

Principe de l'identification numérique

L'identification numérique repose sur le calcul pour le profil observé :

- de sa proximité relative aux différents taxons de la base de données (%id)
- de sa proximité au profil le plus typique dans chaque taxon (indice T).

Le classement des taxons permet ensuite de proposer un résultat d'identification.

L'exemple ci-contre permet de suivre le détail des opérations.

1/ La base de données est constituée des pourcentages (P) de réaction positive de chaque couple taxon/test.

2/ Les fréquences d'apparition des réactions observées (F+ et F-) en sont dérivées :

- à partir des fréquences de réactions positives ($P+ = P / 100$) ou négatives ($P- = 1 - (P / 100)$)
 - en tenant compte de risques d'erreur de lecture : risques $\alpha+$ pour une lecture positive et $\alpha-$ pour une lecture négative
 - les formules de calcul sont les suivantes :
 - réaction lue positive $F+ = P+ (1 - \alpha+) + (\alpha+ \times P-)$
 - réaction lue négative $F- = P- (1 - \alpha-) + (\alpha- \times P+)$
- exemple : base de données $P = 60$; $P+ = 0,60$; $P- = 0,40$; $\alpha+ = 10^{-3}$; $\alpha- = 10^{-2}$
- $F+ = 0,60 (1 - 0,001) + (0,001 \times 0,40) = 0,5998$
- $F- = 0,40 (1 - 0,01) + (0,01 \times 0,60) = 0,402$

Parallèlement sont calculées les fréquences d'apparition des réactions du profil le plus typique, avec P+ lorsque $P \geq 50$ et P- lorsque $P < 50$.

Exemple : $P = 70$ $F+ = 0,699$ $P = 50$ $F+ = 0,50$ $P = 20$ $F- = 0,79$

3/ La fréquence d'apparition du profil dans chaque taxon est égale au produit de l'ensemble des fréquences d'apparition, dans le taxon, des réactions observées. Ces fréquences sont ensuite normées suivant LAPAGE, en divisant chacune par leur somme totale, et une multiplication par 100 donne %id pourcentage d'identification.

Parallèlement est calculée pour chaque taxon une fréquence modale à partir des réactions les plus typiques (DYBOWSKI).

4/ Les taxons sont classés par valeur décroissante des %id

- Pour les 4 premiers taxons est calculé le rapport (R) de leur %id à celui du taxon suivant. Un de ces rapports a une valeur maximale. Sont retenus pour l'identification le taxon ayant en rapport maximal ainsi que, s'il y en a, les taxons situés avant lui dans le classement.
- Un indice T est calculé à partir de chaque fréquence modale (*).

Si un seul taxon est retenu, et si $\%id \geq 80,0$, il est proposé pour l'identification, avec un commentaire sélectionne sur les valeurs de son %id et de son indice T.

Si plusieurs taxons sont retenus et avec une somme des $\%id \geq 80,0$, ils sont proposés pour l'identification avec un commentaire sélectionne sur la valeur de la somme des %id et sur la moyenne ces indices T.

(*) formule de calcul de l'indice T :

$$T = (\log F_m - \log S) / - \log S$$

S est une valeur liée au nombre de tests de la base de données. Dans l'exemple : $S = 10^{-2}$

Utilisation du catalogue analytique

Le profil numérique obtenu est recherché dans le Catalogue.

Cinq cas sont possibles :

- 1 identification au taxon : un seul taxon a été retenu
- 2 identification à l'espèce : 2, 3 ou 4 taxons appartenant à la même espèce ont été retenus
- 3 identification au genre : 5, 3 ou 1 taxons appartenant au même genre ont été retenus
- 4 faible discrimination : 2, 3 ou 4 taxons appartenant à des genres différents ont été retenus
- 5 le profil ne figure pas dans le Catalogue Analytique. Ceci est possible lorsque le profil est trop atypique ou sa fréquence d'apparition très faible.

Un complément d'information peut alors être obtenu auprès du Service d'Aide au Diagnostic API.

Le profil est accompagné des informations suivantes :

a) valeur des deux indices suivants

- Le pourcentage d'identification (%id). Il estime la proximité relative du profil observé aux différents taxons de la base de données.
- L'indice T. Il exprime le caractère typique de la souche à l'intérieur d'un taxon. Sa valeur qui varie entre 0 et 1, reflète le nombre de tests à l'encontre. (Voir explications page XIV)

b) commentaire sur la qualité de l'identification, qui prend en compte le pourcentage d'identification (ou la somme des %id, si plusieurs taxons sont proposés) et l'indice T (ou la moyenne).

EXCELLENTE IDENTIFICATION	%id ≥ 99,9 et T ≥ 0,75
TPES BONNE IDENTIFICATION	%id ≥ 99,0 et T ≥ 0,5
BONNE IDENTIFICATION	%id ≥ 90,0 et T ≥ 0,25
IDENTIFICATION ACCEPTABLE	%id ≥ 80,0 et T ≥ 0

c) tests à l'encontre, s'il y en a, avec pourcentage de positivité (il s'agit de tests dont le résultat est inhabituel pour le taxon retenu. La valeur seuil est de 0,25).

d) dans le cas d'identification à l'espèce, au genre ou de faible discrimination, des TESTS COMPLEMENTAIRES sont présents, pour compléter l'identification. Ces tests sont extraits de la littérature.

(.../...)

PRINCIPE DE L'IDENTIFICATION NUMERIQUE

BASE DE DONNEES

	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
Souche étudiée	-	+	+	+

FREQUENCE D APPARITION DES REACTIONS ET DU PROFIL

Formule pour calculer la fréquence d'apparition d'une réaction

$$F+ = P+ (1 - 10^{-3}) + (10^{-3} \times P-)$$

$$F- = P- (1 - 10^{-2}) + (10^{-2} \times P+)$$

Formule pour calculer la fréquence d'apparition d'un profil

FP = produit des fréquences des différentes réactions composant le profil

Cas du profil typique

Calculer F+, si le test est majoritairement positif dans le taxon étudié (≥ 50)

Calculer F-, si le test est majoritairement négatif dans le taxon étudié (< 50)

	Test 1	Test 2	Test 3	Test 4	FPT
Taxon A	0,500	0,929	0,869	0,970	0,39188
Taxon B	0,990	0,999	0,849	0,979	0,82236
Taxon C	0,843	0,843	0,909	0,961	0,62065
Taxon D	0,990	0,949	0,979	0,980	0,90170

Cas du profil de la souche étudiée (Profil observé)

Calculer F+, si le test est positif pour la souche étudiée

Calculer F-, si le test est négatif pour la souche étudiée

	Test 1	Test 2	Test 3	Test 4	FPT
Taxon A	0,500	0,929	0,869	0,021	0,00846
Taxon B	0,990	0,999	0,849	0,979	0,82236
Taxon C	0,843	0,151	0,909	0,031	0,00357
Taxon D	0,990	0,949	0,979	0,011	0,01010
			Somme des FPO =		0,84450

POURCENTAGES D'IDENTIFICATION

Attention : ces pourcentages se calculent pour la souche étudiée (à partir des FPO)

% identification = $(FPO / \text{Somme des FPO}) \times 100$

Taxon A	1,00
Taxon B	97,38
Taxon C	0,42
Taxon D	1,20
	100,00

FREQUENCE MODALE

Se calcule en faisant le rapport pour chaque taxon : FPO / FPT

Taxon A	0,0216
Taxon B	1,0000
Taxon C	0,0058
Taxon D	0,0112

INDICE DE TYPICITE

Permet d'évaluer la caractère typique de la souche au sein d'un taxon. Reflète le nombre de tests à rencontre.

T =	0,17	
T =	1,00	$T = (\log F_m - \log S) / -\log S$
T =	0,12	
T =	0,02	lci S = 10^{-2}