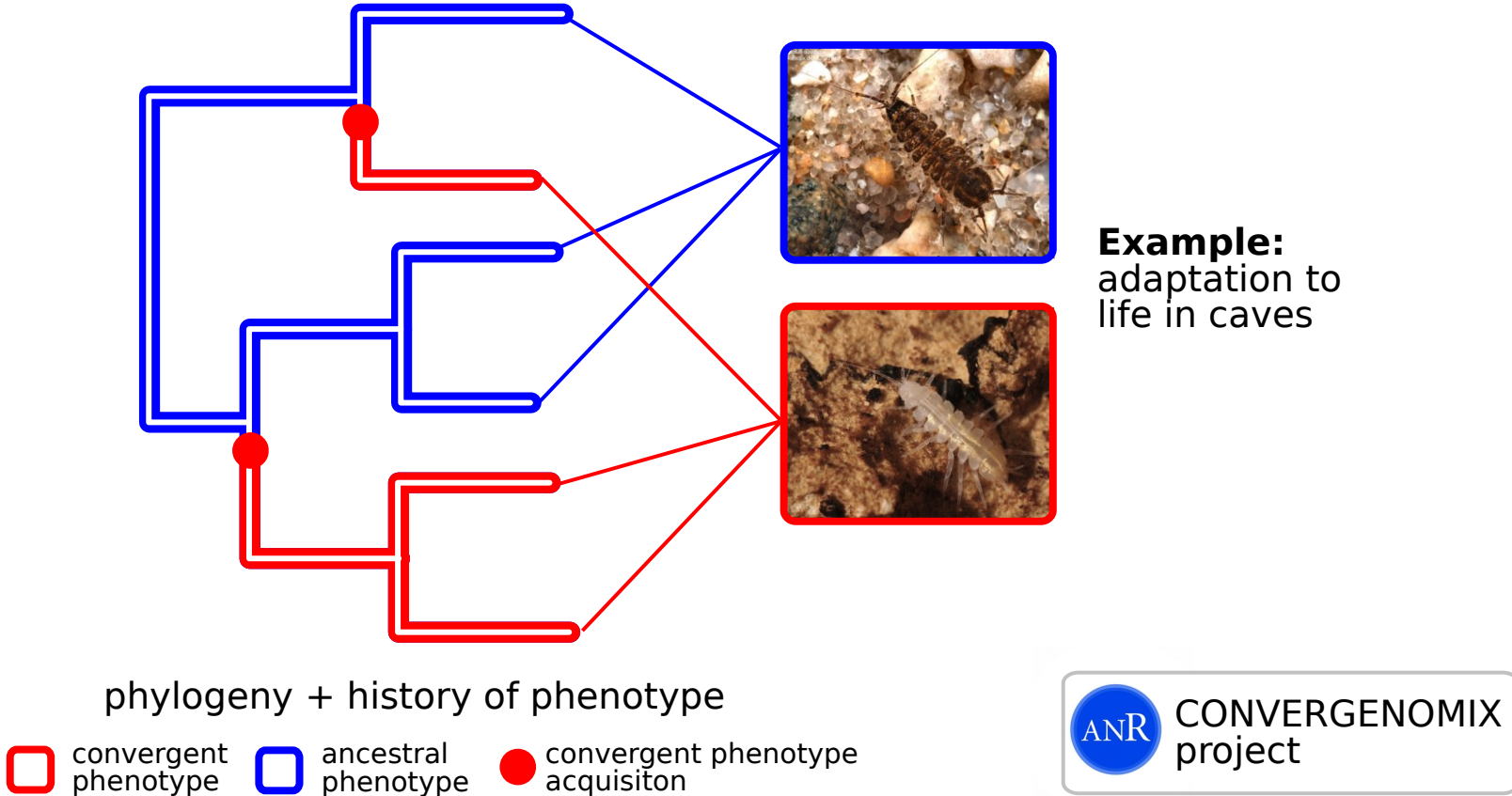
Detecting convergent substitutions with sparse codon-based differential-selection models

Vincent Lanore, Bastien Boussau, Philippe Veber, Nicolas Lartillot

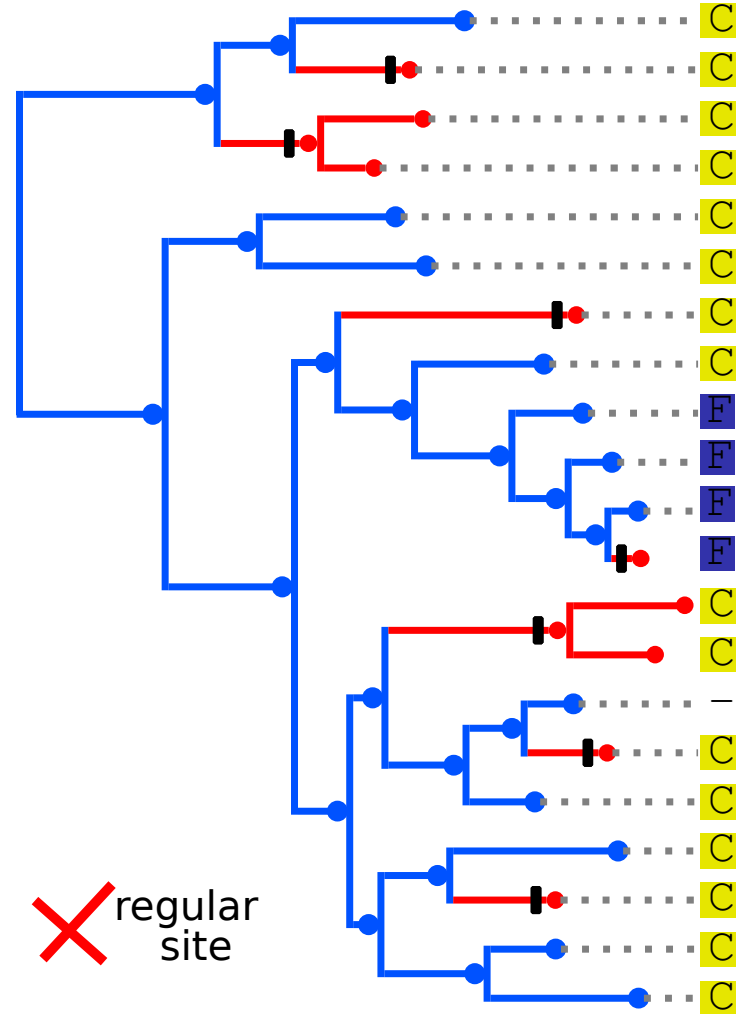
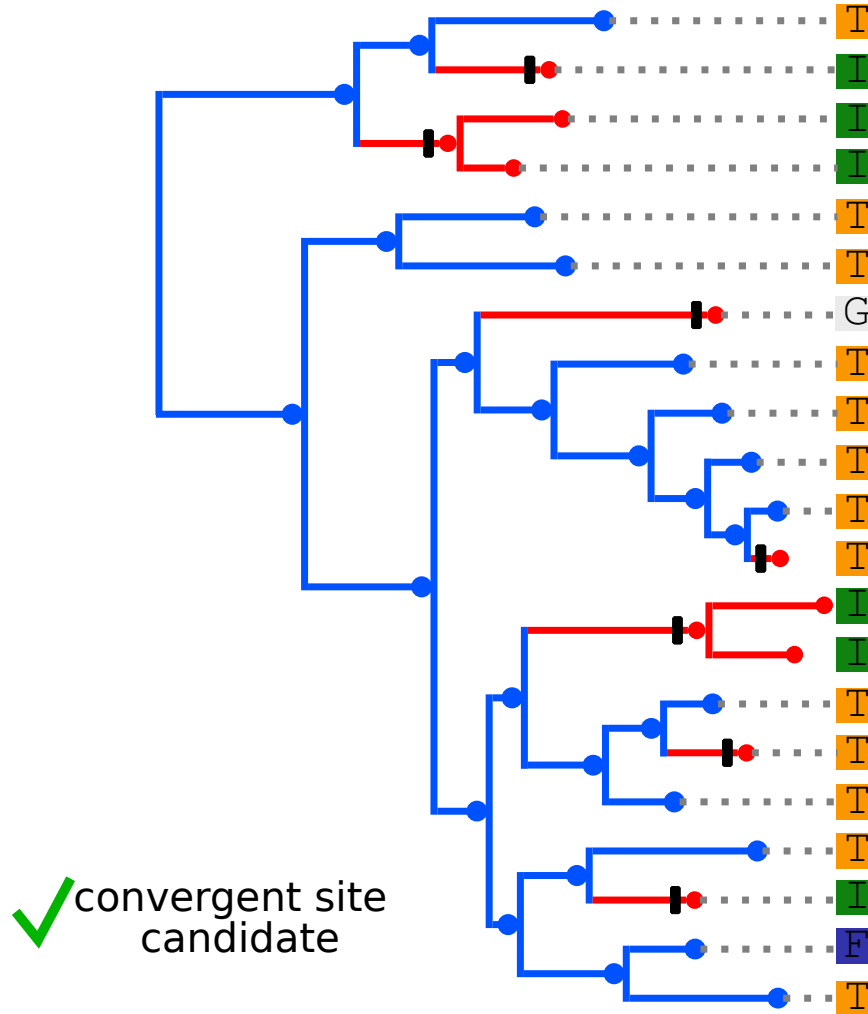


Phenotypic convergence

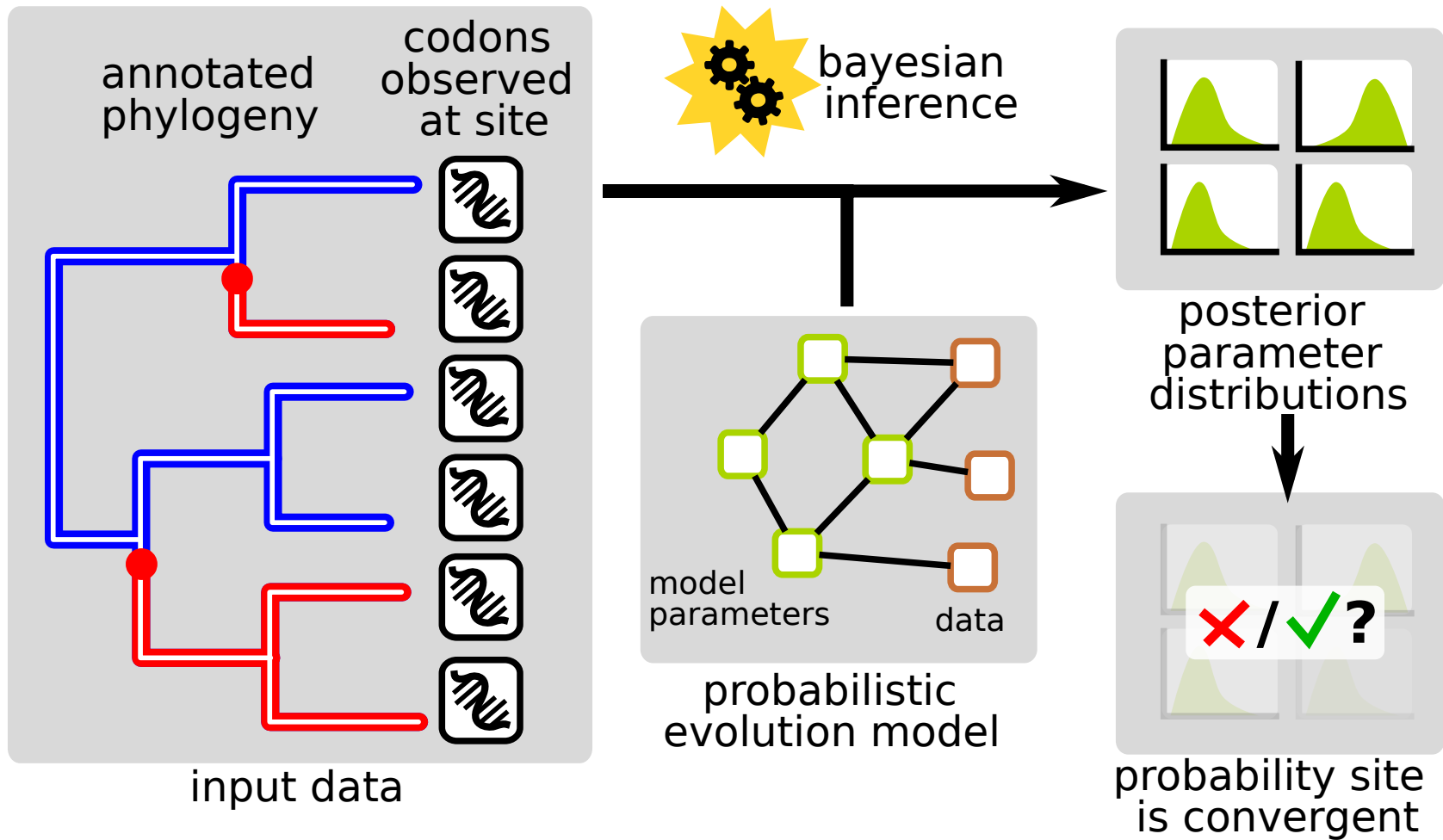
Phenotypic convergence:
independent acquisitions of similar traits



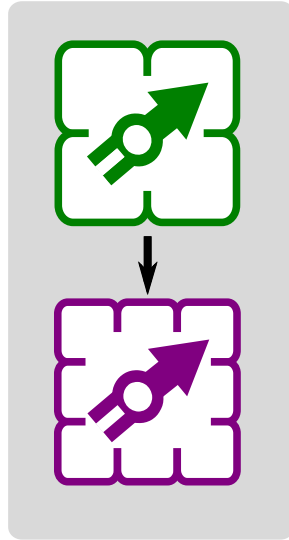
Goal: detect “convergent sites”



Approach



Mutation-selection models



mutation

First level: mutation

from a nucleotide substitution matrix
compute a codon substitution matrix



fitness



substitution
rate



sequence



codon

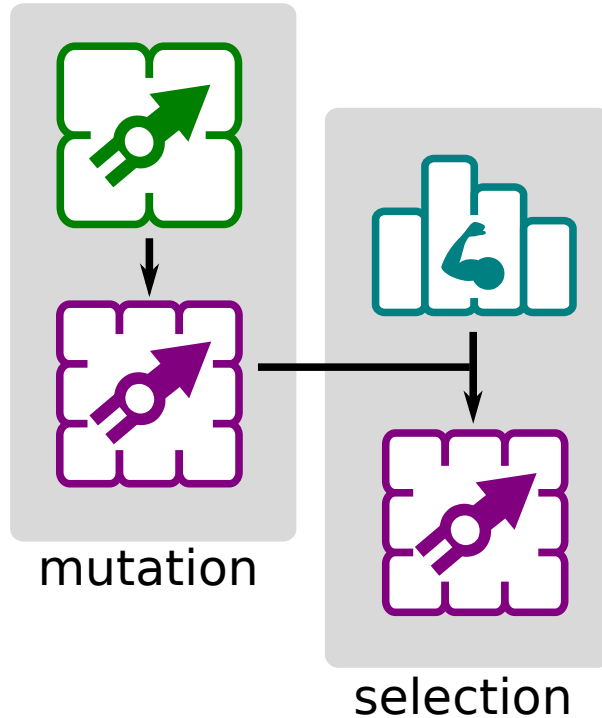


nucleotide



amino
acid

Mutation-selection models



Second level: selection

fixation probability depends on fitness of amino acid before and after substitution



fitness



substitution
rate



sequence



codon

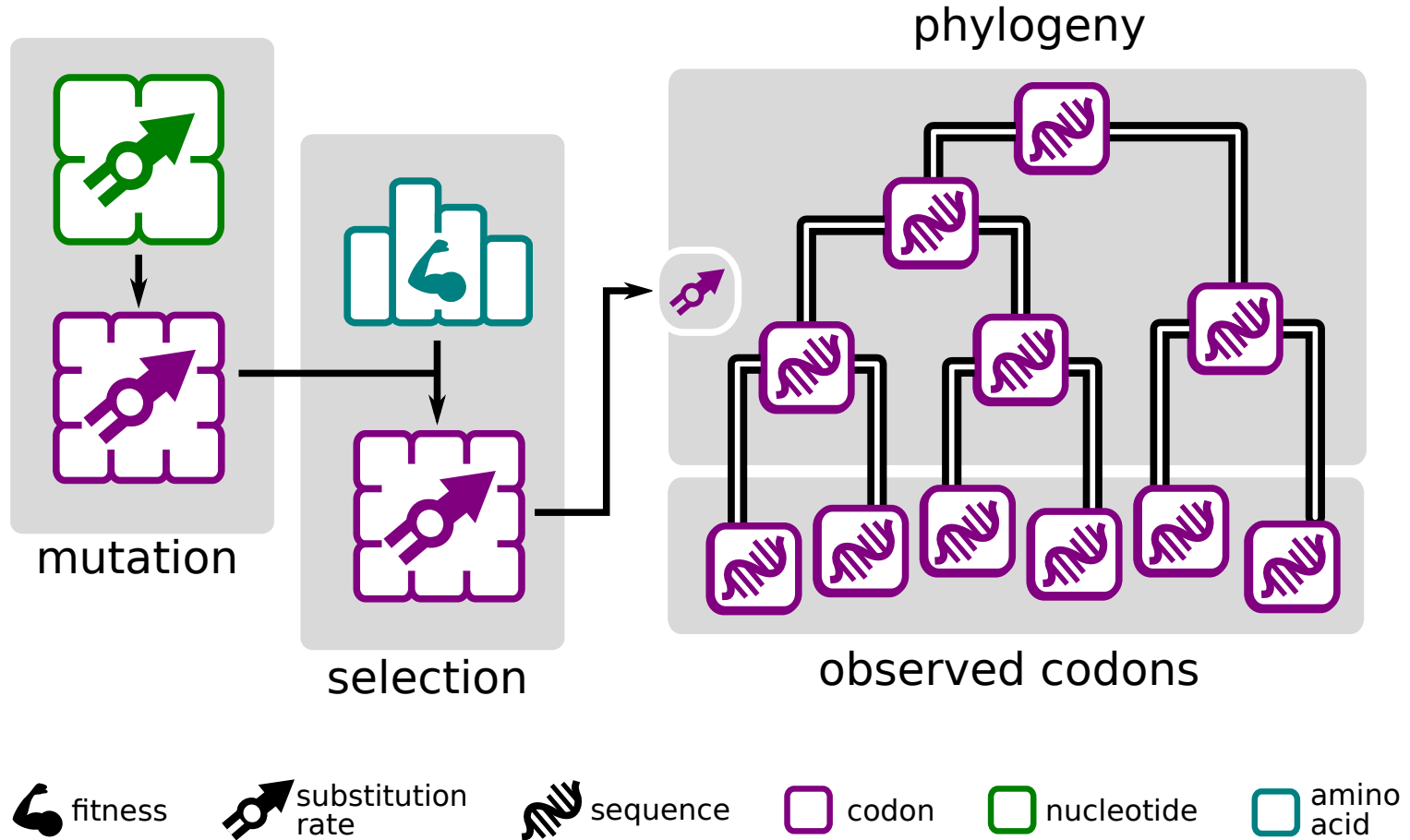


nucleotide

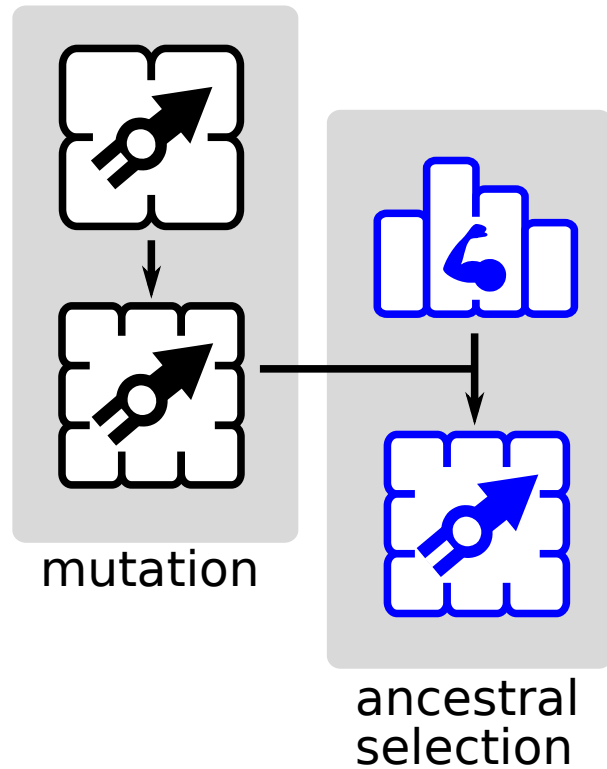


amino
acid

Mutation-selection models



Diffsel: differential-selection model



Mutation-selection base
for ancestral phenotype



fitness



substitution
rate



sequence

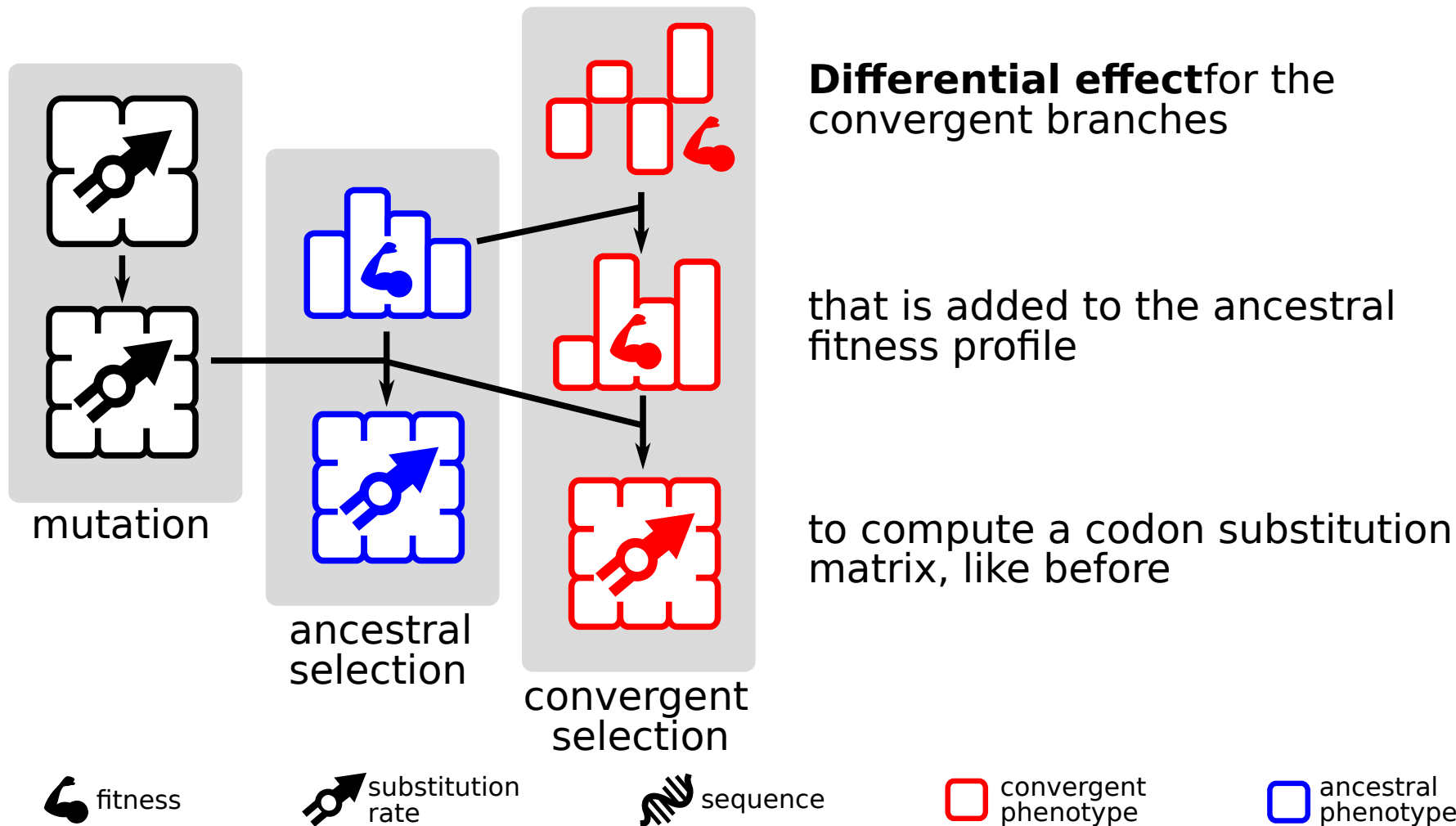


convergent
phenotype

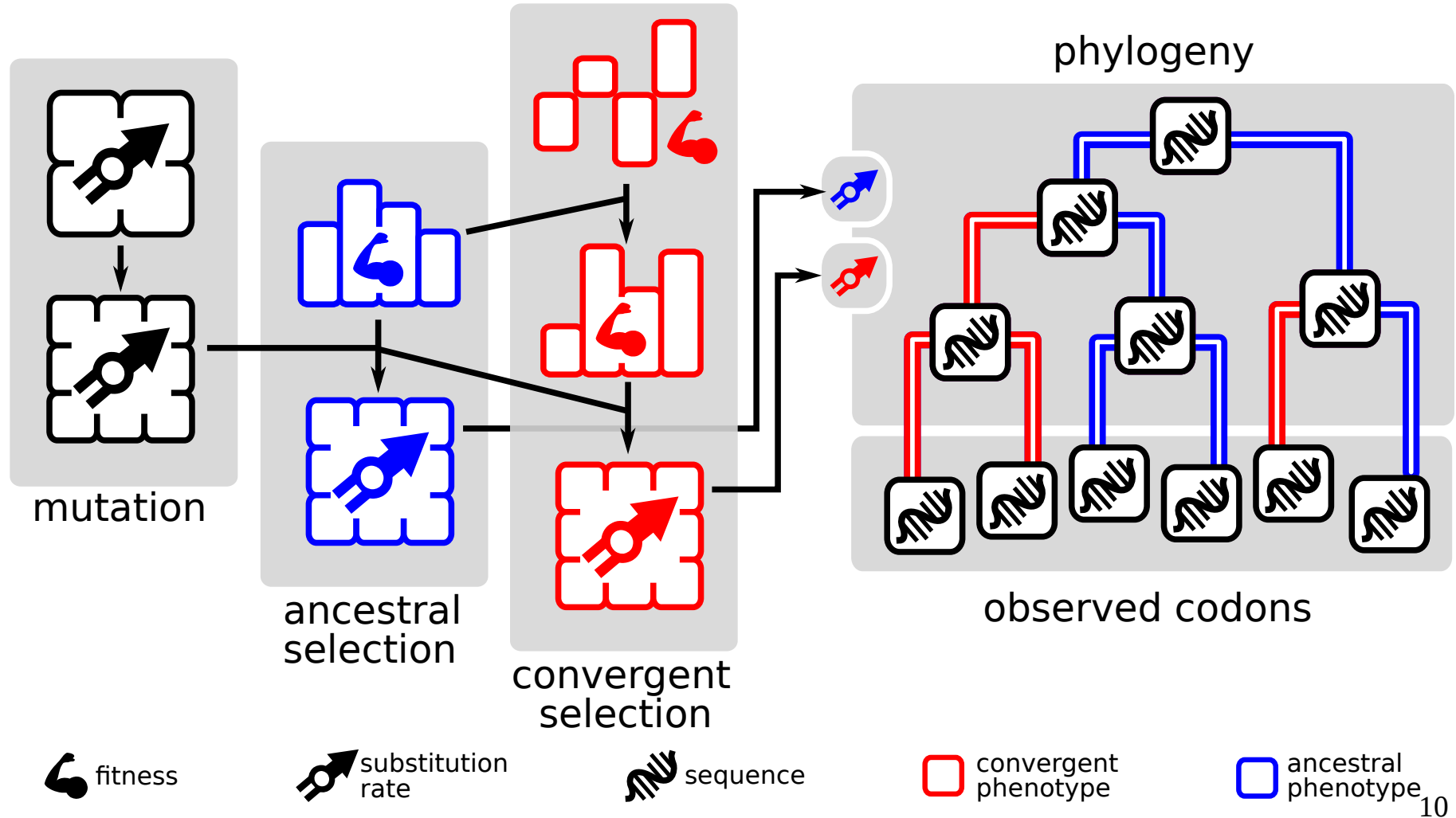


ancestral
phenotype

Diffsel: differential-selection model



Diffsel: differential-selection model



Diffsel: results and limitations

original publication with results on real data



Parto, Lartillot 2018.

"Molecular adaptation in Rubisco:
Discriminating between convergent evolution and positive
selection using mechanistic and classical codon models."
PloS one

comparison with other convergence
detection methods on simulated data



Rey et al. 2019.

"Detecting convergent adaptive amino
acid evolution." Submitted to Philosophical Transactions
of the Royal Society B

recall/precision results on simulated data
show diffsel is **good at ranking sites**

Diffsel: results and limitations

original publication with results on real data



Parto, Lartillot 2018.

"Molecular adaptation in Rubisco:
Discriminating between convergent evolution and positive
selection using mechanistic and classical codon models."
PloS one

comparison with other convergence
detection methods on simulated data

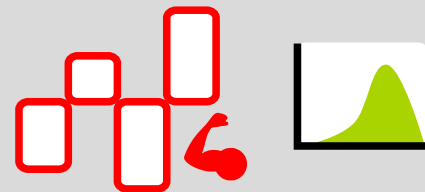


Rey et al. 2019.

"Detecting convergent adaptive amino
acid evolution." Submitted to Philosophical Transactions
of the Royal Society B

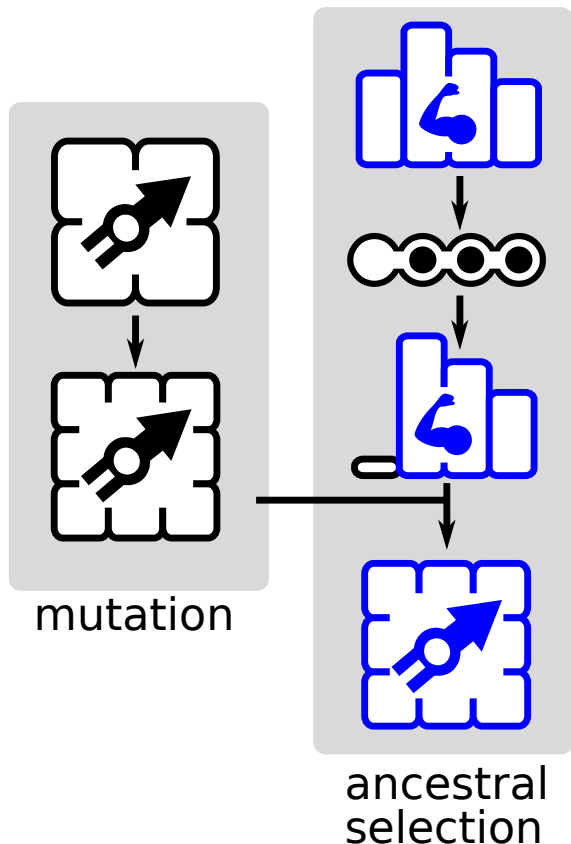
recall/precision results on simulated data
show diffsel is **good at ranking sites**

Problem: scoring sites based on
posterior probability of differential
effects is difficult



difficult to calibrate for detection
on real data

Sparse differential selection



Only certain amino acids
have significant fitness

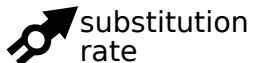
estimating a **profile mask**
(boolean array)

amino acids not in the mask
have a **fixed fitness ϵ**

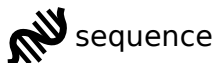
this makes the model **sparse**
(only certain fitnesses are estimated)



fitness



substitution
rate



sequence

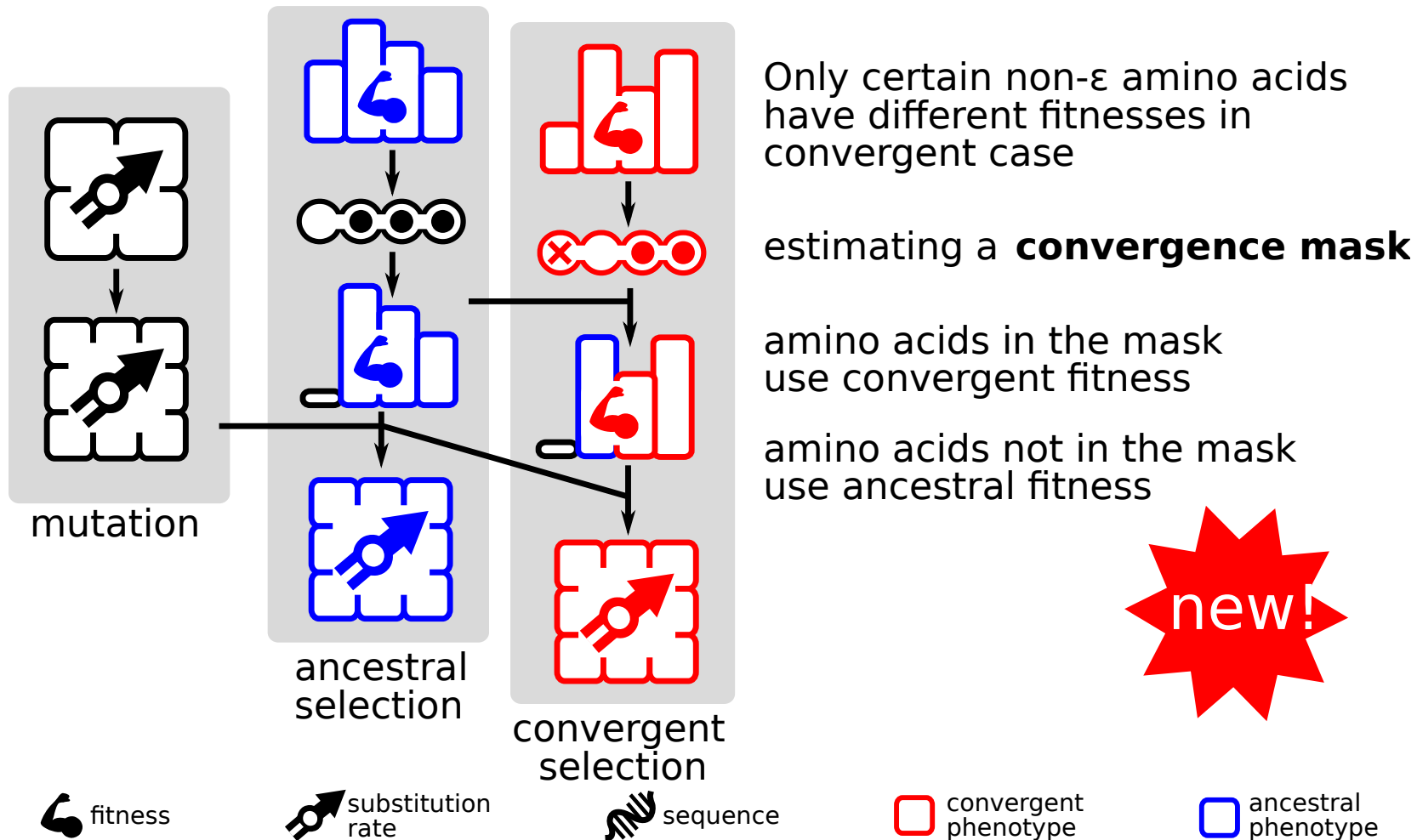


convergent
phenotype

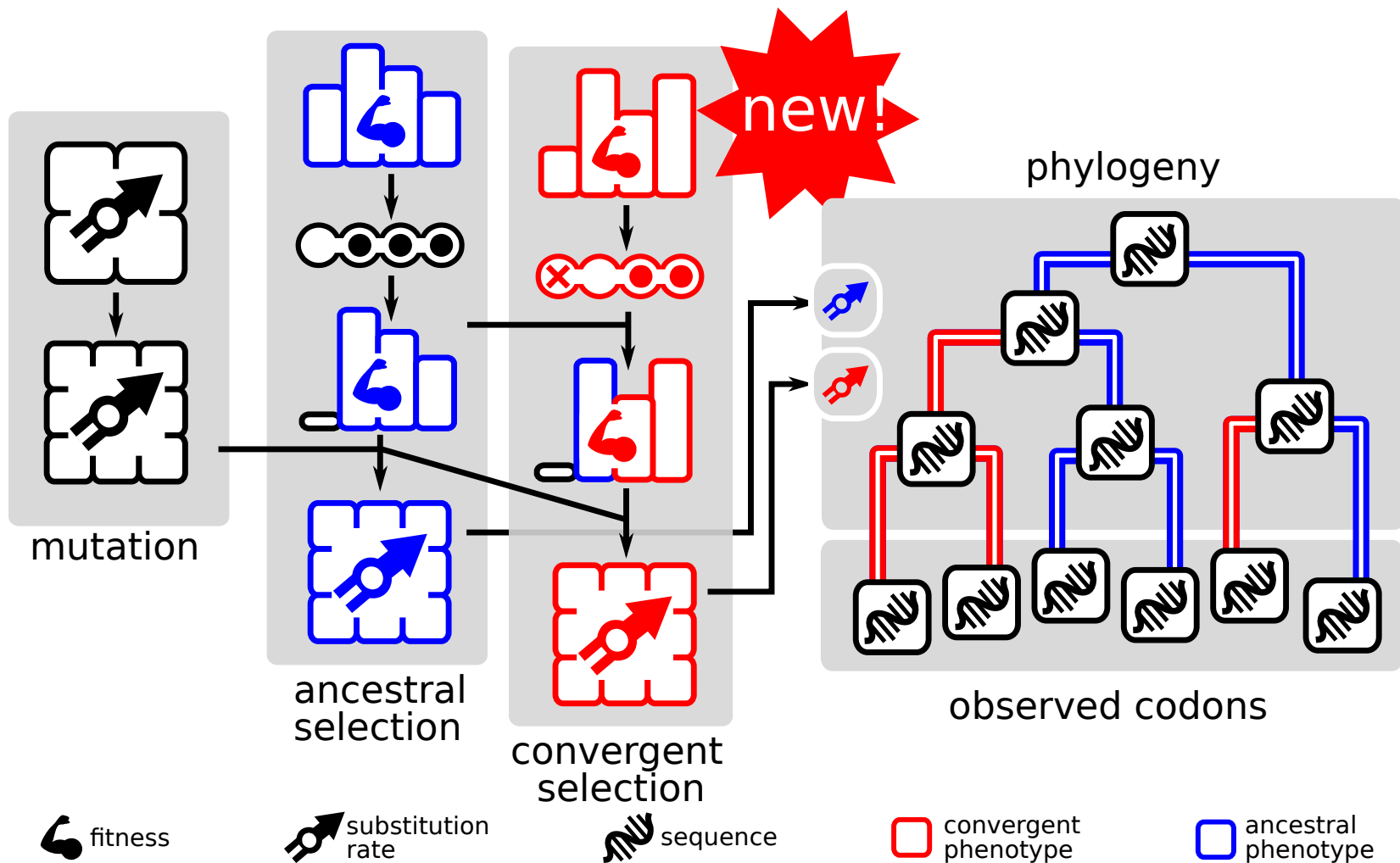


ancestral
phenotype

Sparse differential selection

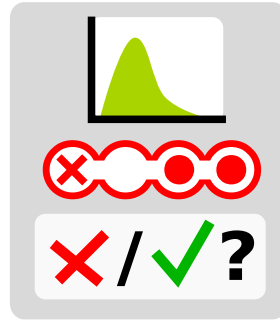


Sparse differential selection



Sparse model benefits

Convergence probability easy to establish:
posterior probability of convergence mask



Also, possibly more efficient because not all amino acid fitnesses need to be evaluated

Results on real data

data taken from



Besnard et al. 2009. “Phylogenomics of C4 Photosynthesis in Sedges (Cyperaceae): Multiple Appearances and Genetic Convergence.” Molecular biology and evolution. 26. 1909-19.10.1093/molbev/msp103.

convergent acquisition of C4
photosynthesis in Cyperaceae
(single gene, 79 species, 458 sites)

Site	Diffsel score/aa	Diffsel sparse score/aa	Identified in Besnard et al
34		0.93 (Q)	
52	0.969 (L)	0.94 (F), 0.93 (L)	yes
57	0.968 (D)		
88	0.984 (T)	0.95 (P)	yes
120	0.94 (Q)	0.99 (E), 0.93 (A)	yes
121	0.96 (N)		yes
126		0.99 (A)	
127	0.955 (T)		
171		0.98 (E)	yes
172		0.91 (A)	
173	0.901 (A)		
175	0.935 (I)	0.91 (V)	
213	0.969 (N)	0.96 (H)	yes
279	0.964 (V)	0.99 (I)	yes
292		0.92 (L)	
299	0.957 (F)		yes
309	0.94 (A)	0.91 (S)	yes
327		0.98 (A)	
328	0.993 (S)		yes
387		0.97 (G)	yes
400	0.922 (E)		
417		0.94 (N)	
427	0.915 (E)		
428	0.934 (I)		
453		0.96 (A)	

Conclusions

Presented a **sparse** differential selection model to detect convergent substitutions

- convergence probability easier to establish
- **real data results:** sparse model consistent with existing publications with some differences

Ongoing work

- evaluate capability to predict false positive rate and number of convergent sites on simulations
- multigene version

Acknowledgements

This work

- Philippe Veber
- Bastien Boussau
- Nicolas Lartillot

Convergence review paper

- Carine Rey
- Philippe Veber
- Laurent Gueguen
- Nicolas Lartillot
- Marie Semon
- Bastien Boussau

Funded by ANR
Convergenomix



Acknowledgements

This work

- Philippe Veber
- Bastien Boussau
- Nicolas Lartillot

Convergence review paper

- Carine Rey
- Philippe Veber
- Laurent Gueguen
- Nicolas Lartillot
- Marie Semon
- Bastien Boussau

Funded by ANR
Convergenomix



Thank you for your attention