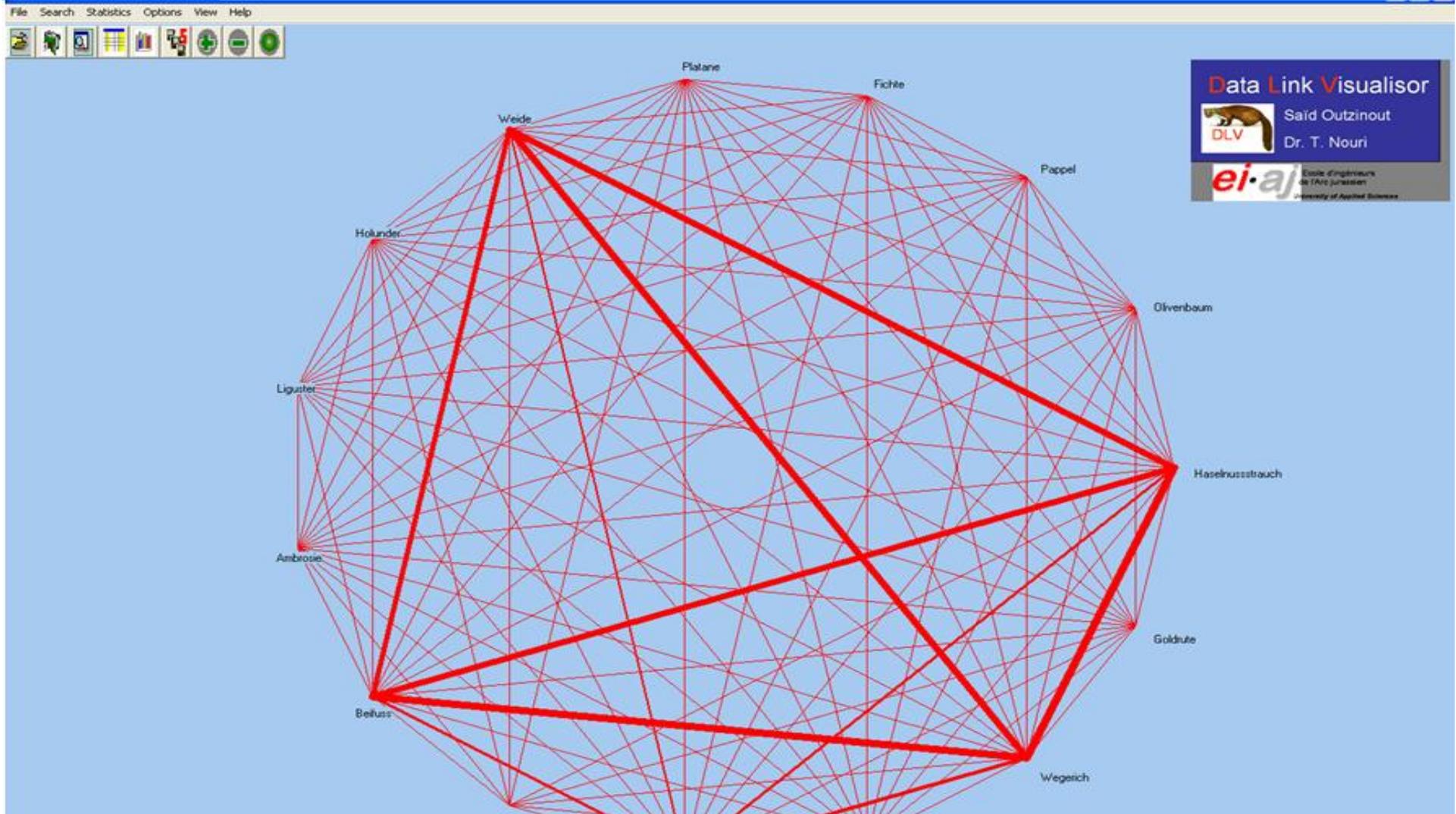


Extraction et visualisation de connaissances biologiques dans des données immunologiques

Prof. Dr. Taoufik Nouri

University of Applied Sciences Northwest Switzerland



Contenu

1. Introduction
2. Méthodologie clinique
3. Data mining au secours des gènes
4. Logiciel de découverte des associations
5. Connaissances découvertes
6. Visualisation des connaissances
7. Conclusion
8. Références



1. Introduction

- Plus de 10% de la population souffrent d'allergies aux animaux, aux plantes, etc.
- Nécessité de faire des vaccins pour chaque allergie.

Ceci:

- Augmente le coût de la santé
- Génère des effets secondaires

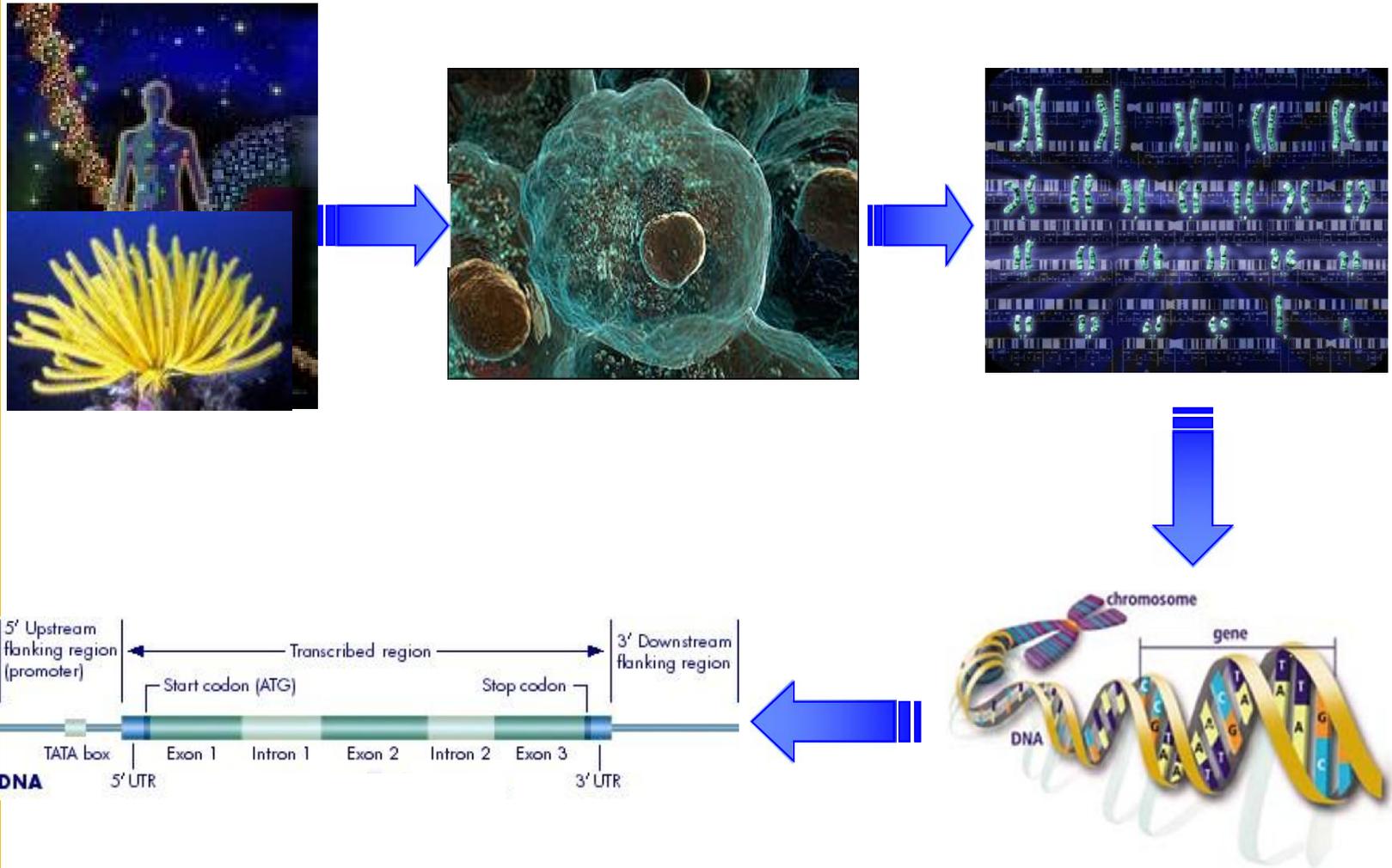
Problème

Comment réduire le nombre de vaccinations contre les anti-gènes pour diminuer les allergies?

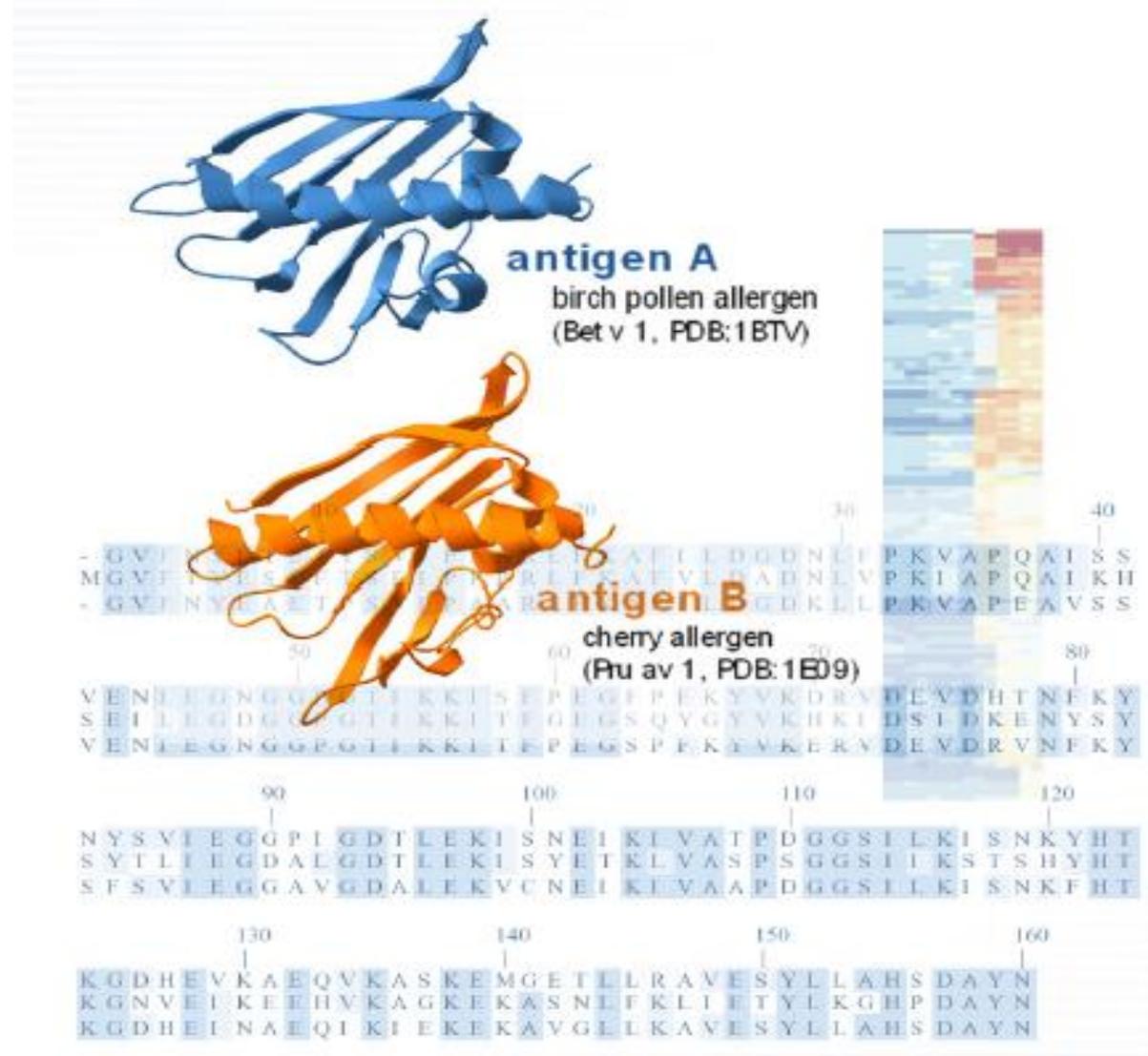
2. Méthodologie clinique

- Comparer les structures au niveau des gènes/ADN de l'agent responsable de l'allergie.
- Si ces structures sont similaires, alors on peut utiliser le même vaccin contre ces agents.
- Ce qui signifie que le même vaccin protège contre un ensemble de gènes.

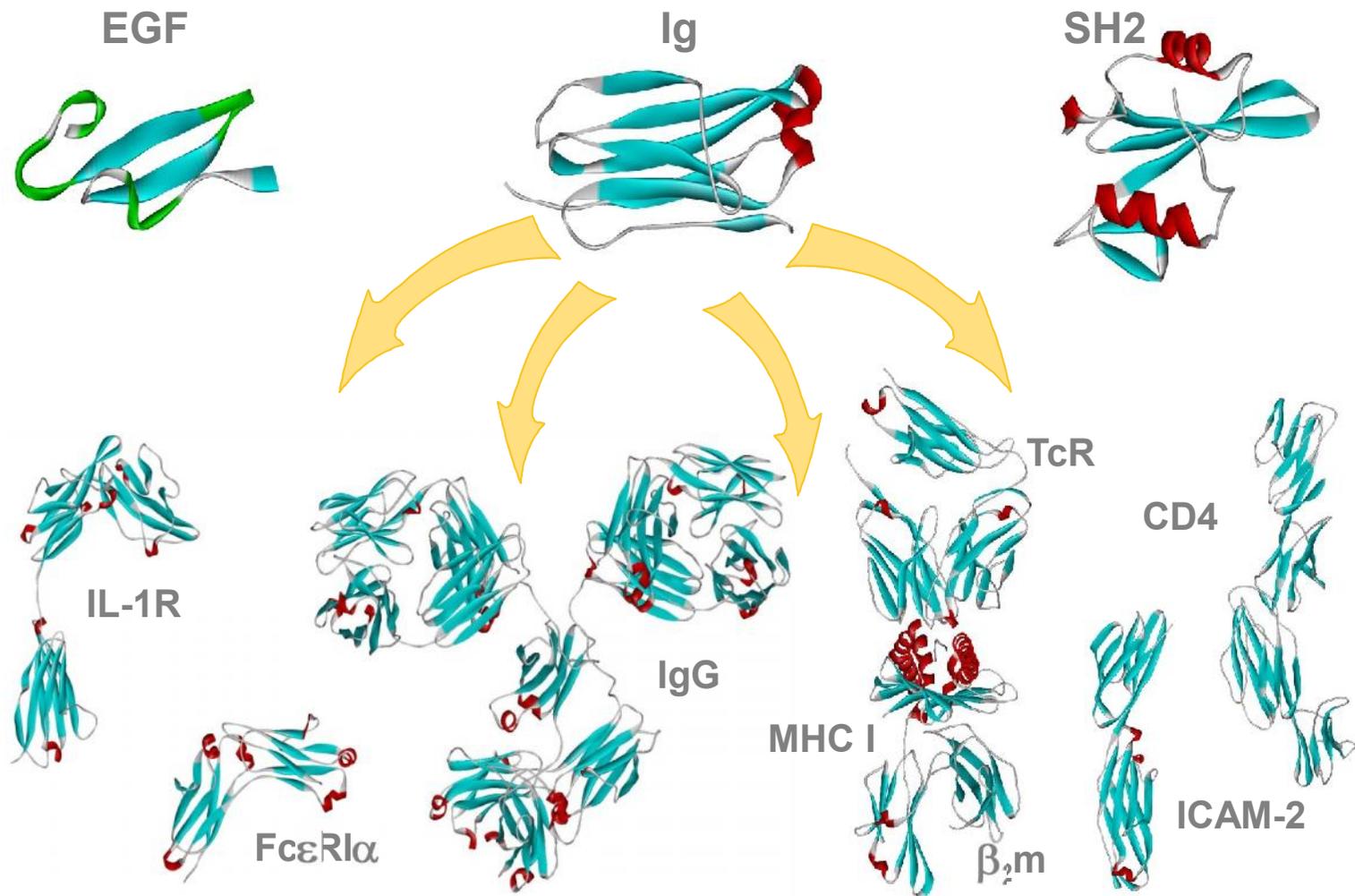
Sequencement de l'AND 1/2



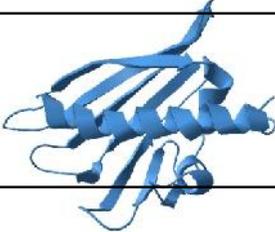
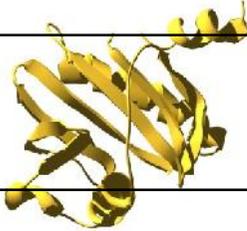
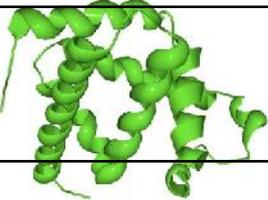
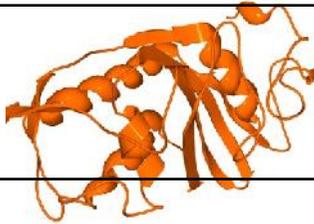
Décodage des gènes



Gènes ayant des structures communes



Exemple de motif d'allergène

<i>Motif No.</i>	<i>Nombre d'allergènes</i>	<i>Structure</i>	<i>Famille de Protéines</i>
AM00001	101		Allergène du pollen de poirier
AM00002	68		Profiline
AM00003	36		Globine
AM00004	35	inconnue	inconnue
AM00005	22		Allergène du venin de guêpe
...

Difficulté

Déterminer la structure ADN de tous les anti-gène (plantes, animaux, insectes, etc.) et les comparer demande **d'énormes ressources!**

3. Data mining au secours des gènes

- Nous avons proposé de collecter des données sur un grand nombre de patients, le type de vaccins qu'ils ont reçus ainsi que les types d'allergie qu'ils ont développés dans leurs vies.
- Ces données portaient sur une période de 10 ans.
- Ces données provenaient de médecins pédiatres.

Hypothèse

- Nous cherchons à savoir si un vaccin contre un anti-gène spécifique a protégé l'enfant (patient) vacciné contre d'autres anti-gènes.
- Donc s'il a empêché l'apparition d'autres allergies.
- Comment démontrer cette hypothèse sur la base des données reçues des pédiatres?

Proposition

- Nous avons développé un logiciel utilisant la technique des règles d'associations sur la base de l'algorithme Apriori.

Principe des règles d'associations

Soit une règle J de la forme:

$$J: \quad LS \quad \longrightarrow \quad RS$$

Nous définissons le support et la confiance comme suit:

$$\text{supp}(J) := \frac{| T \in D \mid J \subseteq T |}{k}$$

$$\text{conf}(R) := \frac{\text{Supp} (LS \cup RS)}{\text{Supp} (LS)}$$

Avec: T le nombre de transactions contenant la règle J et k le nombre de transactions total dans la base de données

4. Logiciel de découverte des associations

Data Link Visualiser

File Search Statistics View Help

DATA

Select All

<input type="checkbox"/> idTest_5110_Res	<input type="checkbox"/> r_t2_classe	<input type="checkbox"/> r_Moisissures1_classe
<input type="checkbox"/> idPatient	<input type="checkbox"/> r_t3_classe	<input type="checkbox"/> r_Moisissures2_classe
<input type="checkbox"/> Sexe	<input type="checkbox"/> r_t4_classe	<input type="checkbox"/> r_e1_classe
<input type="checkbox"/> Adult	<input type="checkbox"/> r_t5_classe	<input type="checkbox"/> r_e5_classe
<input type="checkbox"/> r_a_lgE_classe	<input type="checkbox"/> r_t7_classe	<input type="checkbox"/> r_e2_classe
<input type="checkbox"/> r_g3_classe	<input type="checkbox"/> r_t11_classe	<input checked="" type="checkbox"/> r_e3_classe
<input type="checkbox"/> r_g12_classe	<input type="checkbox"/> r_t14_classe	<input type="checkbox"/> r_e4_classe
<input checked="" type="checkbox"/> r_gx3_classe	<input type="checkbox"/> r_Arbres1_classe	<input type="checkbox"/> r_e6_classe
<input type="checkbox"/> r_Graminees1_classe	<input type="checkbox"/> r_Arbres2_classe	<input checked="" type="checkbox"/> r_e70_classe
<input type="checkbox"/> r_Graminees2_classe	<input type="checkbox"/> r_d1_classe	<input type="checkbox"/> r_e77_classe
<input type="checkbox"/> r_Graminees2_id	<input checked="" type="checkbox"/> r_d2_classe	<input type="checkbox"/> r_e78_classe
<input type="checkbox"/> r_wx3_classe	<input type="checkbox"/> r_d70_classe	<input type="checkbox"/> r_e79_classe
<input checked="" type="checkbox"/> r_w6_classe	<input type="checkbox"/> r_d71_classe	<input checked="" type="checkbox"/> r_e82_classe
<input type="checkbox"/> r_w7_classe	<input checked="" type="checkbox"/> r_d73_classe	<input type="checkbox"/> r_e84_classe
<input checked="" type="checkbox"/> r_w8_classe	<input checked="" type="checkbox"/> r_h1_classe	<input type="checkbox"/> r_e85_classe
<input type="checkbox"/> r_w9_classe	<input checked="" type="checkbox"/> r_h4_classe	<input type="checkbox"/> r_e86_classe
<input checked="" type="checkbox"/> r_w10_classe	<input checked="" type="checkbox"/> r_k73_classe	<input type="checkbox"/> r_e88_classe
<input type="checkbox"/> r_w12_classe	<input type="checkbox"/> r_Acariens1_classe	<input checked="" type="checkbox"/> r_e201_classe
<input type="checkbox"/> r_w19_classe	<input type="checkbox"/> r_Acariens2_classe	<input type="checkbox"/> r_ex71_classe
<input type="checkbox"/> r_Plantes1_classe	<input type="checkbox"/> r_mx1_classe	<input type="checkbox"/> r_Animaux1_classe
<input type="checkbox"/> r_Plantes2_classe	<input type="checkbox"/> r_m1_classe	<input type="checkbox"/> r_Animaux2_classe
<input type="checkbox"/> r_tx5_classe	<input type="checkbox"/> r_m2_classe	<input type="checkbox"/> r_fx1_classe
<input type="checkbox"/> r_tx6_classe	<input type="checkbox"/> r_m3_classe	<input type="checkbox"/> r_fx2_classe
<input type="checkbox"/> r_tx9_classe	<input type="checkbox"/> r_m4_classe	<input type="checkbox"/> r_fx3_classe
<input type="checkbox"/> r_t1_classe	<input type="checkbox"/> r_m6_classe	<input type="checkbox"/> r_fx5_classe

Minimum Support **70**

Minimum Confidence **80**

Search State

Ignored State

Level **2**

Data View

Rules Generation

Only Visualisation

Visualization

Calculate

TestID	idPatient	Katze	Pferd	Ziege	Hund	Hamster	Feder	Rind	Ratte	Maus	pilz II	Roggen	Getreide	Erle	Birke	Eiche	Zypresse	Ahorn	Esche	Pappel
52	7085	0	0	-1	0	-1	0	-1	-1	-1	-1	0	-1	-1	0	0	-1	-1	0	-1
53	6804	0	-1	0	-1	-1	0	-1	-1	-1	-1	1	-1	-1	0	-1	-1	-1	-1	-1
54	1629	1	0	-1	0	-1	0	-1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1
55	8347	0	-1	-1	0	-1	0	-1	-1	-1	-1	0	-1	0	0	0	-1	0	0	-1
56	1064	1	1	-1	0	-1	0	-1	-1	-1	-1	0	-1	0	0	0	-1	-1	-1	-1
57	7212	0	-1	-1	0	-1	0	-1	-1	-1	-1	0	-1	0	0	0	-1	0	0	-1
58	6601	0	0	-1	0	-1	0	-1	-1	-1	0	1	-1	-1	-1	-1	-1	-1	-1	-1
59	7810	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	0	-1	-1	-1	0	-1
60	7689	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	0	0	0	-1	1	1	-1
61	6822	0	0	-1	0	-1	0	-1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1
62	8363	-1	0	-1	-1	-1	0	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1
63	7407	0	0	-1	0	-1	0	-1	-1	-1	-1	1	-1	-1	0	0	-1	-1	0	-1
64	3563	1	1	-1	1	-1	0	-1	-1	-1	0	1	-1	0	0	0	-1	1	1	-1

- Les données sont complètement anonymes.
- Pas de connaissance préalable des données
- Chaque enregistrement contient vers les 700 attributs
- Les données représentent quelques dizaines de milliers de tests
- Nous n'avons aucune influence sur les résultats

5. Connaissances découvertes 1/2

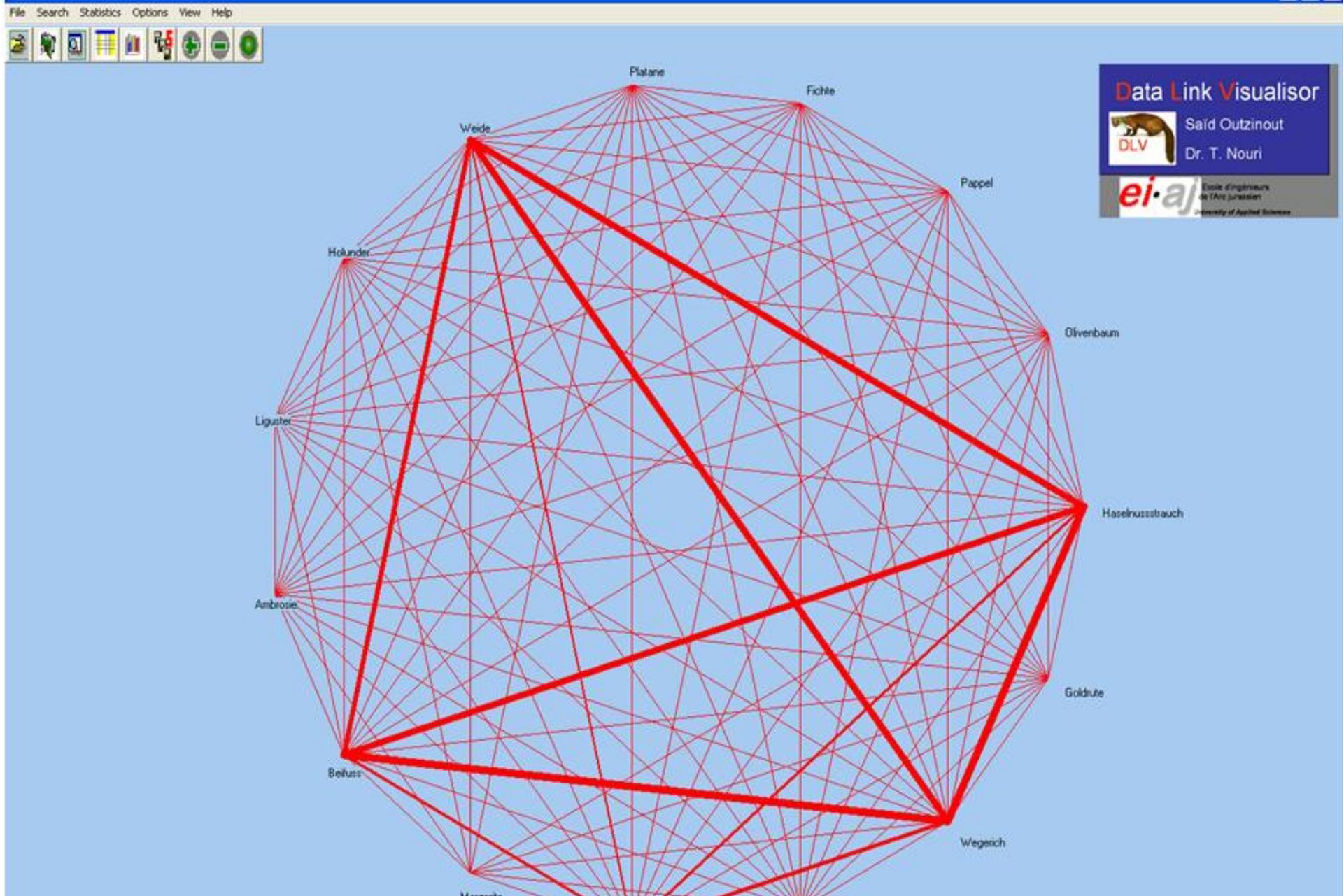
- Nous avons mis en évidence quelques dizaines de groupe d'anti-gènes (doublets, triplets, quadruplets, etc.). Les règles découvertes sont de la forme: $A_1, A_2, \dots, A_n \rightarrow c_1, c_2, c_2, \dots, c_n$
- Ces groupes d'anti-gènes sont susceptibles d'être traiter par le même vaccin.
- La structure DNA de ces groupes d'anti-gènes devait être similaire.
- Quelques-uns de ces groupes sont connus par les médecins (ils ont été découverts par hasard), notre étude les confirme.
- Une étude comparative dans des DB-biologiques devait confirmer nos découvertes puis une étude **clinique**.

5. Connaissances découvertes 2/2

Prémisse ==>	Conséquent	Support	Confidence
Noisetier	Olivier	46.1	72.2
Olivier	Noisetier	46.1	72.2
Noisetier	Peuplier	57.5	47.3
Peuplier	Noisetier	36.3	75
Noisetier	Sapin	52.2	30.4
Sapin	Noisetier	21.5	73.6
Noisetier	Platane	40	41.1

Celui qui a été vacciné contre l'olivier, peut être protégé contre le noisetier avec un support de 66.1% et une confiance de 72.2%

6. Visualisation des connaissances 1/2



6. Visualisation des connaissances 2/2

Sur la base du diagramme de liaisons précédent, nous pouvons déduire visuellement et de façon rapide les groupes de plantes ou de produits contenant des gènes ayant des structures ADN similaires.

7. Conclusion

- Le fait d'avoir des liens forts dans une règle (support et confiance) n'est qu'un indicateur sur la structure ADN des anti-gène responsables de l'allergie.
- Maintenant nous pouvons faire des comparaisons **ciblées** de structures des ensembles de gènes trouvées par notre étude dans des bases de données biologiques contenant ces structures.
- Par la suite, des études cliniques doivent confirmer l'effet du vaccin sur le groupe de gènes découvert par cette étude.
- Ainsi nous pourrions identifier définitivement les groupes d'anti-gènes et les bloquer par un seul vaccin.

8. Références

- <http://www.ensembl.org>
- <http://www.expasy.org>
- <http://www.ncbi.nlm.nih.gov>
- <http://www.ebi.ac.uk>
- <http://www.rcsb.org/pdb>