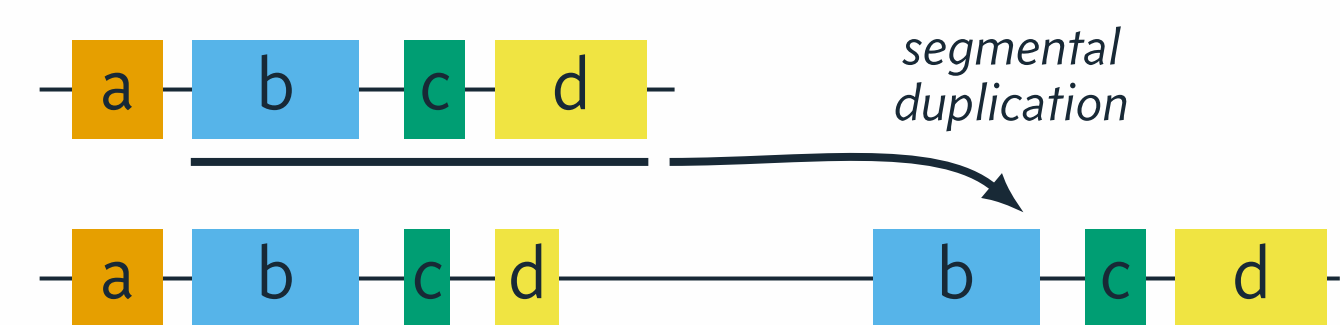


STUDYING THE EVOLUTION OF CRISPR-CAS SYSTEMS USING SUPERDTL RECONCILIATION

Yoann Anselmetti¹ Mattéo Delabre² Nadia El-Mabrouk² | ¹Université de Sherbrooke ²Université de Montréal

SYNTENIES AND SEGMENTAL EVENTS

- Synteny** are groups of genes evolving jointly through segmental events



CRISPR-CAS SYSTEMS

- Adaptive immunity** mechanisms found in bacteria and archaea



- Cas genes** perform three main functions

Adaptation Insertion of new *target sequences* into the CRISPR array
Expression Transcription of the array as crRNA
Interference Degradation of targets, guided by crRNA

CAS SYNTENIES CLASSIFICATION

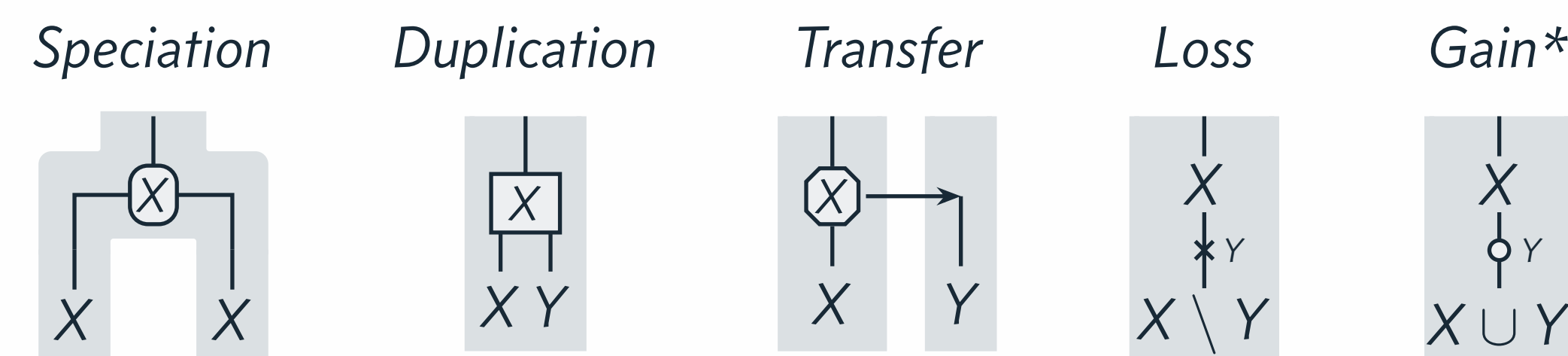
	Adaptation			Expression			Interference				
Class 1	Type I	Cas1	Cas2	Cas4	Cas6	Cas7	Cas5	Cas11	Cas8	Cas3''	Cas3'
	Type III	Cas1	Cas2		Cas6	Cas7	Cas5	Cas11		Cas10	
	Type IV				Cas6	Cas7	Cas5			Csf1	
Class 2	Type II	Cas1	Cas2	Cas4	RNasIII					Cas9	
	Type V	Cas1	Cas2	Cas4						Cas12	
	Type VI	Cas1	Cas2							Cas13	

Figure adapted from [3]

CAN WE INFER A SCENARIO FOR THE EVOLUTIONARY HISTORY OF CAS SYNTENIES?

SYNTENY EVOLUTION MODEL

- Synteny** are modelled as **sets of genes**, i.e., gene order is ignored
- Undergo **unary** (content-changing) and **binary** (copy-creating) **events**



*Gains are non-segmental & each family is gained exactly once

KEY: Species tree Synteny tree

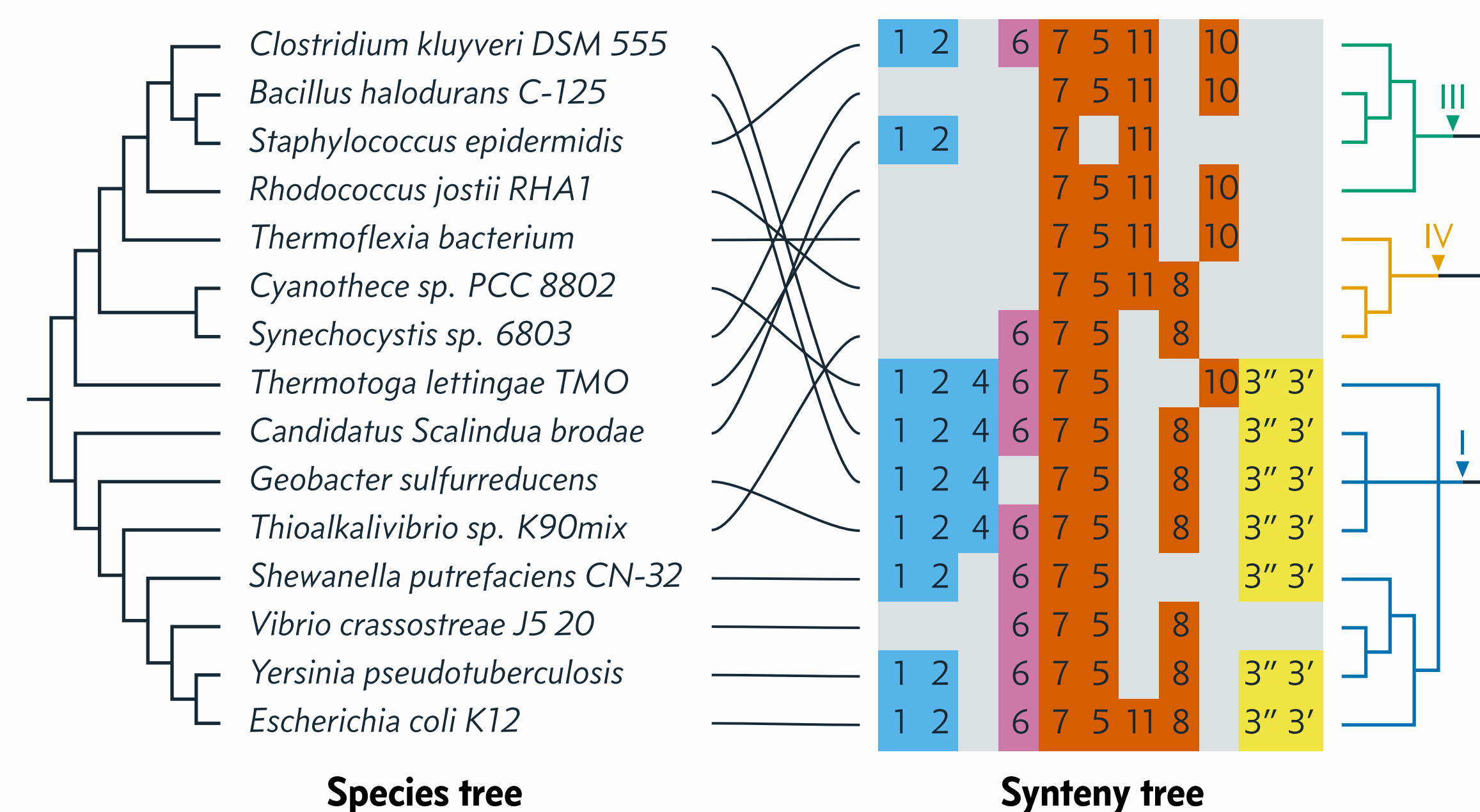
- Evolutionary histories** are branching sequences of events that correspond to a *synteny tree* and fit inside a *species tree*
- Each event type is given a cost, histories of **minimum cost** are sought

SUPERDTL ALGORITHM

- Given *synteny* and *species trees*, computes a minimum-cost history
- Extension of usual DTL reconciliation algorithms [4] with inference of **ancestral synteny and unary events**
- Time complexity $O(NM^3)$, space complexity $O(NM)$
 N : Number of synteny, M : Number of species
- Implementation available** at: github.com/UdeM-LBIT/superrec2

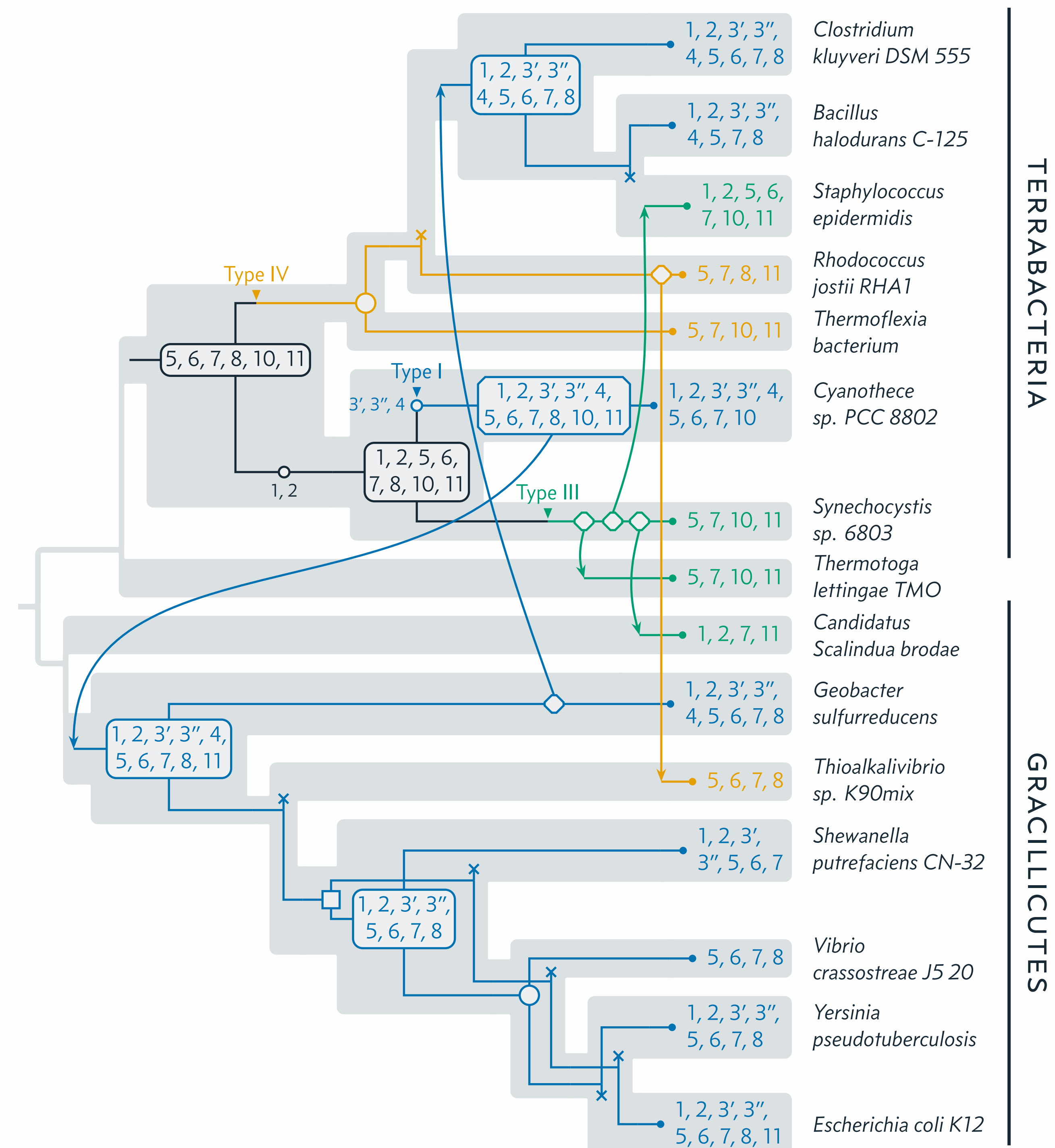
DATASET

- 15 Class-1 Cas synteny** compiled from [3], species tree based on [1]



SCENARIO FOR CLASS-1 CAS SYNTENIES

- Event costs:** Duplication $\in \{1, \dots, 3\}$, Transfer $\in \{1, \dots, 4\}$, Loss 1
- Representative of the 32 solutions for Duplication $\in \{1, \dots, 2.5\}$, Transfer 4 (with non-significant variations) is depicted below



- In line with current hypotheses** for Class-1 Cas synteny evolution [2]
- Consistent resolution** of the multifurcations in the synteny tree
- Multiple losses seen in *Type I* may be due to **unsampled lineages** [5]

[1] Coleman et al. "A rooted phylogeny resolves early bacterial evolution" (May 2021)

[2] Koonin and Makarova. "Evolutionary plasticity and functional versatility of CRISPR systems" (Jan. 2022)

[3] Makarova et al. "Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants" (Feb. 2020)

[4] Tofigh, Hallett, and Lagergren. "Simultaneous identification of duplications and lateral gene transfers" (Mar. 2011)

[5] Weiner and Bansal. "Improved duplication-transfer-loss reconciliation with extinct and unsampled lineages" (Aug. 2021)