



Centre National de Référence  
**CytoMégaloVirus**  
CNR-CMV



Instituts  
thématiques



**Inserm**





Institut national  
de la santé et de la recherche médicale

## Rapid genotyping of cytomegalovirus enveloppe glycoproteins from toddler's saliva.

Grosjean J., Hantz S., Cotin S., Marin B., Pasquier C., Denis F., Alain S.

EA3175, INSERM  
Faculté de Médecine  
2 Av du Dr. Marcland  
87000 Limoges

CNR CytomégaloVirus – CHU de Limoges  
2 Av. M.L. King  
87042 Limoges Cedex



Mesdames, Messieurs les membres du jury j'ai l'honneur de vous présenter le travail de recherche que j'ai effectué, sous la responsabilité du Pr Alain, au sein de la JE – Inserm dirigé par le Pr. M.C. Ploy à la faculté de médecine de Limoges. Ce travail porte sur le typage de protéines d'enveloppe du CytomégaloVirus Humain.

15 secondes

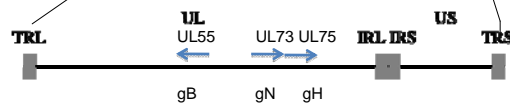
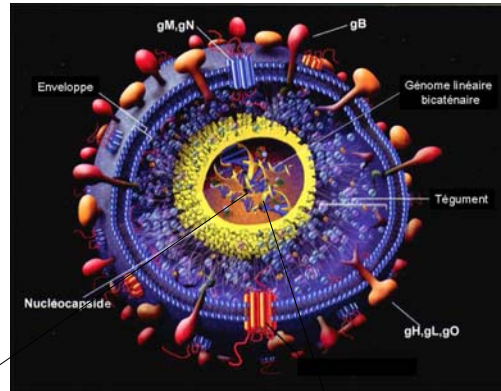
## CMV enveloppe glycoproteins :

Involved in viral entry and propagation in infected cells  
Major targets for neutralizing antibodies (gB, gH)  
⇒ Components of future vaccines

3 functional complexes gCI, gCII, gCIII

Genetic polymorphism used for classification of CMV strains in genotypes

Clinical impact of genotypes ?



Reddehase 2006

### Why choosing the envelope glycoproteins?

First because they are involved in the various patterns of viral entry and propagation. through ... and organized in three complexes GCI gB implicated in entry and propagation GCII gH gM implicated in attachment and GCIII (GH, GO, gL involved in and they are major targets ... which makes them a natural component of future vaccines. Second, because there is an existing classification, though the clinical impact of genotypes remains uncertain.

We choose to work first on the major representant of each group, gB, GH and gN

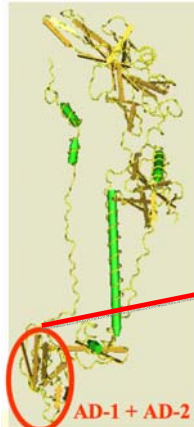
Ces glycoprotéines gB, gH et gN sont portées par l'enveloppe virale. Elles sont codées par les gènes UL55, UL75 et UL73. Ces gènes sont situés sur le fragment long du génome, UL55 sur le brin antisens, UL73 et UL75 sur le brin sens.

Ces trois glycoprotéines induisent des anticorps neutralisants toutefois gB et dans une moindre mesure gH sont les cibles majeures.

Elles sont indispensables à la pénétration du virus dans la cellule. De plus elles permettraient au virus d'avoir un tropisme pour les cellules épithéliales et fibroblastiques et pour les cellules souches hématopoïétiques

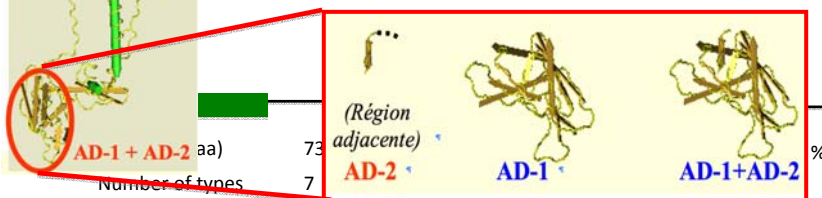
## Variability of envelope glycoproteins :

- Glycoprotein B : homodimer , 2 subunits



- 2 epitopes AD1 and AD2

- 3D structure : AD2 close to AD1  
(personal computerized modelization from G Champier)



(Chou et al., 1992, Pignatelli et al., 2004)

Glycoprotein B gB is polyprotein cleaved at position 460 within a cleavage zone CLZ. gB variability results both from high variability (60%) of two loci : the cleavage zone and the N terminal region, and from possible recombinations between these two loci 1 2 3 4 en N Term correspondent aux séquences associées aux génotypes de Chou définis en CLZ. Il existe une région variable en C-term (2 génotypes définis mais peu variable et pas d'épitopes). Nous ne l'avons pas étudiée. La gB comporte deux zones de variabilité majeure en vert sur le schéma. La première variable à 63% est située sur la partie N terminale de la protéine. La seconde variable à 66 % est situé sur une zone de clivage post traductionnel. Nous la nommerons CLZ pour Cleavage Zone.

Les épitopes des anticorps neutralisant sont indiqués en rouge sur le schéma. L'épitope AD-2, donne d'ailleurs son nom à la zone variable adjacente. L'épitope AD1 hautement conservé, suit immédiatement la zone de variabilité CLZ.

A partir des séquences nucléotidiques correspondants 6 types sont décrits sur la zone AD2 et 7 sur la zone de Clivage. Bien entendu ces types ne sont pas indépendants et majoritairement le même type est trouvé sur les deux zones. Toutefois des phénomène de recombinaison ont été décrit permettant la liaison d'un type AD2 avec un autre type CLZ. Ces phénomènes ont été

## Actual methods for genotyping :

- **Methods**

- Simple RFLP (1) or sequencing (2)
- Not very sensitive
- Can't be use directly on large-scale epidemiologic studies

- **Classification**

- ? Correlation ? with
- Clinical impact
- Immune response

(1) Chou et al. 1992

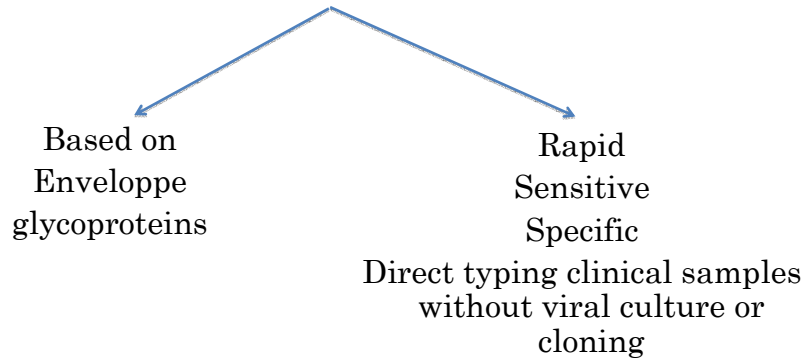
(2) Meyer-könig 1999, Pignatelli et al. 2003

Actuellement les génotypes sont identifiés, soit par étude du profil de restriction, soit par séquençage. Ces techniques de référence, bien qu'étant à l'origine de la nomenclature actuelle sont peu sensibles. Elles ont été définis à partir de souches virales et sont difficilement applicables aux échantillons cliniques, ce qui ne permet pas de faire de larges études de populations.

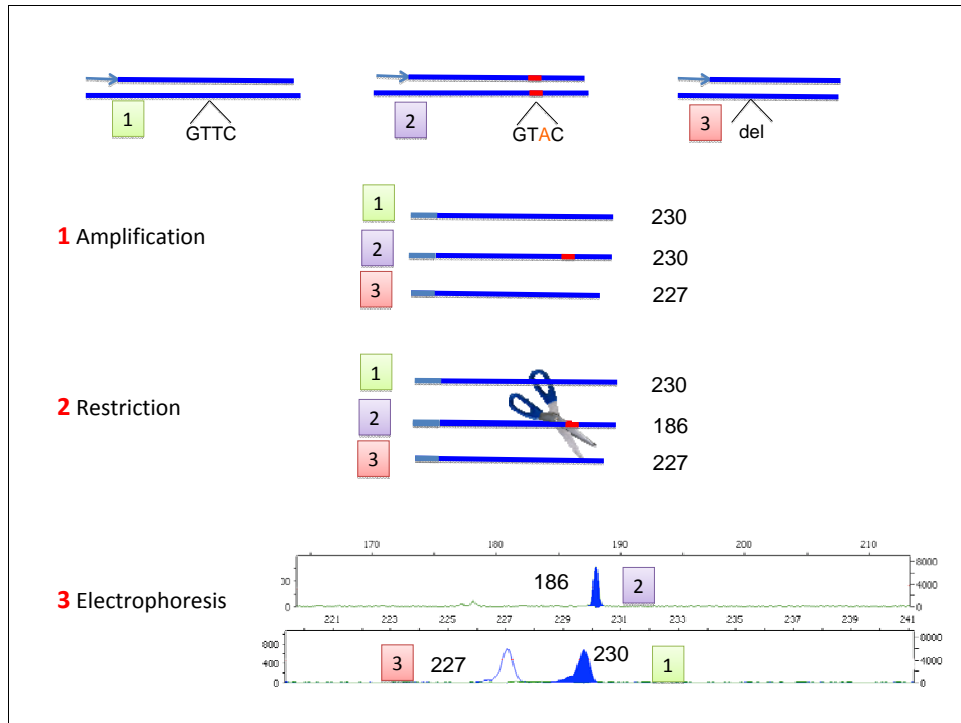
La nomenclature actuelle ne montre pas de corrélation entre le génotype des glycoprotéines d'enveloppe et un éventuel pouvoir pathogène des souches ou une variation de la RI.

## AIM of the study :

To develop an original technique  
for CMV genotyping



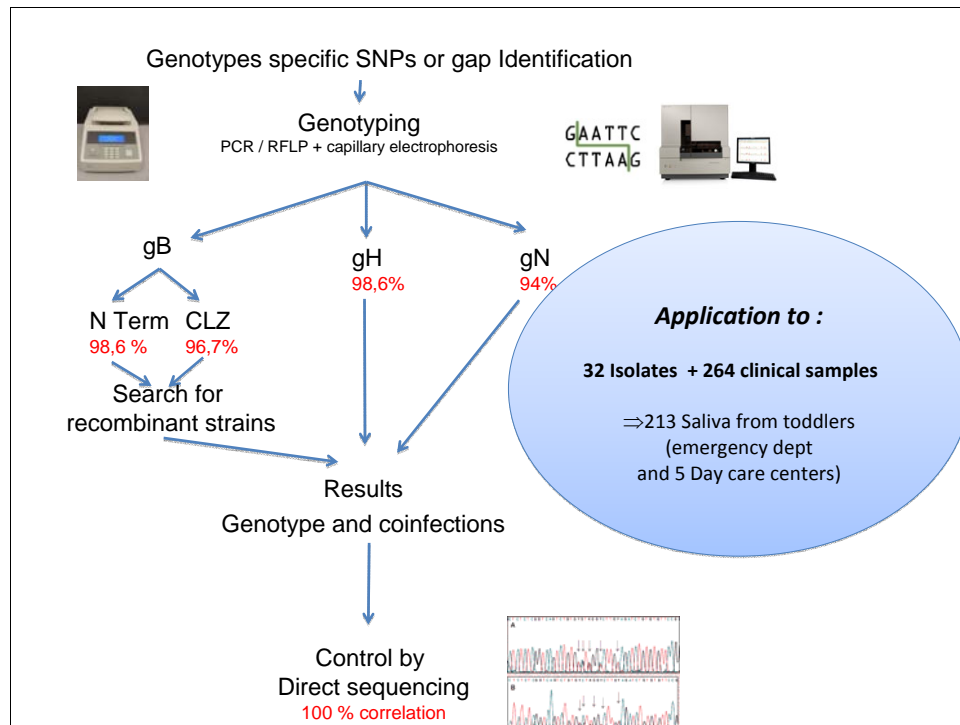
Ainsi l'objectif principal de notre travail a été de mettre au point et de valider une méthode de typages des souches de CMV rapide, sensible et spécifique et ceci à partir des gènes codant les glycoprotéines d'enveloppe.



En voici le principe. Les séquences se différencient soit par une mutation ponctuelle située dans un site de restriction (1 et 2) soit par une délétion comme la séquence 3. Ces séquences sont amplifiées avec des amorces consensus dont l'une est marquée. L'amplicon 3 est directement identifiable par sa taille. Les amplicons 1 et 2 seront différenciés par la taille des fragments obtenus après restriction enzymatique.

La séparation des fragments en électrophorèse capillaire permet de les différencier à une base près.

Voici le résultat : Pour chaque fragment on observe un pic à la taille attendue, en cas de co-infections on observe deux pics correspondant aux deux séquences.



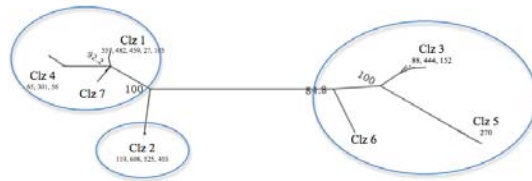
Allows genotyping of more than 90% of routinely positive samples. l'aide de mutations et de délétions spécifique de type que nous avons identifiés nous effectuons un typage en gB, gH et gN. Sur le gène de la glycoprotéine B nous étudions les zones AD2 et CLZ de manière concomitante, ce qui nous permet de rechercher d'éventuels recombinants.

Les résultats obtenus nous permettent de détecter d'éventuels co-infections et sont vérifiés en partie par séquençage afin de tester la spécificité de notre technique de typage.

## gB genotypes

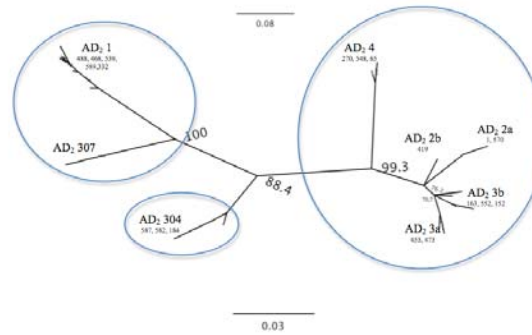
### ○ CLZ (gBcls)

- 1 to 5/7 genotypes found



### ○ N Term (gBn)

- 6 genotypes found
- 2 new genotypes
  - 3b : No 300T>C mutation
  - 2b : 298 bp



Les résultats sur l'ensemble des échantillons étudiés sont les suivants :

Pour la zone CLZ seul les types 1 à 5 ont été retrouvés.

Des séquences ont été effectuées et un arbre phylogénique a été tracé. Comme on le voit sur cette arbre il serait peut être plus judicieux de distinguer d'un point de vue phylogénique 3 types subdivisibles, que 7 types de niveau équivalent.

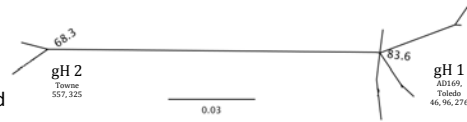
Pour la zone AD2 les 6 types préalablement décrits ont été retrouvés. De la même manière la construction d'un arbre nous permet de penser qu'il existe plutôt 3 types principaux subdivisibles.

Toutefois nous individualisons deux nouveaux types et ceci à la fois en RFLP et de manière phylogénique. Ce sont le type 3b caractérisé par l'absence d'une mutation de restriction, et le type 2b par une délétion supplémentaire faisant varier la taille de l'amplicon.

## gN AND gH GENOTYPES

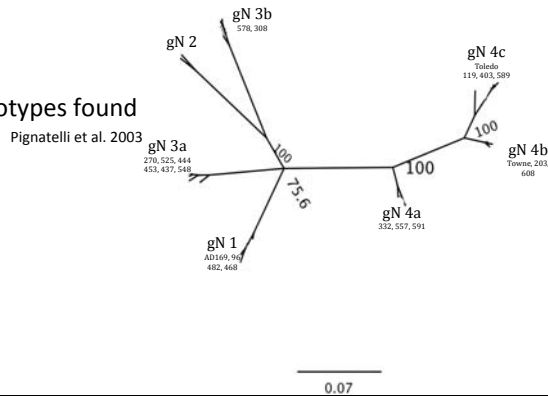
### ○ UL75 (gH)

- 2 genotypes found
- But type 1 is probably composed by 3 sub-types.



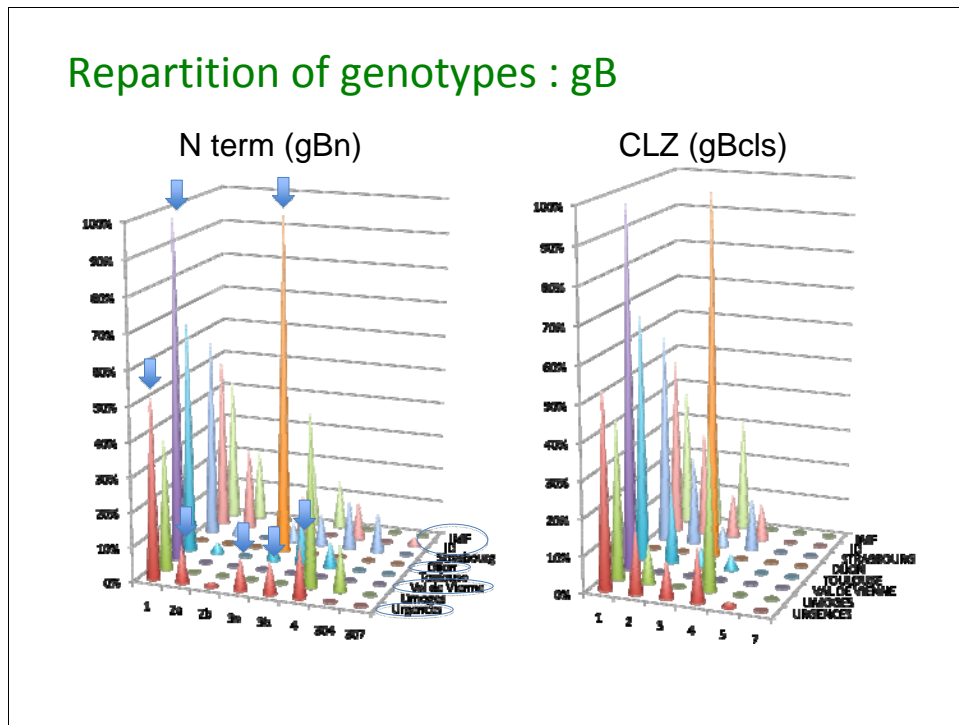
### ○ UL73 (gN)

- All Pignatelli's genotypes found



En UL75 gH les 2 types préalablement décrits ont été retrouvés. Toutefois il paraît évident que le type 1 peut être divisible en sous types ce que personne n'a encore fait dans la littérature. En gN les résultats que nous trouvons sont similaires à ceux publiés en 2003 par Sarah Pignatelli et collaborateur.

## Repartition of genotypes : gB



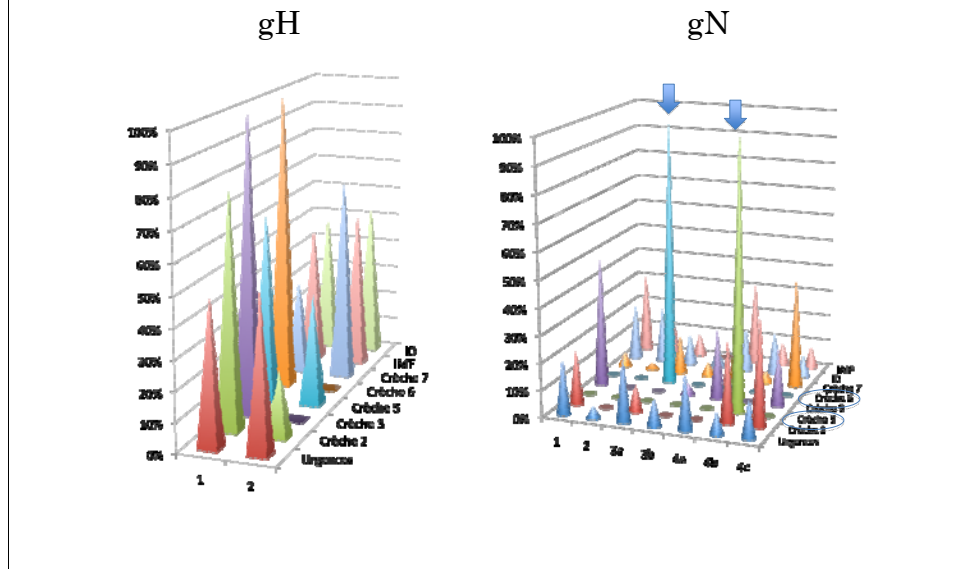
Major types are present At the emergency department which could better represent the general population, Type 1 is predominant, Type 4 represents 15 %, which is higher than previously described. Types (subtypes) 2a and 3a and B are coming after , while 5 is not negligible in CLZ. In day care centers, the center effect is dominant but other genotypes can be found. Ce graphique montre la prévalence de chaque génotype de la glycoprotéine B dans nos différents centres et ceci en comparaison avec deux autres populations, IMF et ID.

Sur la zone AD2 chez les enfants des urgences, population se rapprochant le plus de la population générale on constate que d'une manière globale le type 1 est prédominant, les types 2b, 304 et 307 sont plus rares.

En crèche l'effet centre est majeur. Dans certaines crèches on ne retrouve qu'un seul type et donc très probablement tous les enfants s'y sont contaminés avec la même souche. En effet ces résultats d'effet centre trouvés en AD2 ont été confirmé en CLZ, gH et gN.

Comme attendu la distribution en CLZ varie peu par rapport à la distribution en AD2, ce qui peut paraître logique car ces deux zones sont situés sur le même gène et ne sont pas indépendantes. Les génotypes 6 et 7 sont inexistantes, le 5 est rarissime.

## Repartition of genotypes : gH and gN

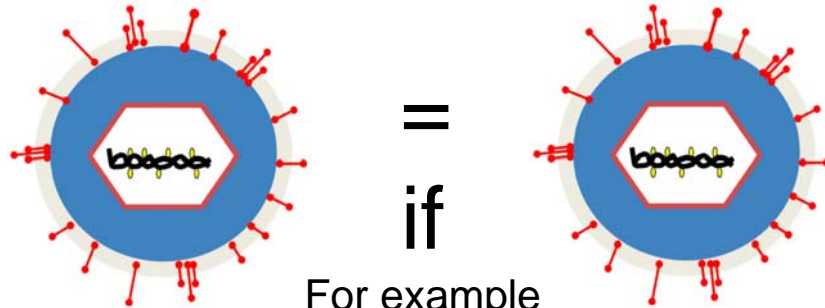


In gH repartition between Gt 1 and 2 is homogenous in all populations

.In gN repartition is also homogenous with exception for two day care centers which seems to harbour only one type of strain  
 En UL75 la répartition entre type 1 et type 2 est homogène, à la fois entre ces deux types et entre les centres.

En UL73, si l'on enlève l'effet centre des crèches 3 et 6 la répartition est conforme a celle décrite par Sarah Pignatelli en Europe, c'est à dire avec une prédominance des types 4.

## Application to circulation of CMV strains

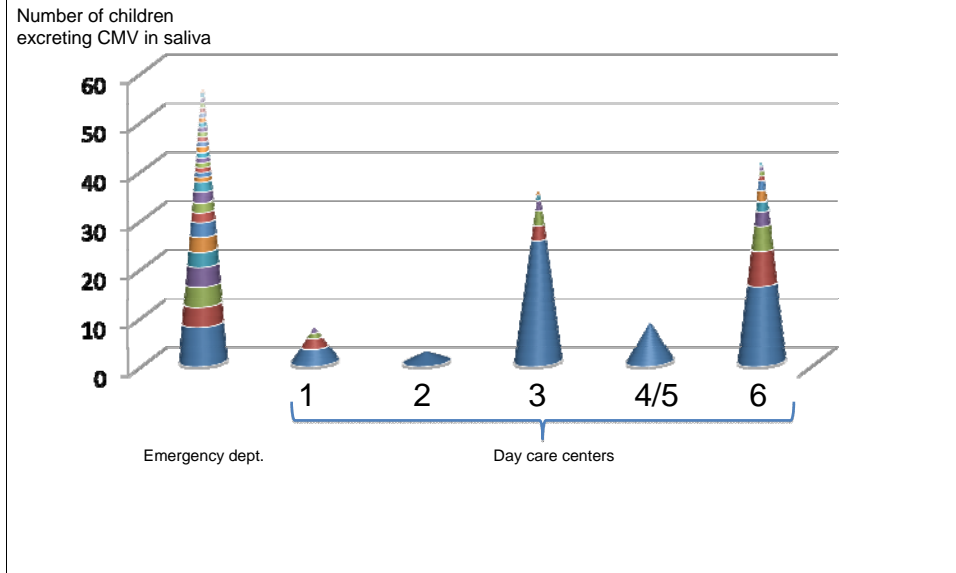


Type	Sample 1	Sample 2
gB N-Term	1	1
gB CLZ	1	1
gH	2	2
gN	3a	3a

By grouping results from Gb Gh and GN we can define prototypic strains as strains with identical gen

La synthèse des génotypes pour les trois glycoprotéines nous a permis de définir la notion de souche. Cette notion est défini à *minima*. Ainsi dans notre étude deux échantillons possèdent la même souche si pour un échantillon simple, c'est à dire sans coinfections on retrouve les 4 mêmes génotypes entre les deux échantillons pour nos 4 zones de typage.

## Clonality in DCCs versus emergency department



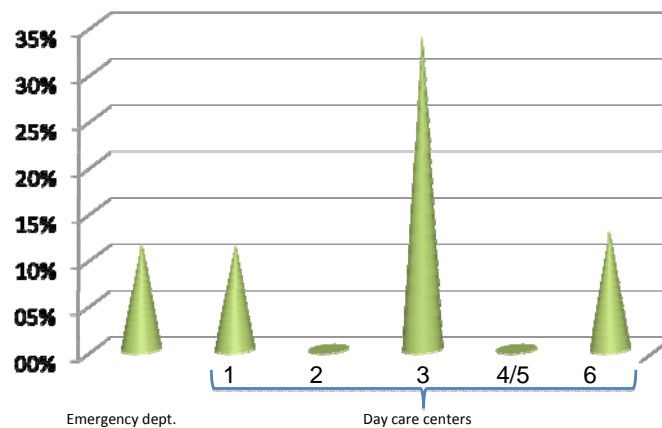
Cette notion de souche nous permet de définir la clonalité de chaque centre. Ainsi au urgences sur 52 échantillons simples, on a retrouvé 30 souches dont la majoritaire n'infectait que 8 enfants.

En ce qui concerne la crèche du CHU de Limoges il n'y a pas d'effet centre véritable. En effet on ne retrouve pas une clonalité supérieure à celle des urgences voisines.

En crèche notamment à Toulouse, on a retrouvé que 6 souches et la majoritaire infectait 25 enfants sur 31. A Strasbourg crèche ou la prévalence est sensiblement identique, seulement la moitié des enfants excrétaient la même souche. A Dijon et à Bosmie l'aiguille tous les enfants possédaient la même souche.

## Coinfections

- From 0 to 33% coinfections in toddlers

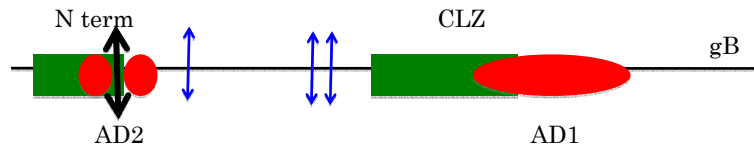


This technique also allows us to identify coinfection with a higher accuracy.

Comme nous l'avons vu notre technique permet de détecter facilement les co-infections, à condition que pour un échantillon donné sur au moins une zone il y ait deux types différents.

Dans notre étude la proportion d'enfants coinfectés variée suivant les centre. Le centre de Toulouse possédaient le plus haut taux de co-infections, soit 33%.

## gB N term/CLZ RECOMBINATION



- 90 % of prototypic sequences
- 10% recombinant sequences
  - Recombinations are always the same
  - Distribution is homogeneous
  - Recombination zone
    - Meyer- König : Between AD2(N Term) and CLZ
    - Our work : Between AD2 epitopes

N-Term	CLZ
3b	2
2b	4
4	5
304	2
307	3

Nous pouvons ainsi imaginer qu'il existerait une ou plusieurs zone de recombinaison majeure à l'intérieur de la zone AD2 et ceci bien en amont de celles décrites par meyer konig (en bleu).

## Conclusion

- **A new method :**
  - Sensitive, low cost, adapted to large scale epidemiologic studies
  - Detects clonality, coinfections, and recombinant strains with a higher accuracy
- **With its limits :**
  - Genotyping based on SNPs
  - Detection of clonality and coinfections from single separate PCRs.

Nous avons développé une méthode sensible permettant le screening direct d'un grand nombre d'échantillons cliniques. Cette méthode est applicable à de larges études de population

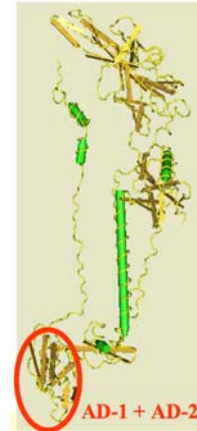
Il s'agit d'une méthode originale à la fois par sa conception et par son caractère discriminant qui permet en très peu de temps de typer 4 régions variables des glycoprotéines d'enveloppe contrairement aux autres techniques ne s'adressant qu'à une région. Cette technique associée au séquençage nous permet une nouvelle nomenclature.

## Perspectives ?

- An extended classification ?
  - Reanalyse clinical outcome with modified gB genotypes classification ?

- Impact of N Term AD1/AD2 recombination ?

- Natural immune response
- Vaccinal immune response



Avec les modifications de nomenclature que nous proposons, intégrant le Reclassement des types gB en fonction de la proximité phylogénique et de la notion de recombinants il serait important de réanalyser les données des différentes études déjà publiées. En effet la plupart des études concluent qu'il n'existe pas de corrélation entre typage et capacité supplémentaire du virus tant au niveau du pouvoir pathogène de la souche que de la réponse au traitement antiviral.

Toutefois l'absence de corrélation ne pourrait elle pas être du à l'inadéquation de la nomenclature utilisée ?

Il nous faudra aussi étudier l'impact qu'a la recombinaison sur la variabilité des épitopes et donc sur la réponse immunitaire et ceci que ce soit au niveau de la réponse naturelle de l'hôte que de la réponse vaccinale.