

Apports et promesses du séquençage génomique en microbiologie clinique.

Philippe Glaser – Institut Pasteur
pglaser@pasteur.fr

Microbiologie clinique – séquençage génomique

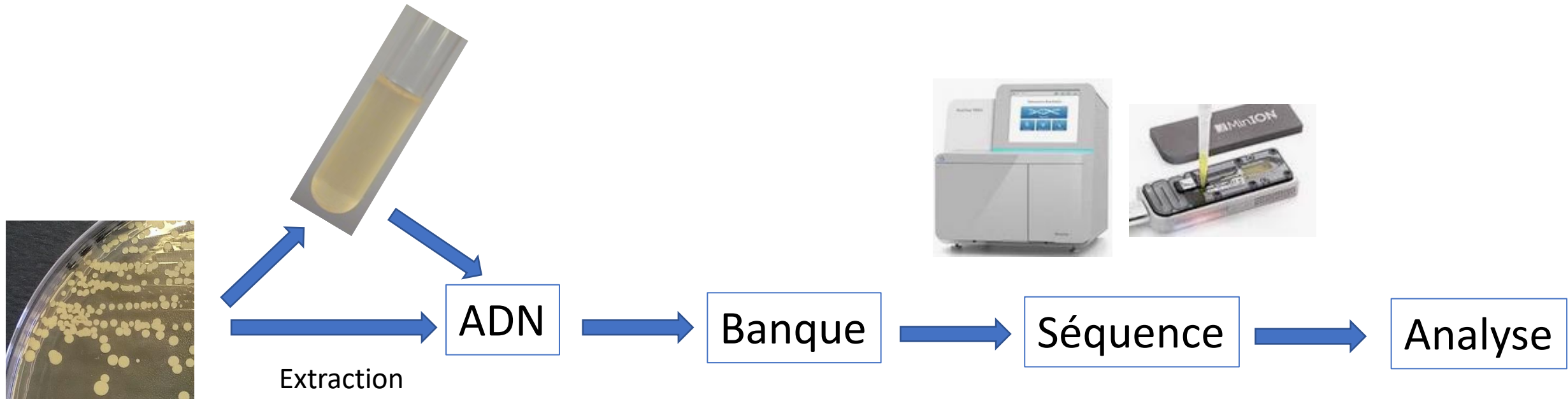
Microbiologie clinique (Hôpital)

- Identification bactérienne
- Antibiogramme
- Epidémiologie moléculaire, analyse de transmission

Séquençage génomique

- Information génomique complète
 - Identification de l'espèce
 - Identification des gènes de résistance:
 - Génotype – phénotype
 - Identification de clones à risque
- Comparaison entre isolats
 - Identification de transmission et de la source

Procédure de séquençage



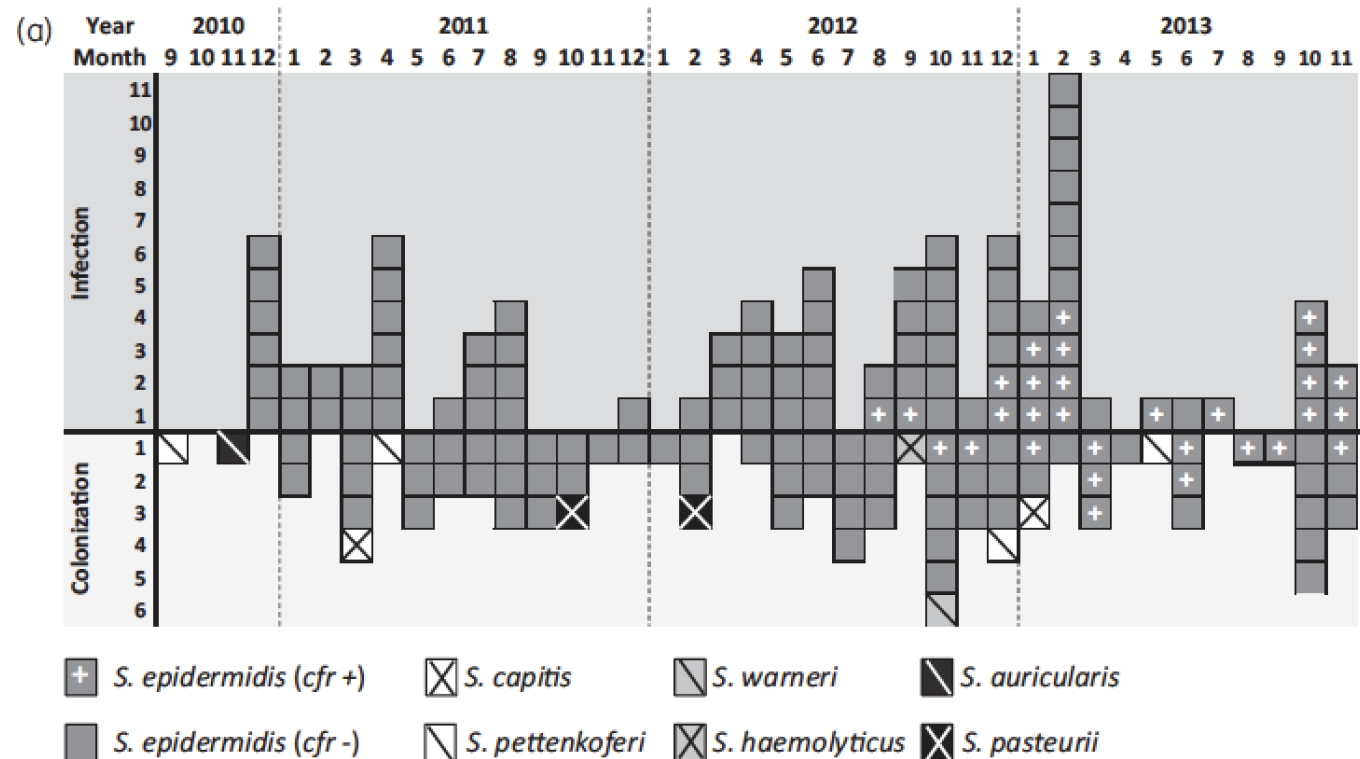
Défis

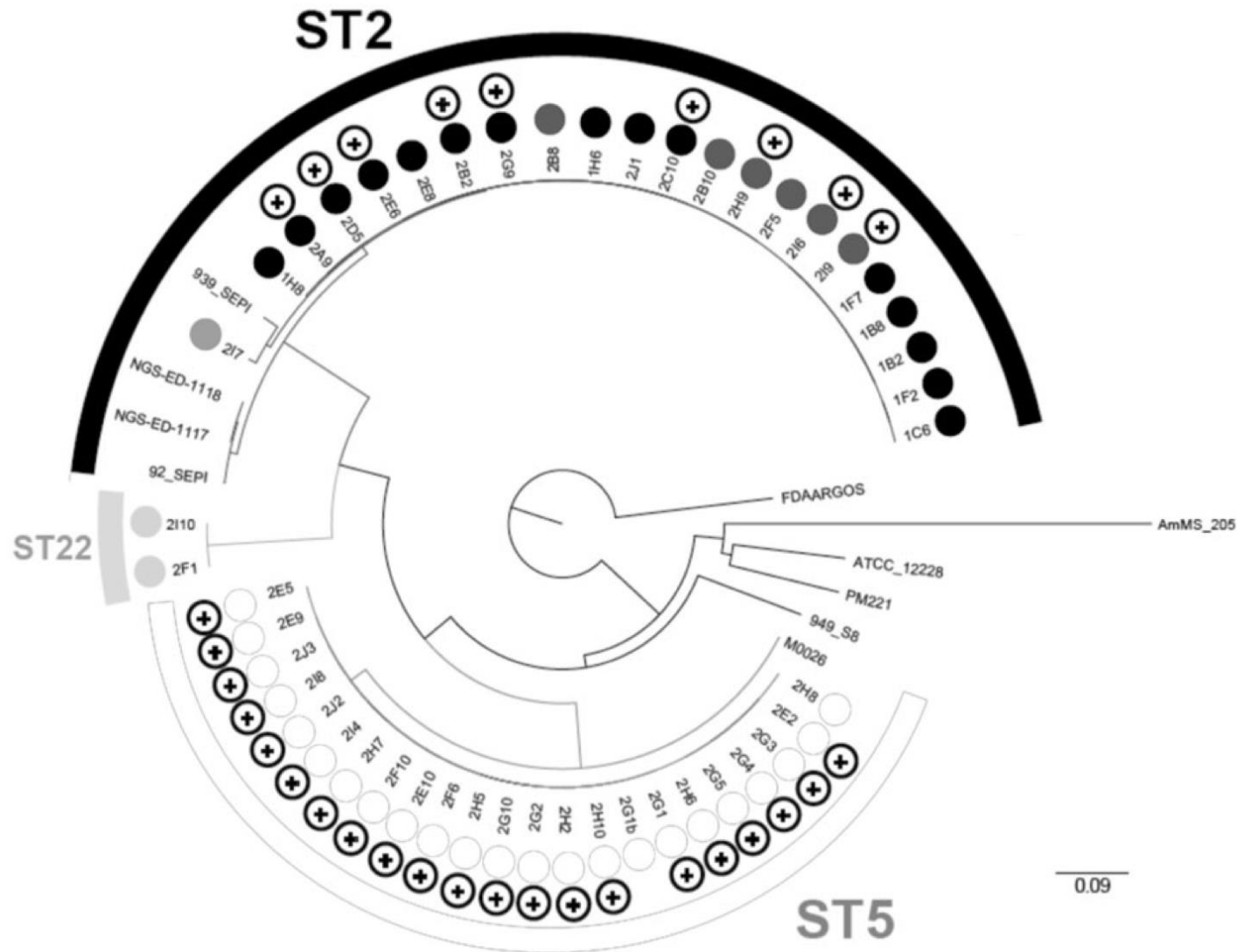
- Temps pour avoir les résultats (> 2-3 jours)
 - Coût: il diminue avec le nombre d'échantillons
- ⇒ Incompatibilité entre temps et coût

Analyse d'une épidémie, des transmissions, de la source

- Cas 1: *Staphylococcus epidermidis*

⇒ Augmentation du nombre de cas d'infection à *S. epidermidis* résistant au Linezolide dans les service de greffe





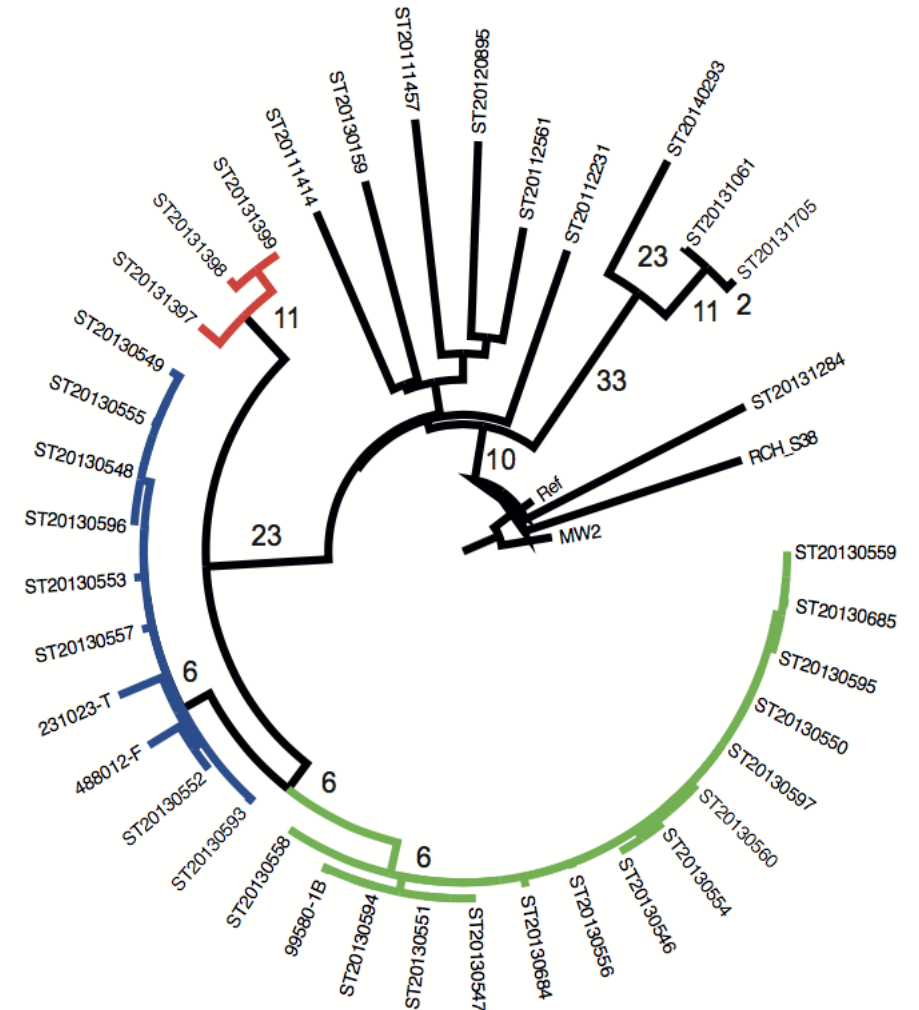
- Séquençage de 47 souches
- ⇒ 3 Clusters (pas une épidémie, trois)
- ⇒ Moins de 10 mutations entre deux isolats
- ⇒ Correspond à des clones à haut risque MDR déjà décrits
- ⇒ Deux plasmide *cfr* (haut niveau de résistance) ont été transférés

Analyse d'une épidémie, de la source d'une contamination

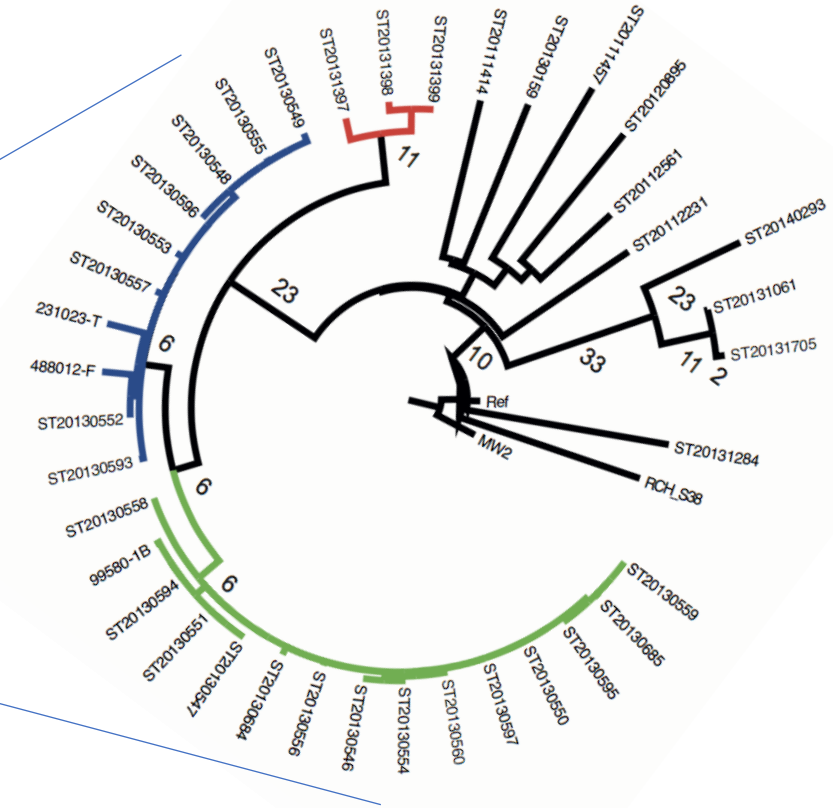
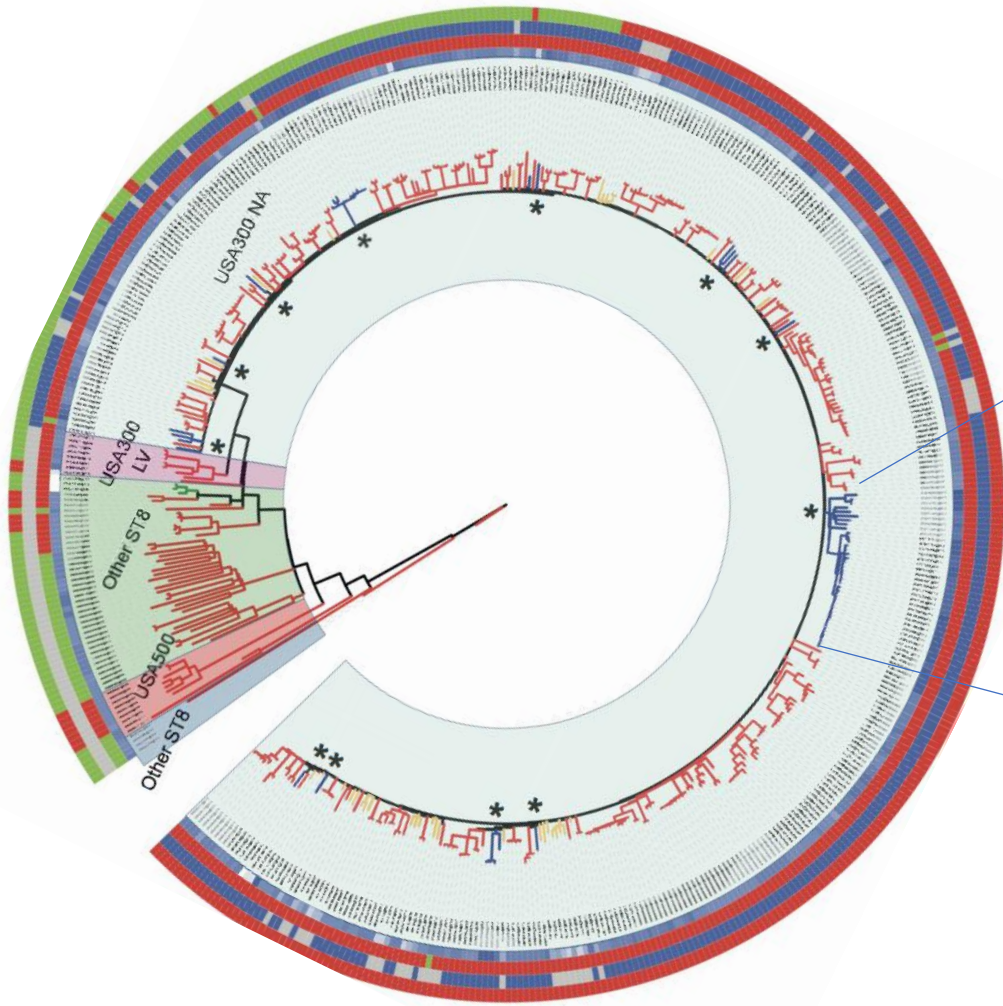
- Cas 2: *Staphylococcus aureus* – USA 300
 - USA300 est le clones de *S. aureus* dominant aux USA.
 - Epidémie à *S. aureus* USA300 dans un hôpital gériatrique
 - Séquençage de 63 souches USA300 du CNR

⇒ Jusqu'à 28 SNPs entre deux isolats de l'hôpital

⇒ Il y a-t-il une seule source?



Analyse dans un contexte global

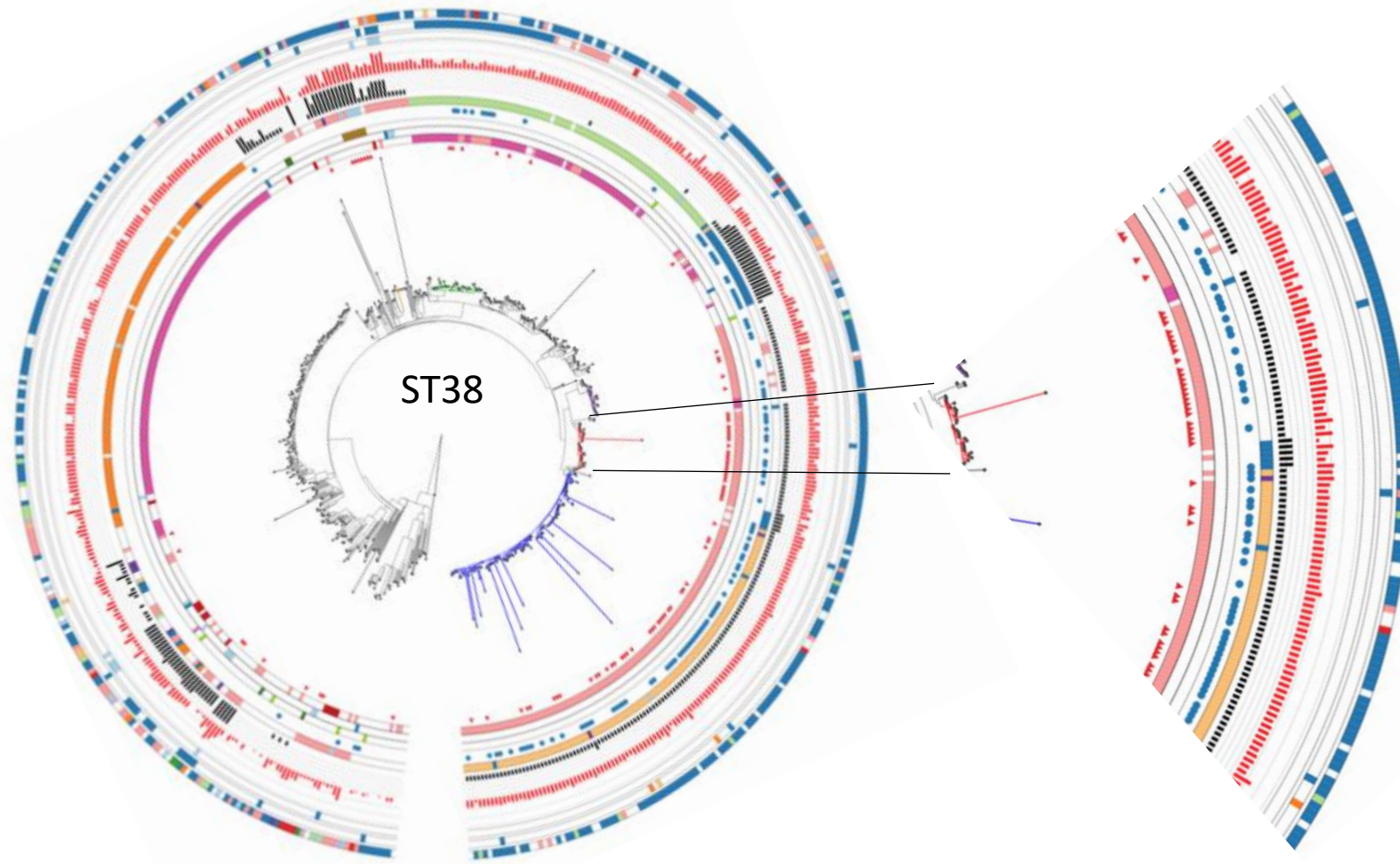


- Monophylétique
⇒ Une seule source probable
⇒ Diversité génétique au sein de la source

Analyse d'une épidémie vs. dissémination

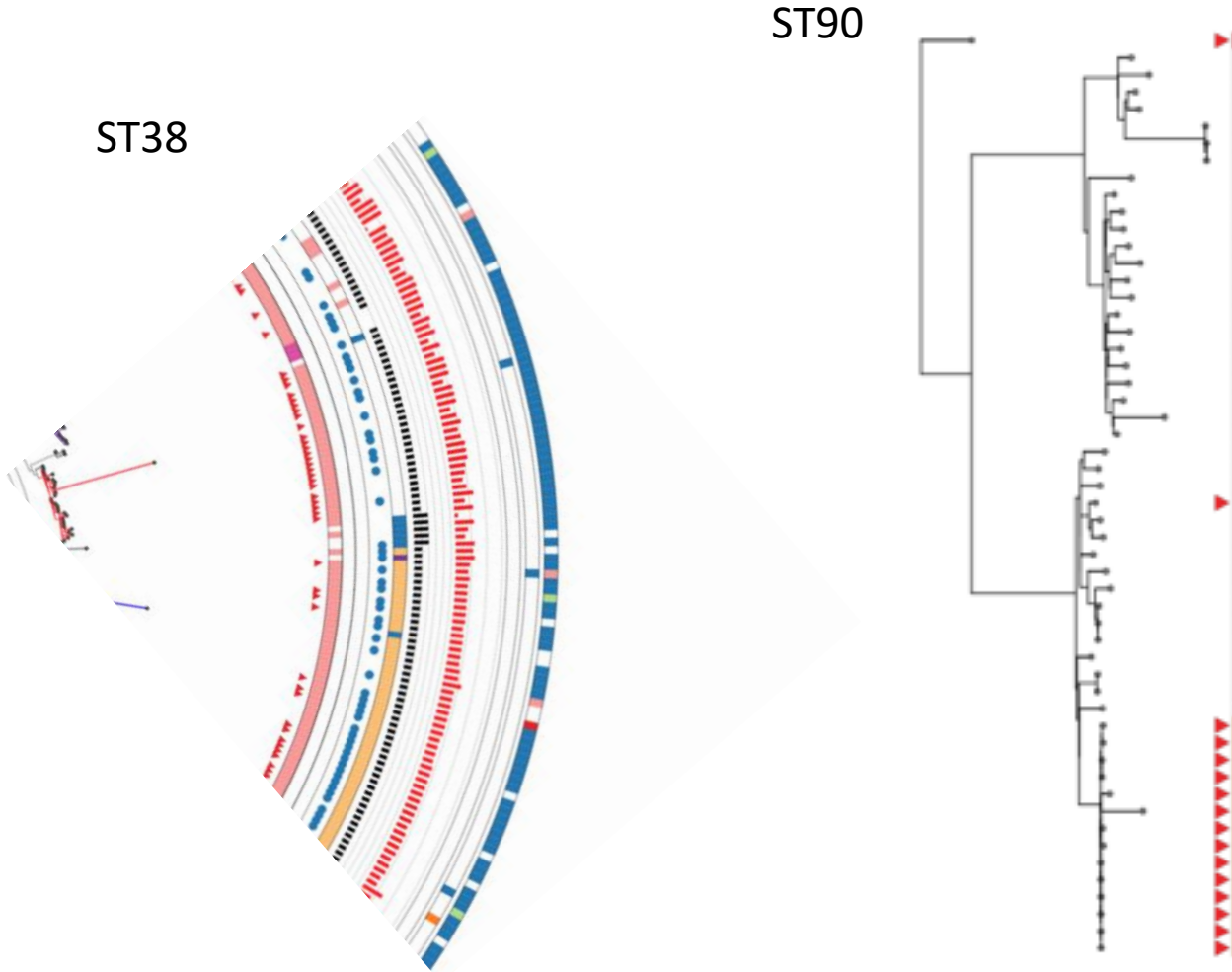
- **Cas 3:** *Escherichia coli* productrices de carbapénémase
 - Bactéries considérées « priorité critique » pour le développement de nouveaux antibiotiques
- ⇒ Séquençage de 712 souches reçues par le CNR avant 2016
- ⇒ Analyse de l'évolution et des cas groupés.

E. Coli producteur de carbapénèmase



- Cluster de souches isolées en France
 - 4 souches des pays bas
 - De 1 à 38 SNPs
 - Clone pandémique

E. Coli producteur de carbapénémase



- Cluster de souches ST38 isolées en France
 - 4 souches des pays bas
 - De 1 à 38 SNPs
 - Clone pandémique
- Souches ST90 (endoscope)
 - De 1 à plus de 135 SNPs (mutator)

⇒ Pour conclure sur une possible transmission et identifier la source, nécessité de combiner données génomiques et épidémiologiques

Contamination supposée à partir d'un endoscope

Prédiction du phénotype de résistance: l'antibiogramme *in silico*

- La résistance aux antibiotiques dépend:
 - De gènes de résistance
 - Nombreux variants pour chaque gène
 - De mutations dans le core génome
 - Mutations répertoriées: fluoroquinolones: *gyrA* et *parC*
 - Mutations prédictibles: inactivation d'un répresseur de pompes d'efflux
 - Mutations inconnues
 - Du fond génétique
 - Du nombre de copies de certains gènes
 - De la combinaison de ces facteurs

Méthodes « rules-based »

Prends en compte

- Les gènes de résistance
- Les mutations répertoriées: fluoroquinolones: *gyrA* et *parC*, porines, *rpoB*

Ne prend pas en compte

- Les mutation non décrites, de nombreuses mutations 'prédictibles'
- Le fond génétique
- Le nombre de copies de certains gènes
- La combinaison de ces facteurs

Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data

N. Stoesser^{1,2*}, E. M. Batty³, D. W. Eyre¹, M. Morgan², D. H. Wyllie⁴, C. Del Ojo Elias¹, J. R. Johnson⁵, A. S. Walker¹,
T. E. A. Peto¹ and D. W. Crook^{1,2}

- 74 *E. coli* et 69 *K. pneumoniae*
- amoxicillin, co-amoxiclav, ceftriaxone, ceftazidime, ciprofloxacin, gentamicin and meropenem.
- Comparaison prédiction basée sur les gènes / CMI (SIR)
- Sensibilité 0.96 et spécificité 0.97

Example: le site Resfinder

Center for Genomic Epidemiology

Chromosomal point mutations ☐

Acquired antimicrobial resistance genes ☐

Select species

Klebsiella* ▼

*Chromosomal point mutation database exists

Select type of your reads

Assembled Genome/Contigs ▼

If you get an "Access forbidden. Error 403": Make sure the start of the web adress is https and not just http. Fix it by clicking [here](#).

 Choose File(s)

Name

Size

Progress

Status

 Upload

 Remove

Sortie: résistance et gènes de résistance

Antimicrobial	Class	WGS-predicted phenotype	Genetic background
vancomycin	glycopeptide	No resistance	aac(6')-Ib-cr (aac(6')-Ib-cr_EF636461), aac(6')-Ib-Hangzhou (aac(6')-Ib-Hangzhou_FJ503047), rmtF (rmtF_JQ808129)
mupirocin	pseudomonic acid	No resistance	
tobramycin	aminoglycoside	Resistant	
virginiamycin m	streptogramin a	No resistance	
isepamicin	aminoglycoside	No resistance	erm(B) (erm(B)_JN899585)
virginiamycin s	streptogramin b	Resistant	
hydrogen peroxide	peroxides	Resistant	sitABCD (sitABCD_AY598030)
butirosin	aminoglycoside	No resistance	blaSHV-106 (blaSHV-106_AM922307), blaOXA-232 (blaOXA-232_JX423831), blaCTX-M-186 (blaCTX-M-186_KX446939), blaTEM-1B (blaTEM-1B_AY458016), blaCTX-M-15 (blaCTX-M-15_AY044436)
ampicillin	beta-lactam	Resistant	

Aminoglycoside								
Resistance gene	Identity	Alignment Length/Gene Length	Position in reference	Contig or Depth	Position in contig	Phenotype	PMID	Accession no.
aac(6')-Ib-cr	99.2292870906	519/519	1..519	KPST231_2	61801..62319	ciprofloxacin	17938184	EF636461
aac(6')-Ib-Hangzhou	99.6146435453	519/519	1..519	KPST231_2	61801..62319	amikacin,tobramycin	19703723	FJ503047
aadA2	100.0	792/792	1..792	KPST231_3	22519..23310	spectinomycin,streptomycin	22511964	JQ364967
rmtF	100.0	780/780	1..780	KPST231_2	62730..63509	amikacin,gentamicin,tobramycin	22547620	JQ808129

Méthodes basées sur l'apprentissage (machine learning)

- Intelligence artificielle

- Principe de la méthode

- Jeu de données d'apprentissage: souches avec génome séquencé et phénotype de résistance (CMI ou SIR)
- Méthode d'apprentissage
- Jeu de données test avec souches séquencées et phénotypes

- **Difficultés:** comment décrire un génomes de 5 Million de bases avec un pan-génome très vaste comparé à la taille du jeu d'apprentissage.

- Méthode basée sur des mots de longueur k (k-mer) $k=13$
- Méthode basée sur la phylogénie.

>30 publications depuis 2018

Bioinformatics, 34, 2018, i89–i95
doi: 10.1093/bioinformatics/bty276
ISMB 201

A pan-genome-based machine learning approach for predicting antimicrobial resistance activities of the *Escherichia coli* strains

Hsuan-Lin Her¹ and Yu-Wei Wu^{2,*}

- 59 génomes d'*E. coli*
- 38 antibiotiques
- *E. coli*, *Salmonella*, *K. pneumoniae*, *S. aureus*, *A. baumannii*, *P. aeruginosa*, *M. tuberculosis* ...

- ⇒ Uniquement des publication décrivant des méthodes / apprentissage
- ⇒ Résultats très « prometteurs »
- ⇒ Pas d'utilisation pour le moment

 **frontiers**
in Cellular and Infection Microbiology

Learning From Limited Data: Towards Best Practice Techniques for Antimicrobial Resistance Prediction From Whole Genome Sequencing Data

OPEN ACCESS

Lukas Lüttinger^{1,2}, Peter Májek¹, Stephan Beisen¹, Thomas Rattei²
and Andreas E. Posch^{1*}

- 8704 génomes de 5 espèces de la base de donnée Patric

Using Genetic Distance from Archived Samples for the Prediction of Antibiotic Resistance in *Escherichia coli*

Derek R. MacFadden,^{a,b,c} Bryan Coburn,^{a,d} Karel Brinda,^{b,e} Antoine Corbeil,^h Nick Daneman,^a David Fisman,^{a,f} Robyn S. Lee,^{e,f}
ID Marc Lipsitch,^b Allison McGeer,^{a,f} ID Roberto G. Melano,^{g,h} Samira Mubareka,^{a,h} William P. Hanage^b



RESEARCH

Predicting Antimicrobial Resistance Using Partial Genome Alignments

ID D. Aytan-Aktug,^a M. Nguyen,^{b,c} ID P. T. L. C. Clausen,^a R. L. Stevens,^{b,d,e} F. M. Aarestrup,^a ID O. Lund,^a ID J. J. Davis^{b,c,f}

Information utilisée?

- Marqueurs de résistance
- Proximité phylogénétique

A venir: site web pour un antibiogramme *in silico*

Chromosomal point mutations ☐

Acquired antimicrobial resistance genes ☐

Select species

Klebsiella* ▼

*Chromosomal point mutation database exists

Select type of your reads

Assembled Genome/Contigs ▼

If you get an "Access forbidden. Error 403": Make sure the start of the web adress is https and not just http. Fix it by clicking [here](#).

 Choose File(s)

Name	Size	Progress	Status

 Upload

 Remove

Conclusions

- Le séquençage génomique est la méthode la plus puissante d'épidémiologie moléculaire
- Doit être combinée à des données épidémiologiques
- Permet de faire des découvertes fonctionnelles
- Promesses pour un antibiogramme *in silico*
- Pour le moment pas de preuve de concept à l'hôpital

Seq2Diag

ABRomics



Plan Prioritaire de Recherche
PIA

Remerciements

Unité EERA, Institut Pasteur

- Isabelle Rosinski-Chupin
- Nicolas Cabanel
- Rafael Patino-Navarrete

CNR – Hôpital Bicêtre

- Thierry Naas
- Laurent Dortet
- Rémy Bonin

CNR – Staph Lyon

- François Vandenesch