

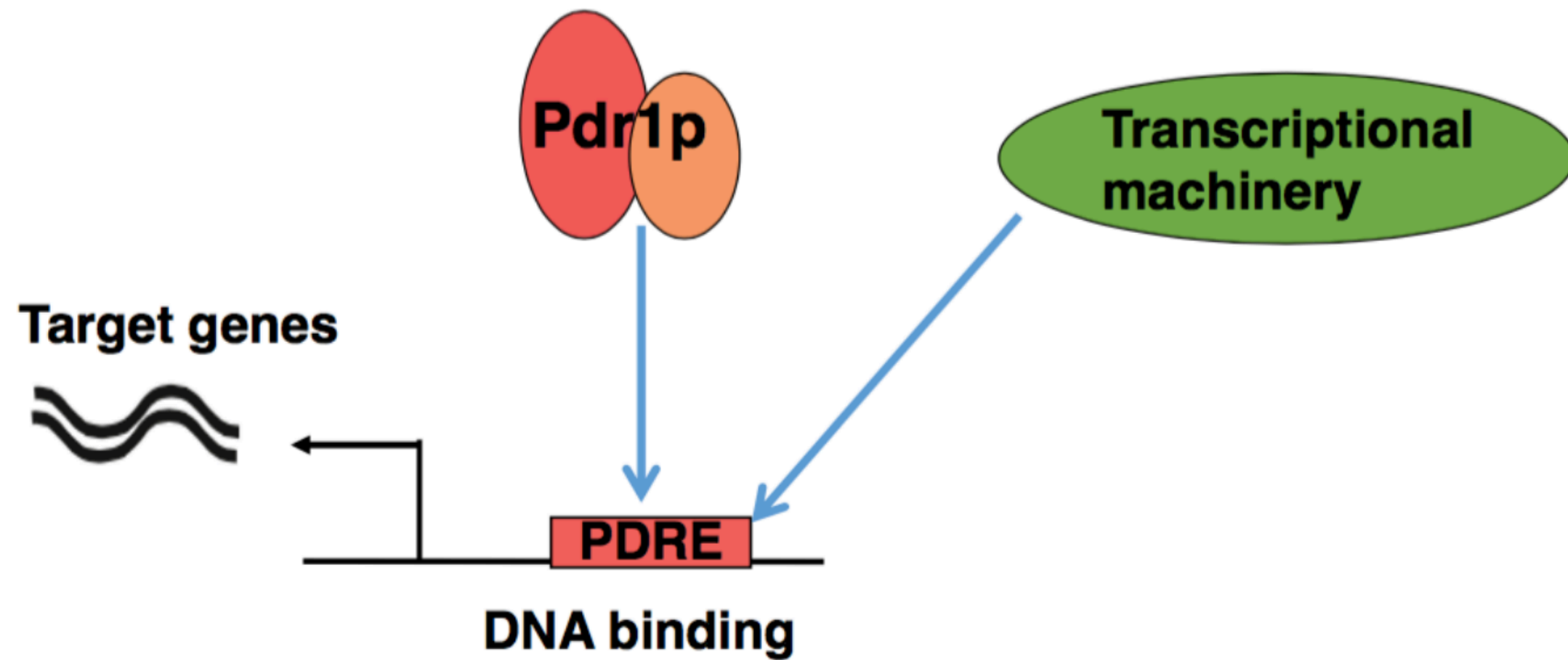


Identification des motifs d'ADN pour le facteur de transcription Pdr1 chez *Candida glabrata*

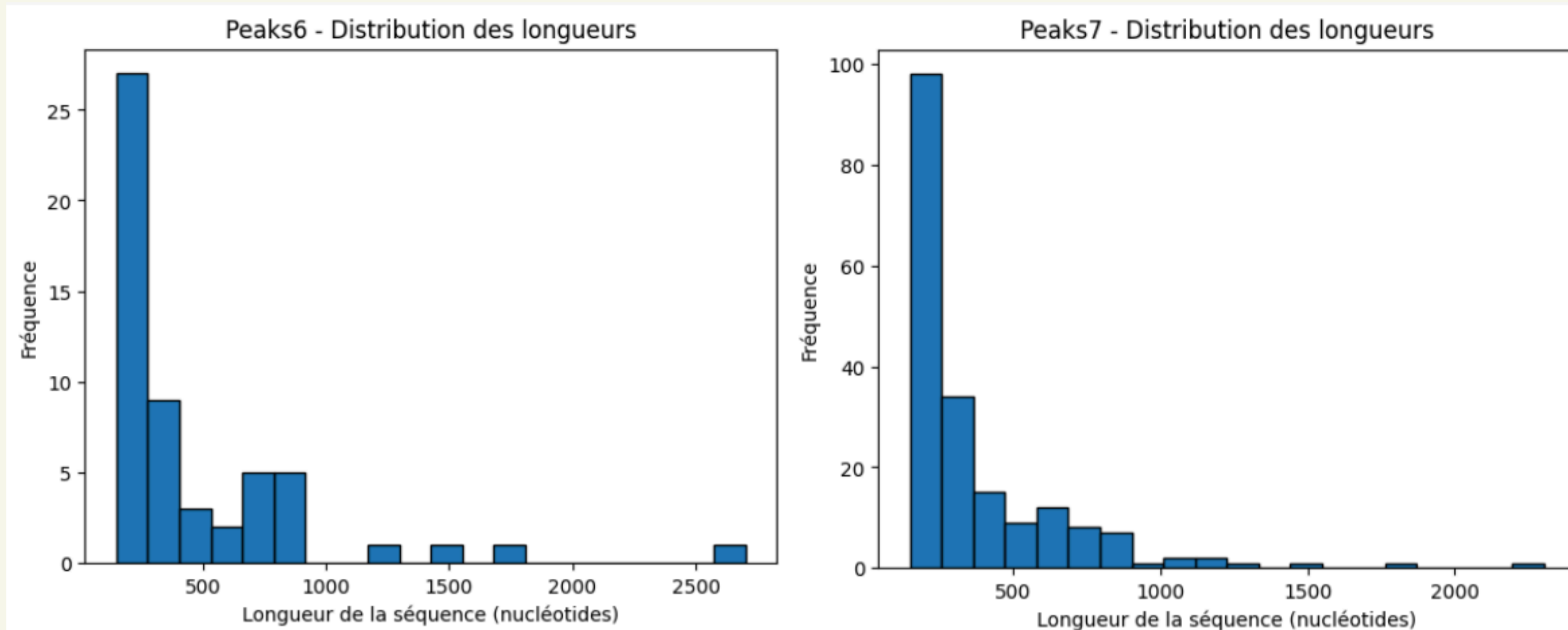
Irina Andriushchenko 21210340

Numéro des fichiers : Sequence_by_Peaks_6.fasta et
Sequence_by_Peaks_7.fasta

CONTEXTE BIOLOGIQUE



DONNÉES UTILISÉES



MÉTHODES UTILISÉES

HASH TABLE

Recherche des k-mers
les plus fréquent

MEDIAN STRING

Recherche de motif le
plus proche de toutes
les séquences

SUFFIX TREE

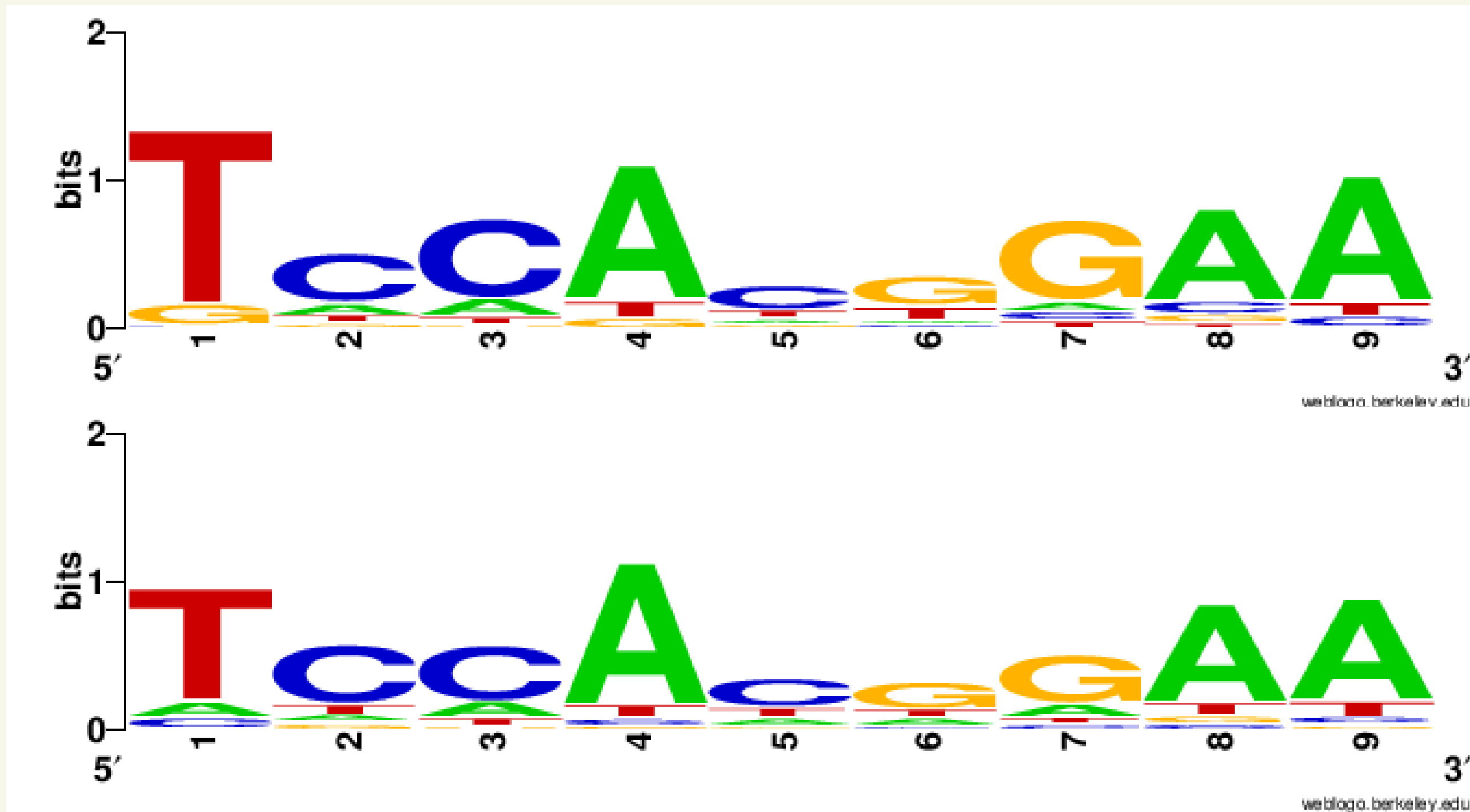
Recherche des plus
longues sous-
séquences communes

RÉSULTATS

Tableau comparatif des motifs identifiés pour Pdr1 (Peaks 6 et 7)

Méthode	Motif(s) identifié(s)	Présent dans fichier(s)	Remarques
Table de Hachage	TCCACGGAA	Peaks 6, Peaks 7	Motif spécifique retrouvé avec $k = 9$
	TTCCACGG, TCCACGGA, CCACGGAA	Peaks 6	Variantes proches observées à $k = 8$
Median String	TTATTTT, TTTATTT, TTATTTA	Peaks 6	Motifs riches en A/T, distances faibles mais peu informatifs
	ATTTTAT, TATTATT, TTATATT	Peaks 7	Motifs simples en A/T, aucune trace de TCCACGGAA
Suffix Tree	TCCACGGAA	Peaks 6, Peaks 7	Confirmé à $k = 9$ dans les deux fichiers
	TTCCACGG, TCCACGGA, CCACGGAA	Peaks 6	Variantes retrouvées à $k = 8$
RSAT PeakMotif	vvtCCACGGA, svttCCrCGGAA (IUPAC)	Peaks 6, Peaks 7	Cœur du motif confirmé : CCACGGAA

VISUALISATION





PROMOTEURS CONTENANT LE MOTIF

35 / 5300

0.66%

CONCLUSION

1. Motif TCCACGGAA
identifié et validé

2. Méthodes efficaces:
Hash Table + Suffix
Tree + RSAT



QUESTION TIME



MERCI