

## Identification probabiliste (numérique)

L'identification numérique repose sur l'utilisation d'une base de donnée.

L'identification d'une souche consiste à comparer de façon mathématique le profil d'une souche inconnue avec celui des différents taxons de la base de données.

1. La base de donnée est constituée des % **de réactions positives (P)** de chaque test pour chaque taxon recensé.
2. La **fréquence d'apparition des réactions (F)** observées est calculée de la façon suivante :
  - si le caractère est + : on retient le % divisé par 100
  - si le caractère est - : on retient le 1 - (% divisé par 100)

Remarque : le vrai calcul diffère très légèrement car tient compte des risques d'erreurs de lecture. Ceci explique que le chiffre 0,99 est retenu pour 100% (au lieu de 1) et de 0,01 pour 0% (au lieu de 0).

3. La **fréquence d'apparition du profil (PO)** est obtenu en faisant le produit de l'ensemble des fréquences d'apparition des réactions pour chaque taxon.

4. Le **% d'identification (%Id)** est obtenu en divisant la fréquence d'apparition du profile par la somme de toutes les fréquences d'apparition des différents taxons.

Le %Id exprime la **proximité relative** du profil aux différents taxons.

5. Les taxons sont ensuite classés par **valeur décroissante du %Id**.

6. Un indice de **typicité T** est calculé à partir des fréquences modales (calcul non décrit ici).

L'indice T varie de 0 à 1 : il exprime la proximité du profil au taxon, autrement dit le nom et l'importance des **tests « à l'encontre »**.

**Exemple** (pour comprendre....) : utiliser la base de données ci-jointe.

On considère les profils suivants

	Test 1	Test 2	Test 3	Test4	commentaire
Profil 1	-	+	+	+	Taxon B 97,35% T=1
Profil 2	+	-	-	+	Aucun taxon retenu
Profil 3	-	+	-	+	<b>B acceptable</b>
Profil 4	-	+	+	-	D ou E

On suppose pour le profil 4 qu'un taxon E supplémentaire soit présent dans la base de données.

## EXEMPLE D'IDENTIFICATION

### 1/ BASE DE DONNEES (4 TESTS, 4 TAXONS)

	Test 1	Test 2	Test 3	Test 4
Taxon A	50	93	87	2
Taxon B	0	100	85	98
Taxon C	15	15	91	3
Taxon D	0	95	98	1
Taxon E	2	98	92	10

### 2/ FREQUENCES D'APPARITION DES REACTIONS

	SOUCHES ETUDIEES				PROFIL LE PLUS TYPIQUE			
	Test 1	Test 2	Test 3	Test 4	Test 1	Test 2	Test 3	Test 4
Profil	-	+	+	+				
Taxon A	0.990	0.929	0.869	0.021	0.50	0.929	0.869	0.970
Taxon B	0.990	0.990	0.849	0.979	0.99	0.99	0.849	0.979
Taxon C	0.843	0.151	0.909	0.031	0.843	0.843	0.909	0.961
Taxon D	0.990	0.949	0.979	0.011	0.99	0.949	0.979	0.980

### 3/ FREQUENCES D'APPARITION DU PROFIL

	PROFIL OBSERVE (PO)	PROFIL LE PLUS TYPIQUE (PT)
Taxon A	$0,500 \times 0,929 \times 0,869 \times 0,021 = 0,00846$	$0,500 \times 0,929 \times 0,869 \times 0,970 = 0,39154$
Taxon B	$0,990 \times 0,990 \times 0,849 \times 0,979 = 0,81463$	$0,990 \times 0,990 \times 0,849 \times 0,979 = 0,81463$
Taxon C	$0,843 \times 0,151 \times 0,909 \times 0,031 = 0,00359$	$0,843 \times 0,843 \times 0,909 \times 0,961 = 0,62079$
Taxon D	$0,990 \times 0,949 \times 0,979 \times 0,011 = 0,01012$	$0,990 \times 0,949 \times 0,979 \times 0,980 = 0,90136$

### 4/ POURCENTAGES D'IDENTIFICATION

Taxon A	$(0.00846 / 0.83680) \times 100 = 1.01$
Taxon B	$(0.81463 / 0.83680) \times 100 = 97.35$
Taxon C	$(0.00359 / 0.83680) \times 100 = 0.43$
Taxon D	$(0.01012 / 0.83680) \times 100 = 1.21$
	100.00

### FREQUENCES MODALES (PO/ PT)

$0.00846 / 0.39154 = 0.0196$
$0.81463 / 0.81463 = 1$
$0.00359 / 0.62079 = 0.0058$
$0.01012 / 0.90138 = 0.0092$

### 5/ CLASSEMENT DES TAXONS ET INDICES T

Taxon B	%id = 97.35	T = 1	$R_1 = 97.35 / 1.21 = 80.45$
Taxon D	%id = 1.21	T = 0	$R_2 = 1.21 / 1.01 = 1.2$
Taxon A	%id = 1.01	T = 0.15	$R_3 = 1.01 / 0.43 = 2.34$
Taxon C	%id = 0.43	T = 0	

Le plus grand rapport est  $R_T$

Seul sera retenu le taxon B : %id = 97.35      T = 1

Le commentaire dans ce cas sera : Bonne identification au taxon B