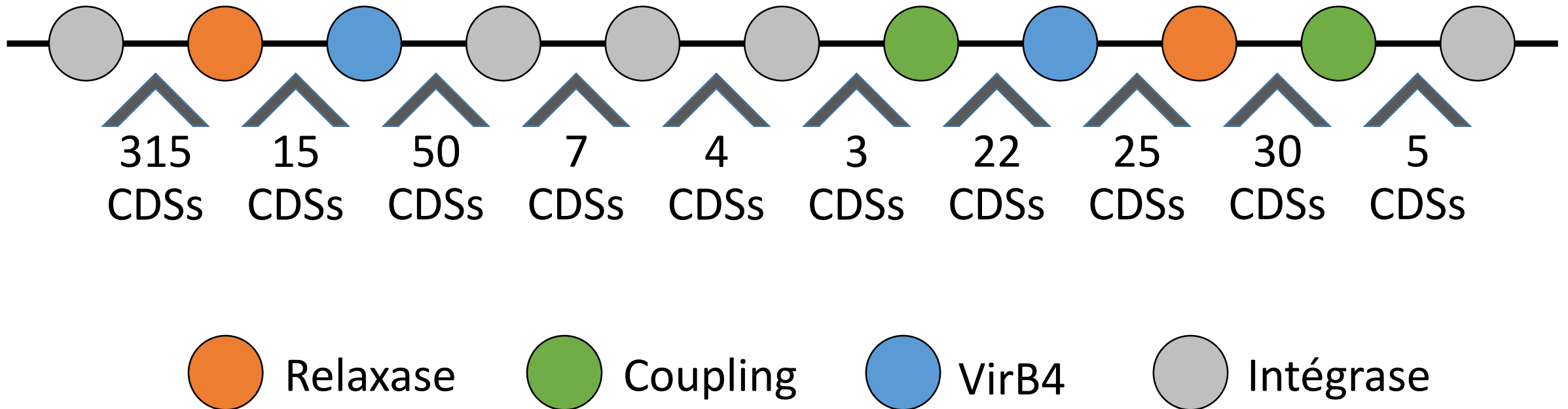


algorithme de détection des structures ICEs/IMEs

- Basé sur l'extension de graine (comme blast) et fusion de graines compatibles.
- Implémentation orientée objet pour faciliter la structuration des données.

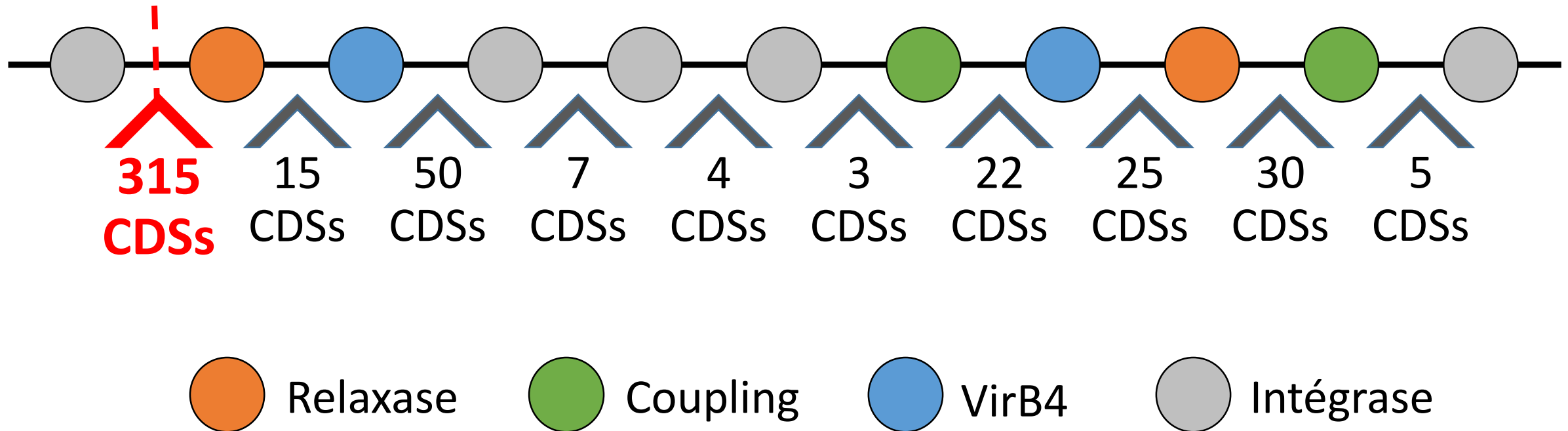
Données en entrées

- Séquence de protéines étiquettes ordonnées sur le génome.



1^{ère} étape : les ICEs/IMEs ne peuvent pas être trop grands

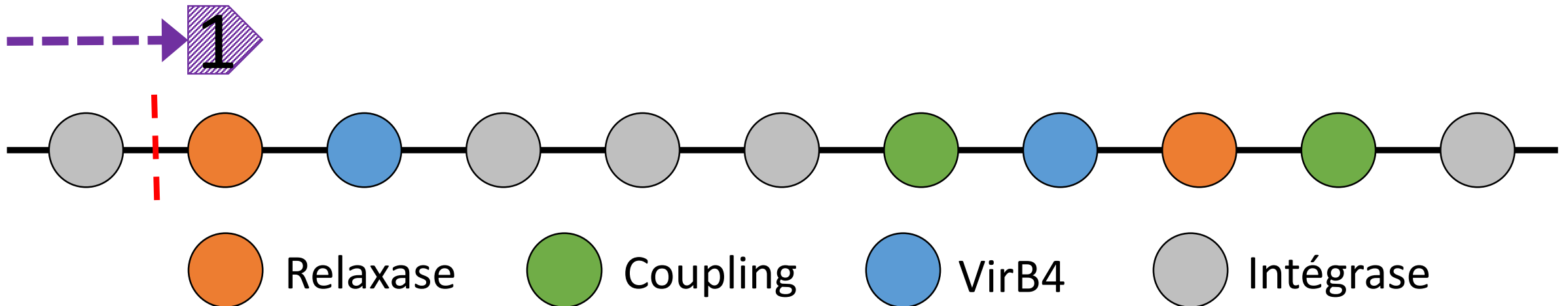
- La séquence est coupée si > 100 CDSs entre 2 protéines étiquettes.



2^{ème} étape : règles pour créer une graine

La séquence est parcourue de gauche à droite (—→). Si une des 3 protéines étiquettes relaxase, coupling, ou virB4 est rencontrée → début graine (1).

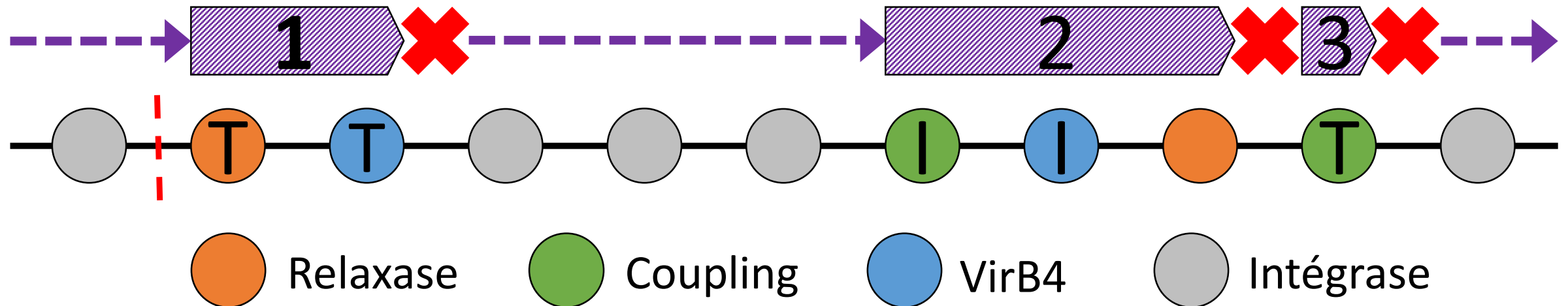
- Si l'élément contient une virB4, cela indique un ICE (complet ou partiel).
- Si l'élément contient une coupling ou relaxase avec au moins une autre protéine étiquette, cela indique un ICE ou IME.
- Si on trouve une intégrase → moins spécifique des ICEs / IMEs (peut être autres EM, i.e. transposons). L'intégrase est toujours en bordure d'élément.



3^{ème} étape : règles pour étendre une graine

La séquence continue d'être parcourue de gauche à droite. Une graine ICE/IME ne peut pas contenir (conditions d'arrêt de l'extension) :

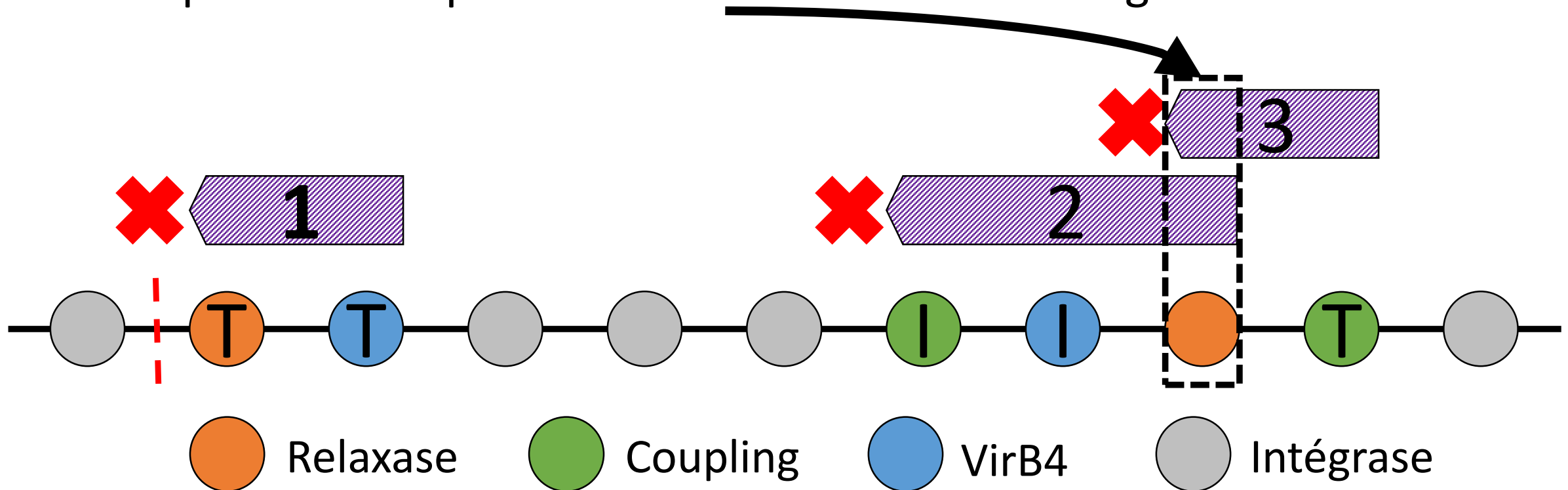
- 2 protéines étiquettes séparé de plus de 100 CDSs (étape 1).
- 2 protéines étiquettes de même type (sauf si elles sont adjacentes sur le génome et de type relaxase ou coupling).
- Protéines étiquettes de familles différentes (i.e. **I** = ICESt3, **T** = Tn916).
- Intégrase : les intégrase sont toujours en bordure d'élément, elles seront traités postérieurement.



4^{ème} étape : étendre graines de droite à gauche

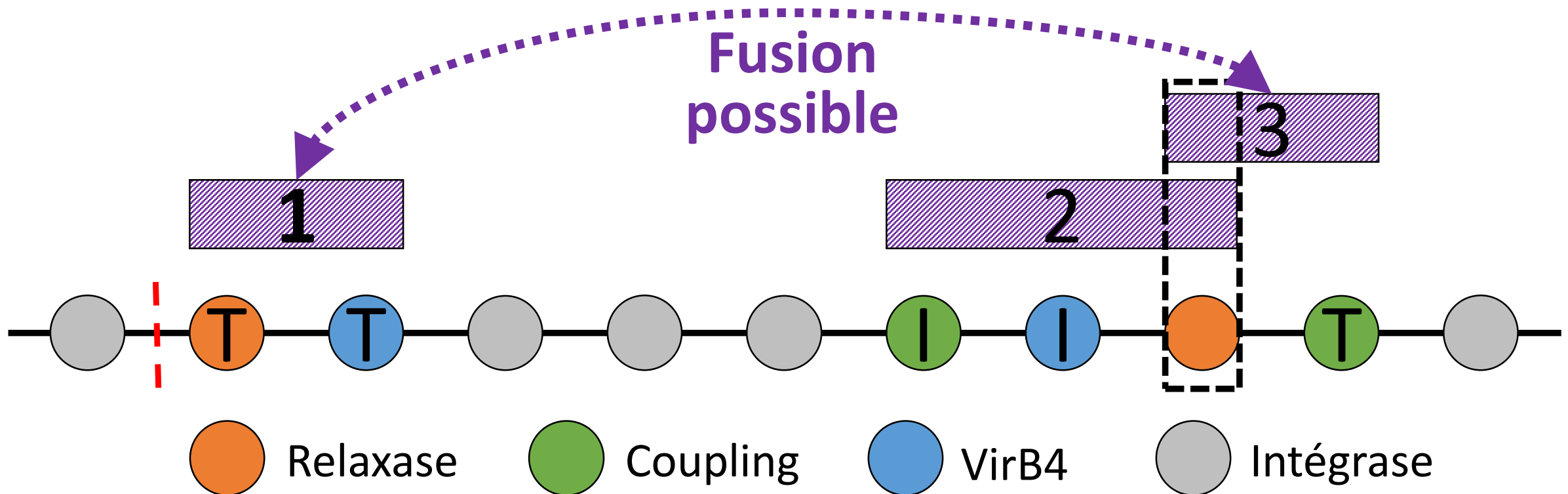
Après la 3^{ème} étape (création et extension des graines de gauche à droite), chaque graine est étendue de droite à gauche (même conditions d'arrêt).

- Les ICEs/IMEs n'ont pas de direction.
- L'algorithme est consistant et indépendant du choix du sens de lecture.
- Possible protéine étiquette en "conflit" rattachée à 2 graines différentes.



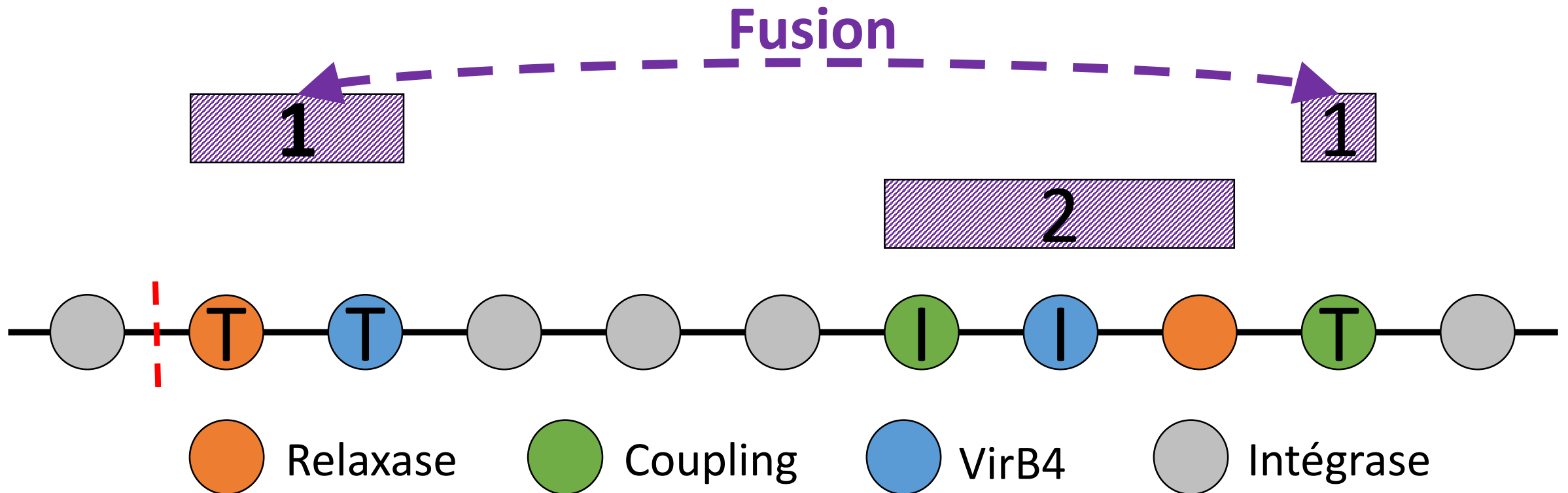
5^{ème} étape : fusion des graines

- Exhaustif : toutes les combinaisons de fusion sont testées. Priorité aux fusions des graines les plus proches si plusieurs possibilités.
- Récursif : détection des cas avec plusieurs niveaux d'emboitements et des cas ICEs/IMEs "découpés" en plus de 2 parties (cas rare).
- Les règles de fusion sont similaires aux règles pour étendre une graine.



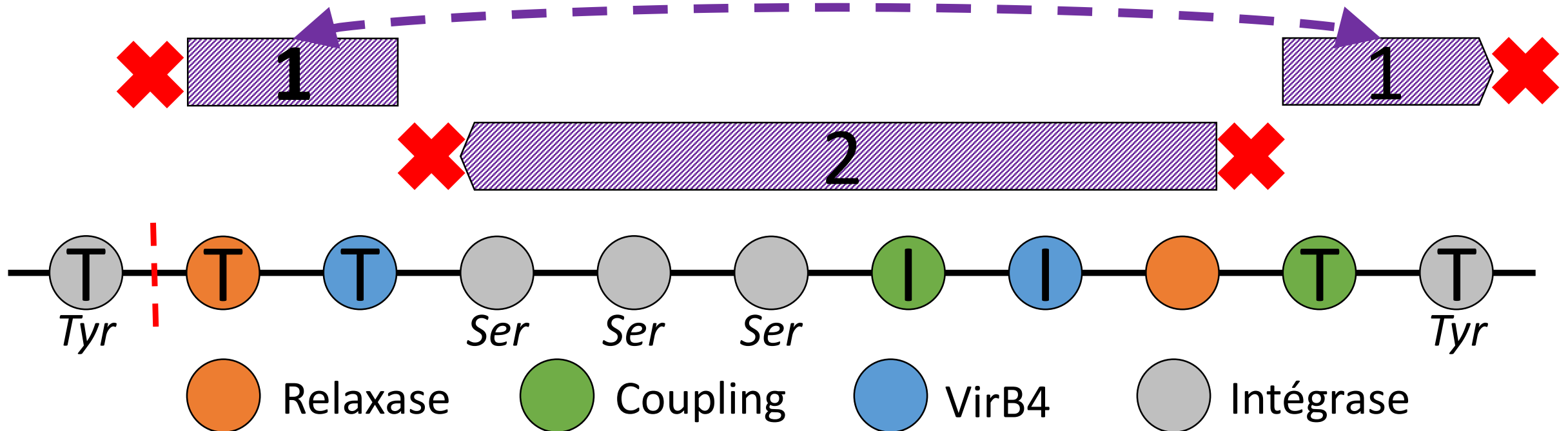
5^{ème} étape : fusion des graines

- Cette étape peut permettre de résoudre des protéines étiquettes en "conflit« (rattachées à 2 graines différentes).



6^{ème} étape : règles ajout intégrase à une graine

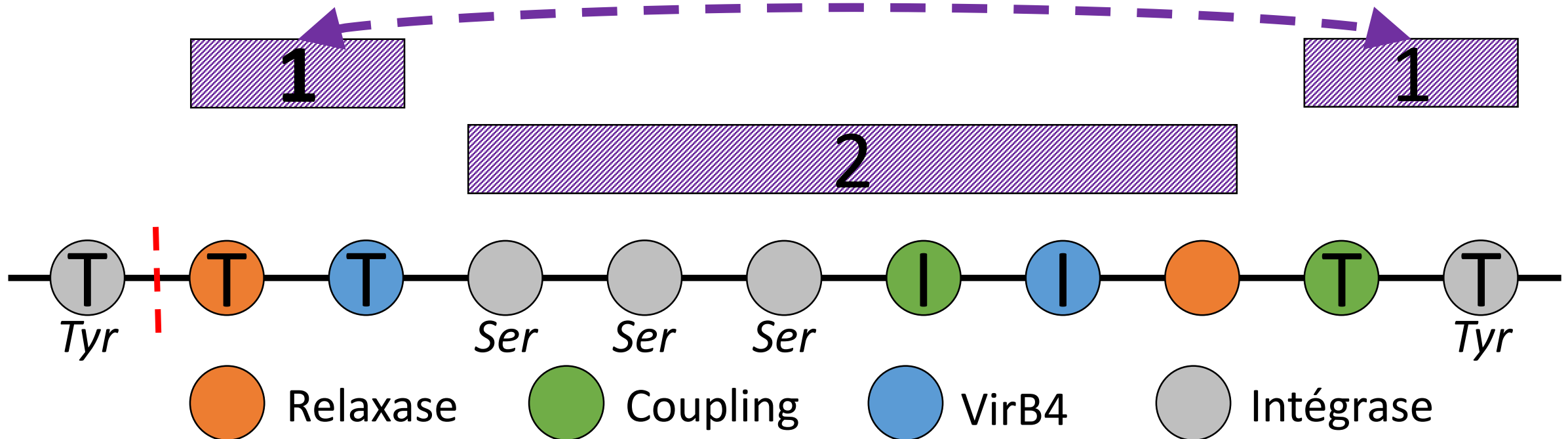
- Intégrase peut être en amont ou aval dans la limite 100 CDS (étape 1).
- La priorité est donnée aux (1) intégrases de même famille (i.e. I = ICESt3, T = Tn916) que les autres protéines étiquettes de la graine ou (2) aux intégrases adjacentes à la graine (possibilité d'intégrase éloignée si emboîtement).



6^{ème} étape : règles ajout intégrase à une graine

Cas spéciaux :

- Intégrases Ser adjacentes sur le génome.
- Si ICE (virB4), orientation intégrase : amont → brin -, aval → brin +.
- Il est possible que l'algorithme ne puisse pas choisir entre une intégrase amont ou aval.



7^{ème} étape : classification différents types ICEs/IMEs

Complet, partiel, à vérifié expérimentalement, emboîté, etc. :

- ICE complet : R+C+V+I
- Module de conjugaison : R+C+V
- ICE partiel : V + autres protéines étiquettes
- IME complet : R+I ou R+C+I avec distance < 10 CDS
- Élément mobilisable : R+C avec distance < 10 CDS
- Autre élément partiel : R+C>10 CDS, R+V, V+C

Jeux de tests de 89 ICEs/IMEs

Adaptés manuellement de cas réels pour tester l'algorithme sur une grande diversité de cas complexes :

- Protéines signatures : 356
- ICEs complets : 23
- Modules conjugaisons : 8
- ICEs partiels : 11
- IMEs complets : 37
- Éléments mobilisables ($R+C < 10$ CDS) : 3
- Autres éléments partiels ($R+C > 10$ CDS, $R+V$, $V+C$) : 7
- Éléments emboîtés : 47