

# Minimisation du problème de caractérisation multiple

J-M. Chantrein  
sous la direction de  
F.Lardeux F.Saubion

LERIA

04 Juillet 2014

Master 2 Intelligence décisionnelle

- Certaines bactéries sont responsables de pathologies sur une large gamme de culture économiquement importante.

- Certaines bactéries sont responsables de pathologies sur une large gamme de culture économiquement importante.
- Ces pathologies induisent des pertes de rendement et diminuent la valeur marchande des semences.

- Certaines bactéries sont responsables de pathologies sur une large gamme de culture économiquement importante.
- Ces pathologies induisent des pertes de rendement et diminuent la valeur marchande des semences.

### Pathovar phaseoli de *Xanthomonas Axonopodis* (Xap)

- Il regroupe toutes les souches pathogènes sur le haricot.

- Certaines bactéries sont responsables de pathologies sur une large gamme de culture économiquement importante.
- Ces pathologies induisent des pertes de rendement et diminuent la valeur marchande des semences.

### Pathovar phaseoli de *Xanthomonas Axonopodis* (Xap)

- Il regroupe toutes les souches pathogènes sur le haricot.
- Il n'est pas endémique en Europe.

- Certaines bactéries sont responsables de pathologies sur une large gamme de culture économiquement importante.
- Ces pathologies induisent des pertes de rendement et diminuent la valeur marchande des semences.

### Pathovar phaseoli de *Xanthomonas Axonopodis* (Xap)

- Il regroupe toutes les souches pathogènes sur le haricot.
- Il n'est pas endémique en Europe.
- Mais pour limiter son introduction, il est inscrit sur la liste des agents pathogènes de quarantaine.

- Certaines bactéries sont responsables de pathologies sur une large gamme de culture économiquement importante.
- Ces pathologies induisent des pertes de rendement et diminuent la valeur marchande des semences.

### Pathovar phaseoli de *Xanthomonas Axonopodis* (Xap)

- Il regroupe toutes les souches pathogènes sur le haricot.
- Il n'est pas endémique en Europe.
- Mais pour limiter son introduction, il est inscrit sur la liste des agents pathogènes de quarantaine.
- La mise en quarantaine des containers de haricots induit de forts coûts de stockage.

## Identification des souches bactériennes

- Obtenir un répertoire de gènes de virulence auprès de biologistes.

## Identification des souches bactériennes

- Obtenir un répertoire de gènes de virulence auprès de biologistes.
- Chercher la plus petite combinaison de gène de virulence spécifique aux souches bactériennes.

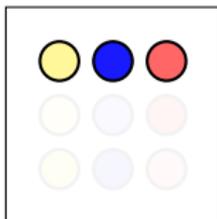
## Identification des souches bactériennes

- Obtenir un répertoire de gènes de virulence auprès de biologistes.
- Chercher la plus petite combinaison de gène de virulence spécifique aux souches bactériennes.
- Utiliser cette combinaison pour la mise au point d'un test PCR<sup>a</sup> Multiplex (Micropuce ADN).

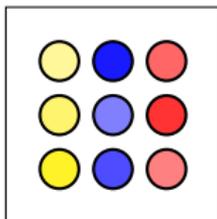
---

a. Polymerase Chain Reaction

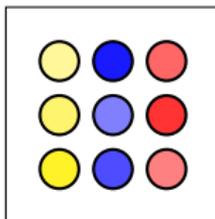
# PCR Multiplex



# PCR Multiplex



# PCR Multiplex



## Remarque

- Nous cherchons la plus petite combinaison de gène de virulence afin de minimiser la taille de la puce et donc son coût.

# Problème de caractérisation multiple (PCM)

## Données du PCM

- Nous disposons d'un ensemble d'entités (souches bactériennes) regroupées en groupes (pathovars).

# Problème de caractérisation multiple (PCM)

## Données du PCM

- Nous disposons d'un ensemble d'entités (souches bactériennes) regroupées en groupes (pathovars).
- Chaque entité est définie par la présence/absence de caractères(gènes).

# Problème de caractérisation multiple (PCM)

## Données du PCM

- Nous disposons d'un ensemble d'entités (souches bactériennes) regroupées en groupes (pathovars).
- Chaque entité est définie par la présence/absence de caractères(gènes).
- Une entité peut être vue comme étant une interprétation booléenne sur les gènes.

# Problème de caractérisation multiple (PCM)

## Données du PCM

- Nous disposons d'un ensemble d'entités (souches bactériennes) regroupées en groupes (pathovars).
- Chaque entité est définie par la présence/absence de caractères(gènes).
- Une entité peut être vue comme étant une interprétation booléenne sur les gènes.
- Les gènes sont donc considérés comme les variables du problème.

# Problème de caractérisation multiple (PCM)

## Données du PCM

- Nous disposons d'un ensemble d'entités (souches bactériennes) regroupées en groupes (pathovars).
- Chaque entité est définie par la présence/absence de caractères(gènes).
- Une entité peut être vue comme étant une interprétation booléenne sur les gènes.
- Les gènes sont donc considérés comme les variables du problème.
- Chaque groupe fournit une table de vérité partielle d'une fonction booléenne dont l'interprétation est vraie pour le groupe en question et fausse pour tous les autres.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

Une caractérisation de taille  $k = n$  est triviale :

- $g_1 : (\neg a \wedge \neg b \wedge \neg c) \vee (\neg a \wedge \neg b \wedge c)$

## Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

Une caractérisation de taille  $k = n$  est triviale :

- $g_1 : (\neg a \wedge \neg b \wedge \neg c) \vee (\neg a \wedge \neg b \wedge c)$
- $g_2 : a \wedge b \wedge c$

## Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

Une caractérisation de taille  $k = n$  est triviale :

- $g_1 : (\neg a \wedge \neg b \wedge \neg c) \vee (\neg a \wedge \neg b \wedge c)$
- $g_2 : a \wedge b \wedge c$
- $g_3 : (a \wedge b \wedge \neg c) \vee (\neg a \wedge b \wedge \neg c)$

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

## Résolution pour $k=2$

Il nous faut parcourir au maximum  $C_3^2$  combinaisons :

- $\{a, b\}$

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

## Résolution pour $k=2$

Il nous faut parcourir au maximum  $C_3^2$  combinaisons :

- $\{a, b\}$
- $\{b, c\}$

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

## Résolution pour $k=2$

Il nous faut parcourir au maximum  $C_3^2$  combinaisons :

- $\{a, b\}$
- $\{b, c\}$
- $\{a, c\}$

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

## Résolution pour $k=2$

Combinaisons  $\{a, b\}$  :

- $e_1(a) \neq e_3(a)$   
 $\Rightarrow$  On passe à l'entité suivante.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

## Résolution pour $k=2$

Combinaison  $\{a, b\}$  :

- $e_1(a) \neq e_4(a)$   
 $\Rightarrow$  On passe à l'entité suivante.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

## Résolution pour $k=2$

Combinaison  $\{a, b\}$  :

- $e_1(a) = e_5(a)$   
⇒ Identique sur le gène *a*, on doit observer le gène *b*.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
e <sub>1</sub>	g <sub>1</sub>	0	0	0
e <sub>2</sub>		0	0	1
e <sub>3</sub>	g <sub>2</sub>	1	1	1
e <sub>4</sub>	g <sub>3</sub>	1	1	0
e <sub>5</sub>		0	1	0

## Résolution pour k=2

Combinaison  $\{a, b\}$  :

- $e_1(b) \neq e_5(b)$   
⇒ Nous devons maintenant observer e<sub>2</sub>.

## Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

Résolution pour  $k=2$ Combinaison  $\{a, b\}$  :

- $e_2(a) \neq e_3(a)$   
 $\Rightarrow$  On passe à l'entité suivante.

## Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

Résolution pour  $k=2$ Combinaison  $\{a, b\}$  :

- $e_2(a) \neq e_4(a)$   
 $\Rightarrow$  On passe à l'entité suivante.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

## Résolution pour $k=2$

Combinaison  $\{a, b\}$  :

- $e_2(a) = e_5(a)$   
⇒ Identique sur le gène *a*, on doit observer le gène *b*.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
e <sub>1</sub>	g <sub>1</sub>	0	0	0
e <sub>2</sub>		0	0	1
e <sub>3</sub>	g <sub>2</sub>	1	1	1
e <sub>4</sub>	g <sub>3</sub>	1	1	0
e <sub>5</sub>		0	1	0

## Résolution pour k=2

Combinaison  $\{a, b\}$  :

- $e_2(b) \neq e_5(b)$   
⇒ La combinaison  $\{a, b\}$  permet de caractériser g<sub>1</sub> des autres groupes.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
e <sub>1</sub>	g <sub>1</sub>	0	0	0
e <sub>2</sub>		0	0	1
e <sub>3</sub>	g <sub>2</sub>	1	1	1
e <sub>4</sub>	g <sub>3</sub>	1	1	0
e <sub>5</sub>		0	1	0

## Résolution pour k=2

Combinaison  $\{a, b\}$  :

- $e_2(b) \neq e_5(b)$   
⇒ La combinaison  $\{a, b\}$  permet de caractériser g<sub>1</sub> des autres groupes.
- Nous devons maintenant observer g<sub>2</sub> vis à vis des autres groupes.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

## Résolution pour $k=2$

Combinaison  $\{a, b\}$  :

- $e_3(a) = e_4(a)$   
⇒ Identique sur le gène *a*, on doit observer le gène *b*.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
e <sub>1</sub>	g <sub>1</sub>	0	0	0
e <sub>2</sub>		0	0	1
e <sub>3</sub>	g <sub>2</sub>	1	1	1
e <sub>4</sub>	g <sub>3</sub>	1	1	0
e <sub>5</sub>		0	1	0

## Résolution pour k=2

Combinaison  $\{a, b\}$  :

- $e_3(b) = e_4(b)$   
⇒ Identique sur le gène  $b$

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
e <sub>1</sub>	g <sub>1</sub>	0	0	0
e <sub>2</sub>		0	0	1
e <sub>3</sub>	g <sub>2</sub>	1	1	1
e <sub>4</sub>	g <sub>3</sub>	1	1	0
e <sub>5</sub>		0	1	0

## Résolution pour k=2

Combinaison  $\{a, b\}$  :

- $e_3(b) = e_4(b)$   
⇒ Identique sur le gène  $b$
- Échec de la caractérisation avec la combinaison  $\{a, b\}$ .

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

## Résolution pour $k=2$

Combinaison  $\{b, c\}$  :

- $e_1(b) \neq e_3(b)$   
 $\Rightarrow$  On passe à l'entité suivante.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

## Résolution pour $k=2$

Combinaison  $\{b, c\}$  :

- $e_1(b) \neq e_4(b)$   
⇒ On passe à l'entité suivante.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

## Résolution pour $k=2$

Combinaison  $\{b, c\}$  :

- $e_1(b) \neq e_5(b)$   
 $\Rightarrow$  Nous devons maintenant observer  $e_2$ .

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

## Résolution pour $k=2$

Combinaison  $\{b, c\}$  :

- $e_2(b) \neq e_3(b)$   
 $\Rightarrow$  On passe à l'entité suivante.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

## Résolution pour $k=2$

Combinaison  $\{b, c\}$  :

- $e_2(b) \neq e_4(b)$   
⇒ On passe à l'entité suivante.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

## Résolution pour $k=2$

Combinaison  $\{b, c\}$  :

- $e_2(b) \neq e_5(b)$   
 $\Rightarrow$  La combinaison  $\{b, c\}$   
permet de caractériser  $g_1$   
des autres groupes.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
e <sub>1</sub>	g <sub>1</sub>	0	0	0
e <sub>2</sub>		0	0	1
e <sub>3</sub>	g <sub>2</sub>	1	1	1
e <sub>4</sub>	g <sub>3</sub>	1	1	0
e <sub>5</sub>		0	1	0

## Résolution pour k=2

Combinaison  $\{b, c\}$  :

- $e_2(b) \neq e_5(b)$   
⇒ La combinaison  $\{b, c\}$  permet de caractériser g<sub>1</sub> des autres groupes.
- Nous devons maintenant observer g<sub>2</sub> vis à vis des autres groupes.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

## Résolution pour $k=2$

Combinaison  $\{b, c\}$  :

- $e_3(b) = e_4(b)$   
⇒ Identique sur le gène  $b$ , on doit observer le gène  $c$ .

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

## Résolution pour $k=2$

Combinaison  $\{b, c\}$  :

- $e_3(c) \neq e_4(c)$   
⇒ On passe à l'entité suivante.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

## Résolution pour $k=2$

Combinaison  $\{b, c\}$  :

- $e_3(b) = e_5(b)$   
⇒ Identique sur le gène  $b$ , on doit observer le gène  $c$ .

## Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
e <sub>1</sub>	g <sub>1</sub>	0	0	0
e <sub>2</sub>		0	0	1
e <sub>3</sub>	g <sub>2</sub>	1	1	1
e <sub>4</sub>	g <sub>3</sub>	1	1	0
e <sub>5</sub>		0	1	0

## Résolution pour k=2

Combinaison {b, c} :

- $e_3(c) \neq e_5(c)$   
⇒ La combinaison {b, c} permet de caractériser g<sub>2</sub> des autres groupes.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

## Résolution pour $k=2$

Combinaison  $\{b, c\}$  :

- $e_3(c) \neq e_5(c)$   
 $\Rightarrow$  La combinaison  $\{b, c\}$  permet de caractériser  $g_2$  des autres groupes.
- $\Rightarrow$  La combinaison  $\{b, c\}$  permet de caractériser cette instance.

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

Formules de caractérisation de l'instance :

- $g_1 : (\neg b \wedge \neg c) \vee (\neg b \wedge c)$

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

Formules de caractérisation de l'instance :

- $g_1 : (\neg b \wedge \neg c) \vee (\neg b \wedge c)$
- $g_2 : b \wedge c$

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

Formules de caractérisation de l'instance :

- $g_1 : (\neg b \wedge \neg c) \vee (\neg b \wedge c)$
- $g_2 : b \wedge c$
- $g_3 : b \wedge \neg c$

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

Formules de caractérisation de l'instance :

- $g_1 : (\neg b \wedge \neg c) \vee (\neg b \wedge c)$
- $g_2 : b \wedge c$
- $g_3 : b \wedge \neg c$

Entité  $e_x = 1\ 0\ 0?$

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		a	b	c
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

Formules de caractérisation de l'instance :

- $g_1 : (\neg b \wedge \neg c) \vee (\neg b \wedge c)$
- $g_2 : b \wedge c$
- $g_3 : b \wedge \neg c$

Entité  $e_x = 1\ 0\ 0\ ?$

- Soit elle fait partie du groupe  $g_1$ .

# Représentation et résolution d'une instance PCM

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
<i>e</i> <sub>1</sub>	<i>g</i> <sub>1</sub>	0	0	0
<i>e</i> <sub>2</sub>		0	0	1
<i>e</i> <sub>3</sub>	<i>g</i> <sub>2</sub>	1	1	1
<i>e</i> <sub>4</sub>	<i>g</i> <sub>3</sub>	1	1	0
<i>e</i> <sub>5</sub>		0	1	0

Formules de caractérisation de l'instance :

- $g_1 : (\neg b \wedge \neg c) \vee (\neg b \wedge c)$
- $g_2 : b \wedge c$
- $g_3 : b \wedge \neg c$

Entité  $e_x = 1\ 0\ 0\ ?$

- Soit elle fait partie du groupe  $g_1$ .
- Soit les biologistes créent un nouveau groupe pour  $e_x$  et il faut recaractériser la nouvelle instance.

## Minimisation du problème de caractérisation multiple(MIN-PCM)

- Consiste à chercher la caractérisation de taille  $k$  contenant le moins de variables.

## Minimisation du problème de caractérisation multiple(MIN-PCM)

- Consiste à chercher la caractérisation de taille  $k$  contenant le moins de variables.
- Permet de minimiser la taille du PCR-Multiplex et donc son coût.

## Minimisation du problème de caractérisation multiple(MIN-PCM)

- Consiste à chercher la caractérisation de taille  $k$  contenant le moins de variables.
- Permet de minimiser la taille du PCR-Multiplex et donc son coût.
- Lors de la recherche, nous essayons de caractériser à partir de  $k = n - 1$  jusqu'à ce qu'il n'y ait plus de caractérisation possible.

## Complexité

- Le PCM appartient à la classe de complexité NP-complet.

## Complexité

- Le PCM appartient à la classe de complexité NP-complet.
- MIN-PCM appartient à la classe de complexité NP-difficile.

## Complexité

- Le PCM appartient à la classe de complexité NP-complet.
- MIN-PCM appartient à la classe de complexité NP-difficile.
- Plus spécifiquement, le PCM appartient à la classe de complexité  $W[2]$ -complet.

## Complexité

- Le PCM appartient à la classe de complexité NP-complet.
- MIN-PCM appartient à la classe de complexité NP-difficile.
- Plus spécifiquement, le PCM appartient à la classe de complexité  $W[2]$ -complet.
- MIN-PCM appartient à la classe de complexité  $W[2]$ -difficile.

## Complexité

- Le PCM appartient à la classe de complexité NP-complet.
- MIN-PCM appartient à la classe de complexité NP-difficile.
- Plus spécifiquement, le PCM appartient à la classe de complexité  $W[2]$ -complet.
- MIN-PCM appartient à la classe de complexité  $W[2]$ -difficile.
- La seule possibilité d'améliorer significativement la résolution complète est d'utiliser des heuristiques sur les choix de variables.

## Complexité

- Le PCM appartient à la classe de complexité NP-complet.
- MIN-PCM appartient à la classe de complexité NP-difficile.
- Plus spécifiquement, le PCM appartient à la classe de complexité  $W[2]$ -complet.
- MIN-PCM appartient à la classe de complexité  $W[2]$ -difficile.
- La seule possibilité d'améliorer significativement la résolution complète est d'utiliser des heuristiques sur les choix de variables.

## Problématique

- A partir des conclusions des précédents travaux, nous cherchons à savoir quels sont les gènes à examiner en priorités lors de la résolution d'une instance MIN-PCM.

# Plan

- Définition d'une instance difficile
- Heuristiques pour une recherche exacte
- Recherche approchée
- Conclusions et perspectives

Instances	Entités	Gènes	Borne min connue
s3836-0	1000	1000	16
rch10	173	98	<b>10</b>

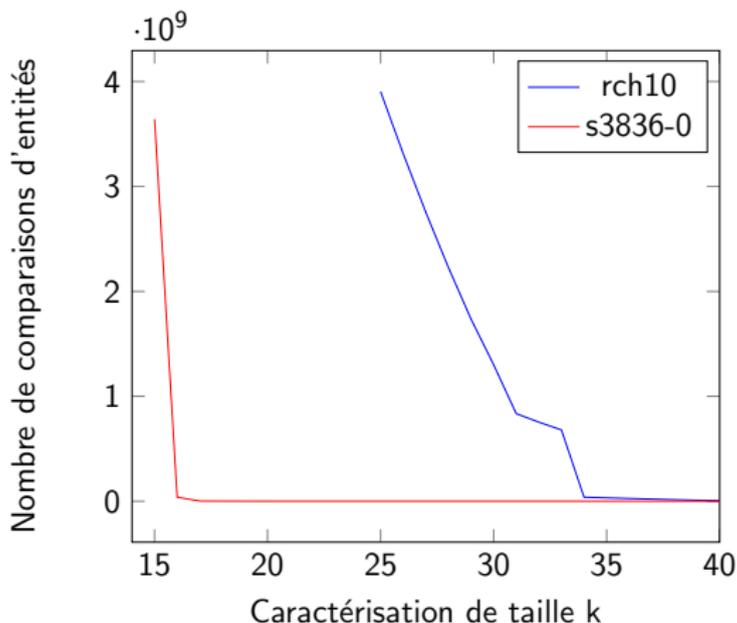
Instances	Entités	Gènes	Borne min connue
s3836-0	1000	1000	16
rch10	173	98	<b>10</b>

- Borne minimum connue de s3836-0 obtenue par une méthode de type recherche arborescente.

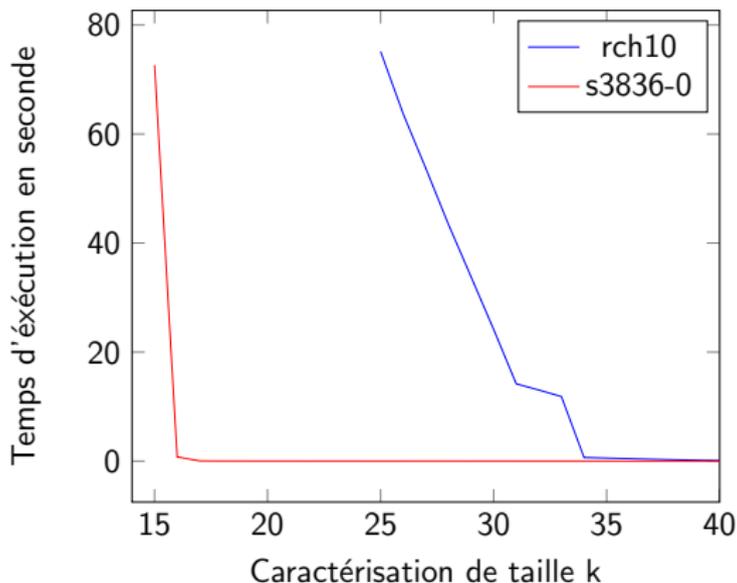
Instances	Entités	Gènes	Borne min connue
s3836-0	1000	1000	16
rch10	173	98	<b>10</b>

- Borne minimum connue de s3836-0 obtenue par une méthode de type recherche arborescente.
- Borne minimum connue de rch10 obtenue par une transformation du MIN-PCM en MIN-ONE, résolue par le solveur *cplex* d'IBM.

Instances	Entités	Gènes	Borne min connue
s3836-0	1000	1000	16
rch10	173	98	<b>10</b>



Instances	Entités	Gènes	Borne min connue
s3836-0	1000	1000	16
rch10	173	98	<b>10</b>



## Masque et ratio d'un groupe

Soit le groupe suivant :

Entités	Gènes			
	g0	g1	g2	g3
e1	1	0	1	1
e2	1	0	1	0
e3	1	0	0	0
e4	1	0	1	0
e5	1	0	1	1
Masque	1	0	0.8	0.4

## Masque et ratio d'un groupe

Soit le groupe suivant :

Entités \ Gènes	Gènes			
	g0	g1	g2	g3
e1	1	0	1	1
e2	1	0	1	0
e3	1	0	0	0
e4	1	0	1	0
e5	1	0	1	1
Masque	1	0	0.8	0.4

Le ratio  $r$  de ce groupe est  $r = 2/4$  soit  $r = 0.5$

## Masque et ratio d'un groupe

Soit le groupe suivant :

Entités \ Gènes	Gènes			
	g0	g1	g2	g3
e1	1	0	1	1
e2	1	0	1	0
e3	1	0	0	0
e4	1	0	1	0
e5	1	0	1	1
Masque	1	0	0.8	0.4

Le ratio  $r$  de ce groupe est  $r = 2/4$  soit  $r = 0.5$

- Les instances réelles (resp. aléatoire) sont constituées de groupes ayant un fort (resp. faible) ratio.

Image d'une instance et taux de similarité  $\mathcal{T}$  des gènes

Soit l'image suivante :

Groupes \ Gènes	Gènes			
	g0	g1	g2	g3
Masque de g1	0.9	0.1	0.5	0.6
Masque de g2	0.9	0.1	0.5	0.6
Masque de g3	0.9	0.1	0.5	0.6
$\mathcal{T}$	0.8	0.8	0	0.2

## Image d'une instance et taux de similarité $\mathcal{T}$ des gènes

Soit l'image suivante :

Gènes \ Groupes	g0	g1	g2	g3
Masque de $g_1$	0.9	0.1	0.5	0.6
Masque de $g_2$	0.9	0.1	0.5	0.6
Masque de $g_3$	0.9	0.1	0.5	0.6
$\mathcal{T}$	0.8	0.8	0	0.2

- Plus la moyenne des taux  $\mathcal{T}$  est élevée, plus l'instance sera difficile à résoudre.

Image d'une instance et taux de similarité  $\mathcal{T}$  des gènes

Soit l'image suivante :

Groupes \ Gènes	Gènes			
	g0	g1	g2	g3
Masque de $g_1$	0.9	0.1	0.5	0.6
Masque de $g_2$	0.9	0.1	0.5	0.6
Masque de $g_3$	0.9	0.1	0.5	0.6
$\mathcal{T}$	0.8	0.8	0	0.2

- Plus la moyenne des taux  $\mathcal{T}$  est élevée, plus l'instance sera difficile à résoudre.
- Les instances réelles présentent une moyenne de taux  $\mathcal{T}$  beaucoup plus élevée que les instances aléatoires (i.e : les groupes sont fortement similaires).

## Image d'une instance et taux de similarité $\mathcal{T}$ des gènes

Soit l'image suivante :

Gènes \ Groupes	g0	g1	g2	g3
Masque de $g_1$	0.9	0.1	0.5	0.6
Masque de $g_2$	0.9	0.1	0.5	0.6
Masque de $g_3$	0.9	0.1	0.5	0.6
$\mathcal{T}$	0.8	0.8	0	0.2

## Heuristique de trie des gènes par $\mathcal{T}$

- Les gènes présentant des taux de similarité faible ont plus de chance de caractériser une instance.

## Image d'une instance et taux de similarité $\mathcal{T}$ des gènes

Soit l'image suivante :

Gènes \ Groupes	g0	g1	g2	g3
Masque de $g_1$	0.9	0.1	0.5	0.6
Masque de $g_2$	0.9	0.1	0.5	0.6
Masque de $g_3$	0.9	0.1	0.5	0.6
$\mathcal{T}$	0.8	0.8	0	0.2

## Heuristique de trie des gènes par $\mathcal{T}$

- Les gènes présentant des taux de similarité faible ont plus de chance de caractériser une instance.
- L'heuristique de trie par  $\mathcal{T}$  consiste à examiner en priorité les gènes ayant de faible taux de similarité.

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

### Heuristique de trie des groupes par $\Gamma$

- Une caractérisation exige un parcours exhaustif de paires de groupes ( $[g_1, g_2]$ ,  $[g_1, g_3]$ ,  $[g_2, g_3]$ ).

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

### Heuristique de trie des groupes par $\Gamma$

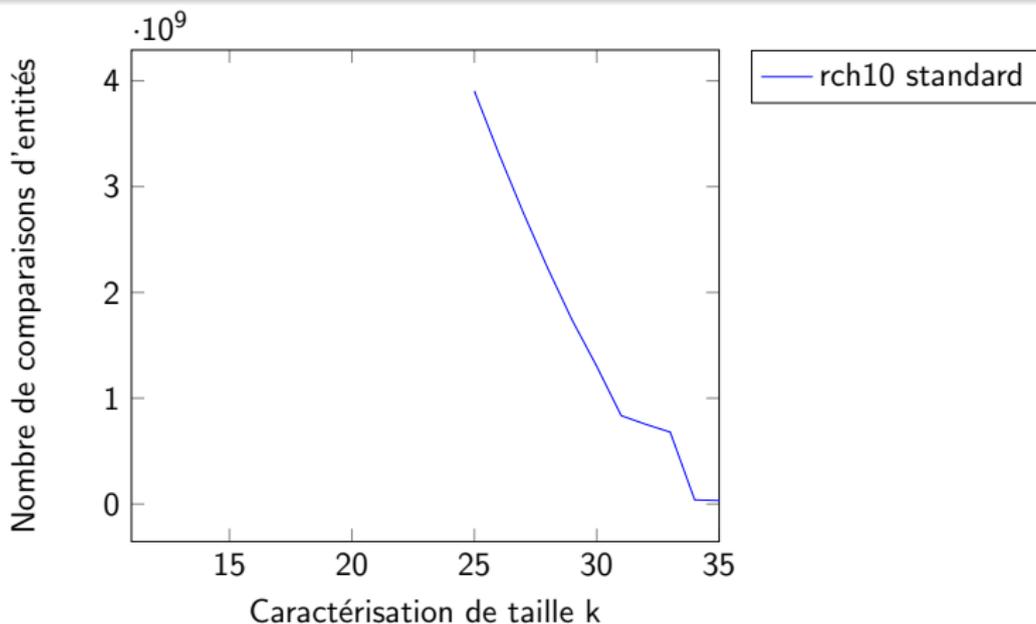
- Une caractérisation exige un parcours exhaustif de paires de groupes ( $[g_1, g_2]$ ,  $[g_1, g_3]$ ,  $[g_2, g_3]$ ).
- Idée : Parcourir en priorités les paires de groupes ayant des forts taux de similarité  $\Gamma$  entre groupe ( $[g_2, g_3]$ ,  $[g_3, g_1]$ ,  $[g_1, g_2]$ ).

Souches	Groupes	Gènes		
		<i>a</i>	<i>b</i>	<i>c</i>
$e_1$	$g_1$	0	0	0
$e_2$		0	0	1
$e_3$	$g_2$	1	1	1
$e_4$	$g_3$	1	1	0
$e_5$		0	1	0

### Heuristique de trie des groupes par $\Gamma$

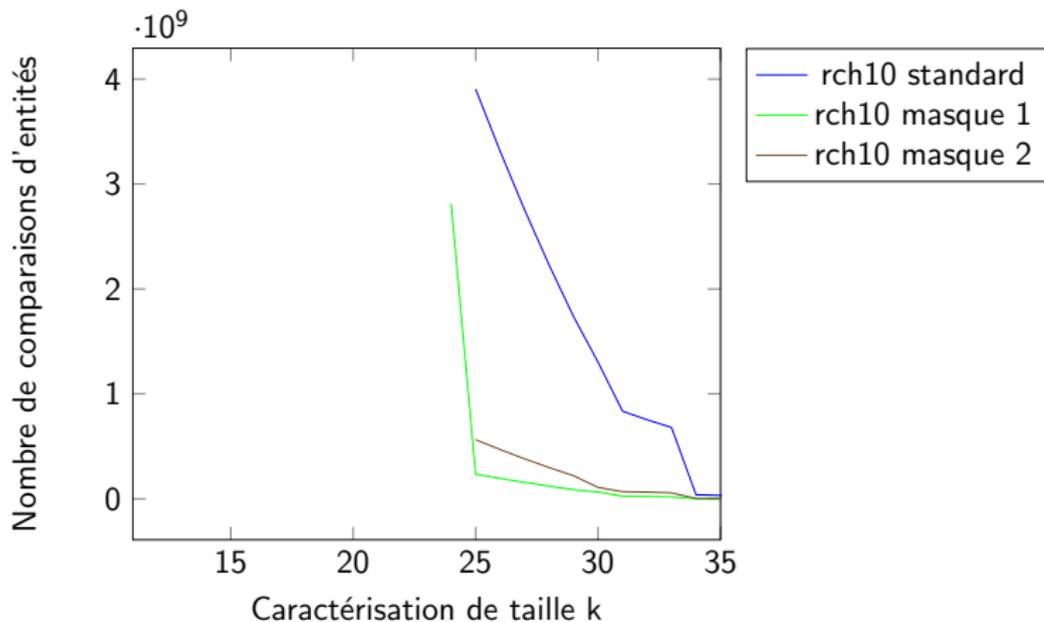
- Une caractérisation exige un parcours exhaustif de paires de groupes ( $[g_1, g_2]$ ,  $[g_1, g_3]$ ,  $[g_2, g_3]$ ).
- Idée : Parcourir en priorités les paires de groupes ayant des forts taux de similarité  $\Gamma$  entre groupe ( $[g_2, g_3]$ ,  $[g_3, g_1]$ ,  $[g_1, g_2]$ ).
- Elimination rapide des combinaisons infructueuses.

# Heuristiques mises en œuvre sur une instance réelle

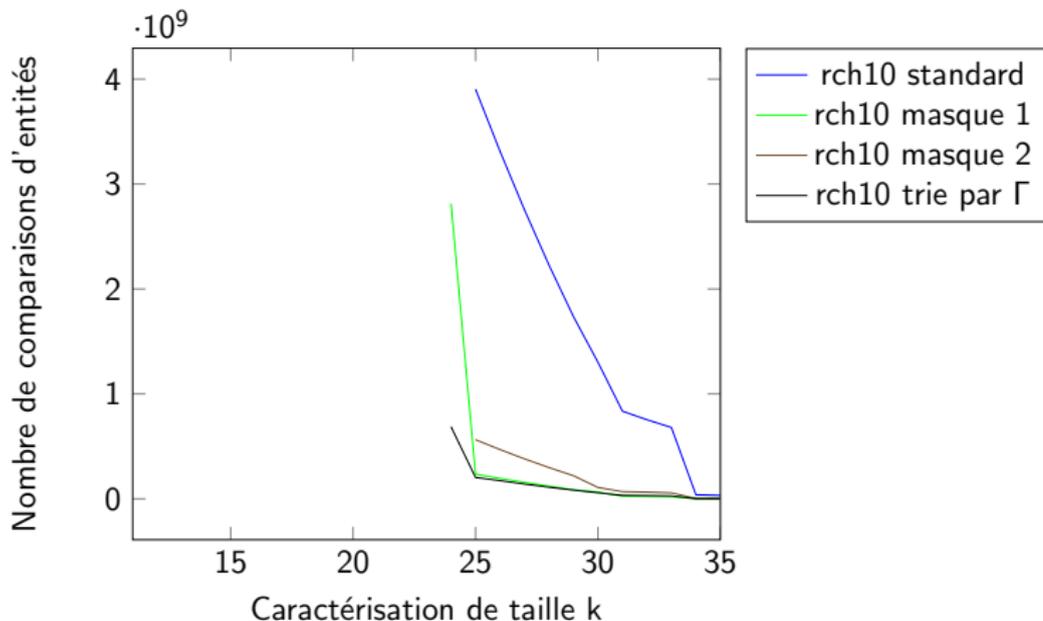


- standard : parcours par ordre lexicographique des caractères.

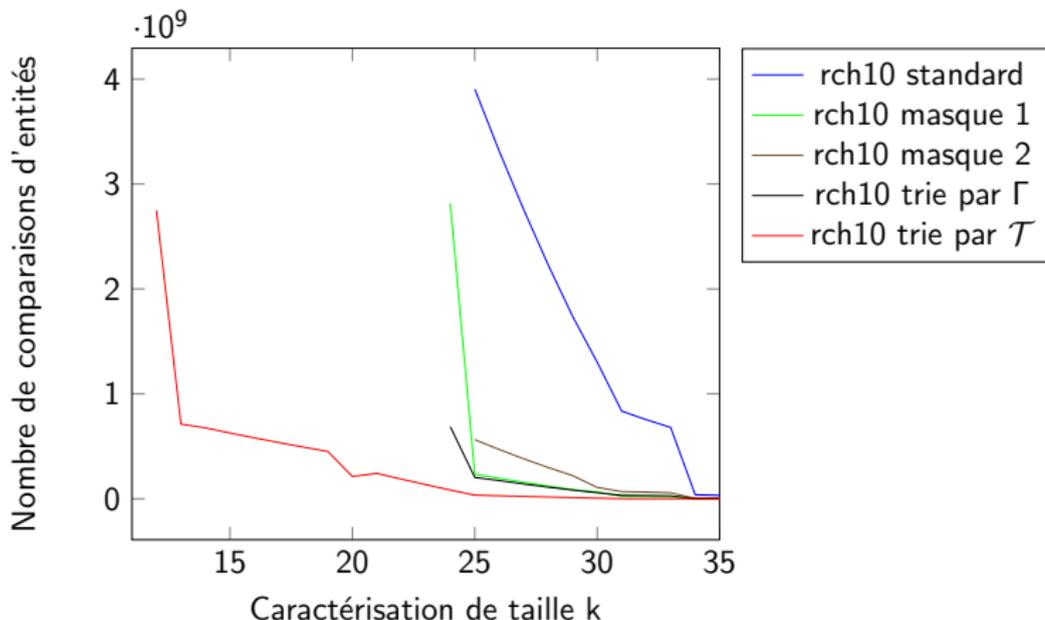
# Heuristiques mises en œuvre sur une instance réelle



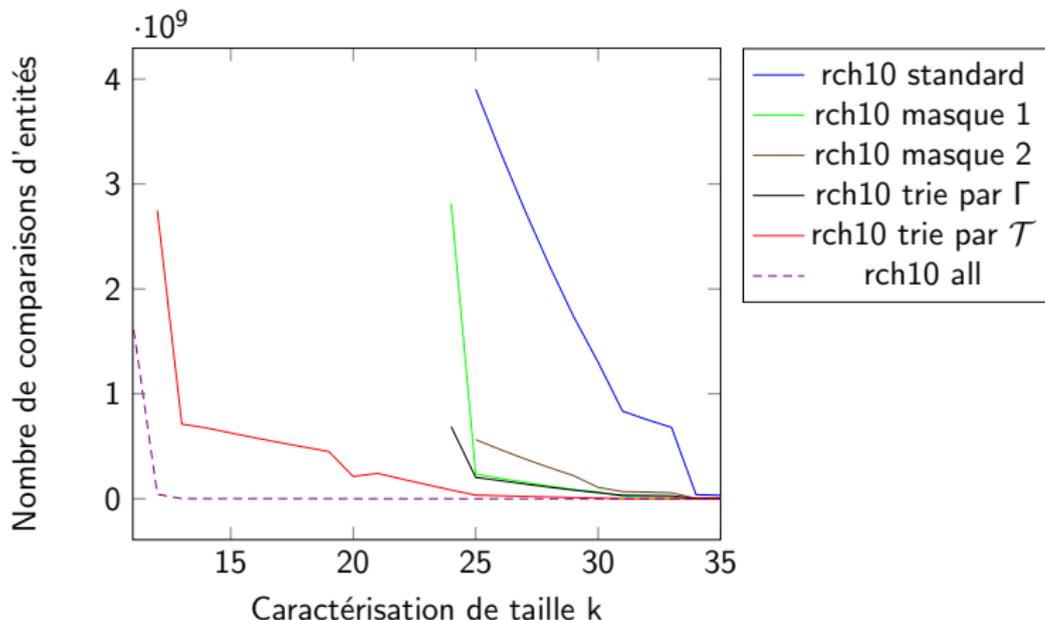
# Heuristiques mises en œuvre sur une instance réelle



# Heuristiques mises en œuvre sur une instance réelle



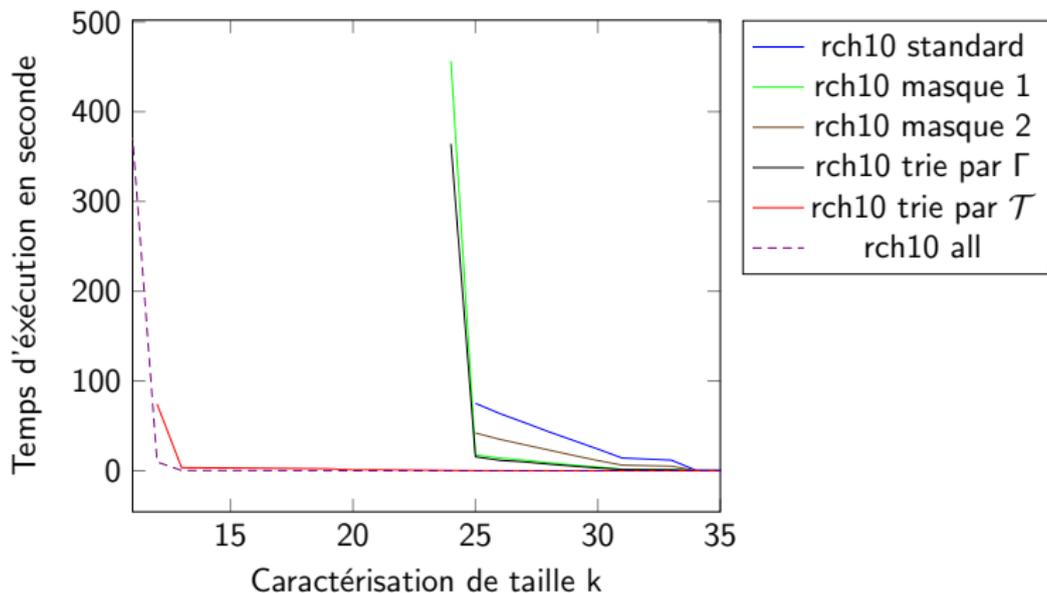
# Heuristiques mises en œuvre sur une instance réelle



# Heuristiques mises en œuvre sur une instance réelle

- Observons maintenant les temps d'exécutions.

# Heuristiques mises en œuvre sur une instance réelle



# Algorithme de recherche approchée par roulette proportionnelle

**roulette** (*taux*, *Groupes*, *k*, *n*)

Réel *proba*[]  $\leftarrow$  initialisation\_proba(*taux*)

# Algorithme de recherche approchée par roulette proportionnelle

**roulette** (*taux*, *Groupes*, *k*, *n*)

Réel *proba*[]  $\leftarrow$  *initialisation\_proba*(*taux*)

**tant que** *VRAI* **faire**

|

# Algorithme de recherche approchée par roulette proportionnelle

**roulette** (*taux*, *Groupes*, *k*, *n*)

Réel *proba*[]  $\leftarrow$  *initialisation\_proba*(*taux*)

**tant que** *VRAI* **faire**

Entier *combinaison*  $\leftarrow$  *genereCombinaisonAvecProba*(*proba*)

# Algorithme de recherche approchée par roulette proportionnelle

**roulette** (*taux*, *Groupes*, *k*, *n*)

Réel *proba*[]  $\leftarrow$  *initialisation\_proba*(*taux*)

**tant que** *VRAI* **faire**

Entier *combinaison*  $\leftarrow$  *genereCombinaisonAvecProba*(*proba*)

**si** *caracterise\_instance*(*l*, *combinaison*) **alors**

afficher("La combinaison ", *combinaison*, " permet de caractériser l'instance.")

*k*  $\leftarrow$  *k* - 1

*combinaison*  $\leftarrow$  {}

**fin**

**fin**

## Recherche approchée

## Résultats sur instances réels

Instances	Entités	Gènes	B. Min	Roulette proportionnelle	
				k	temps
rch8	56	27	<b>9</b>	9	0.031
raphv	108	68	<b>6</b>	6	0.657
raphy	112	70	<b>6</b>	6	0.873
rarep	112	72	<b>12</b>	14	36.627
rch10	112	86	<b>10</b>	12	65.615

# Conclusions

- Définitions de critères permettant de déterminer la difficulté d'une instance.

# Conclusions

- Définitions de critères permettant de déterminer la difficulté d'une instance.
- Mise en place d'heuristiques permettant des résolutions beaucoup plus efficaces sur des instances réelles.

# Conclusions

- Définitions de critères permettant de déterminer la difficulté d'une instance.
- Mise en place d'heuristiques permettant des résolutions beaucoup plus efficaces sur des instances réelles.
- Mise en place d'une recherche approchée fournissant d'excellents résultats.

# Perspectives

- Améliorer la recherche approchée.

# Perspectives

- Améliorer la recherche approchée.
- Développer un générateur d'instances simulant des instances réelles.

# Perspectives

- Améliorer la recherche approchée.
- Développer un générateur d'instances simulant des instances réelles.
- Travailler sur des instances réelles issues d'autres disciplines que celle de la biologie végétale (médecine, ...).

Merci pour votre attention.