

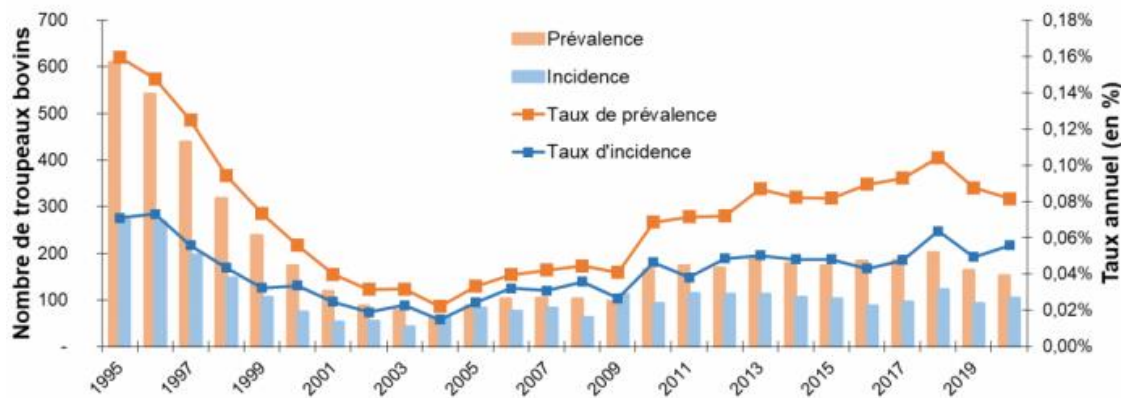
Outbreak reconstruction with a slowly evolving multi-host pathogen: a comparative study of three existing methods on *Mycobacterium bovis* outbreaks.

Hélène Duault, Benoit Durand and Laetitia Canini.



La bTB en France

- En Europe = enjeu économique (EU directive 64/432/EEC)
- Persistance de foyers actifs (Sud-Ouest) → statut officiellement indemne menacé
- Composition majoritaire du système, en plus des bovins:
 - Blaireaux, Sangliers et Cervidés (Côte d'Or et Nouvelle-Aquitaine)
 - Cerf élaphe et Sanglier (Normandie)


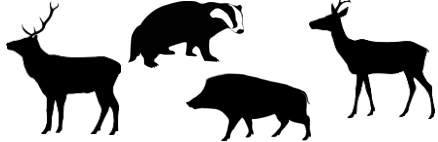


Evolution de la prévalence et de l'incidence de la tuberculose bovine de 1995 à 2020 (© <https://www.platforme-esa.fr>)



La bTB en France: surveillance incomplète

Modalités de surveillance des bovins et de la faune sauvage

Population surveillée		
Totalité de la population	Oui	Non
Variations selon la région	Oui	Oui
Mise en place	Début des années 60	Récente (2011)

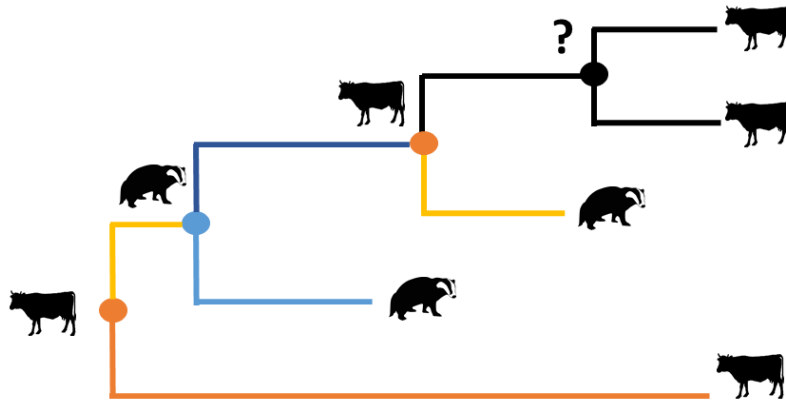
Populations sauvages vs. bovins: différence de surveillance

→ Biais d'échantillonnage

- Données recueillies par la surveillance:
 - Epidémiologiques
 - Génétiques (génotypage/WGS)

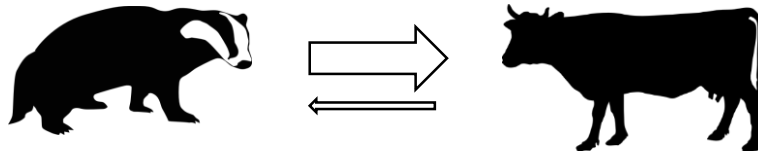
La bTB en France: étude précédente

- Reconstruction phylogénétique



- Modèle d'évolution

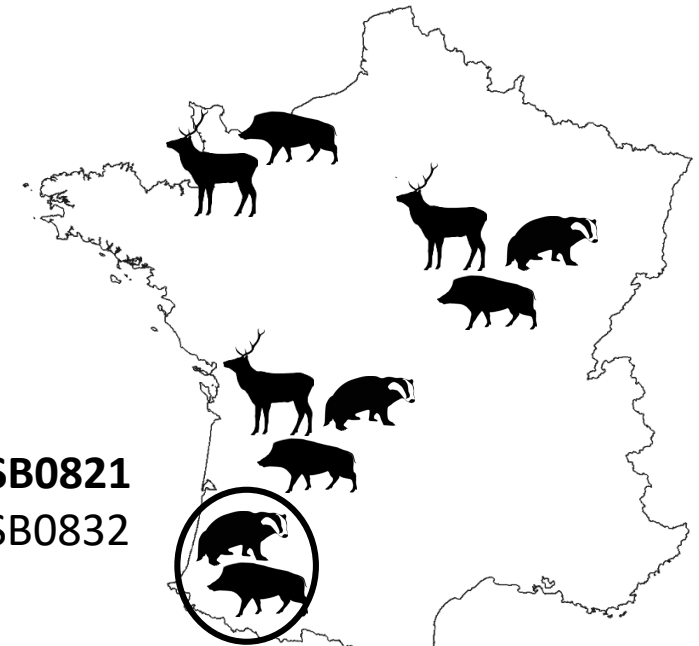
Taux de transition inter-espèces



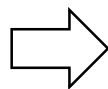
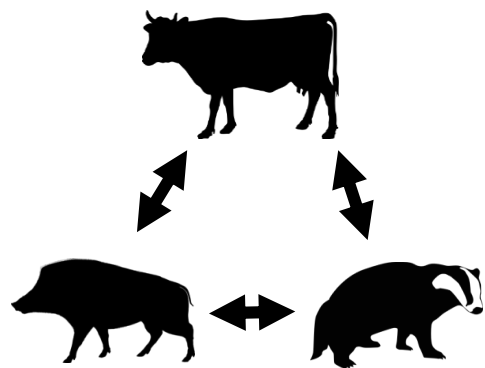
→ Autre méthode pour analyser ces données?

SB0821

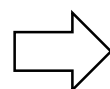
SB0832



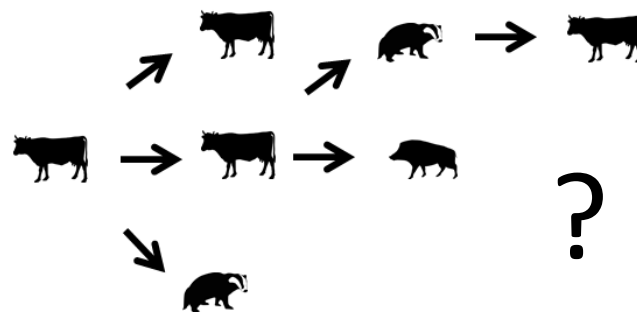
Objectif



Données
biaisées



Arbre de transmission



Temps

Arrive-t-on à reconstruire correctement l'épidémie?

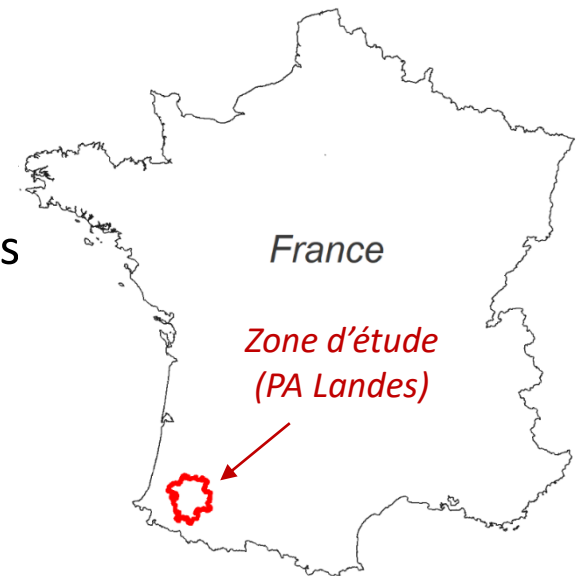
Comment le biais d'échantillonnage influence-t-il la reconstruction d'arbres de transmission de *M. bovis* ?

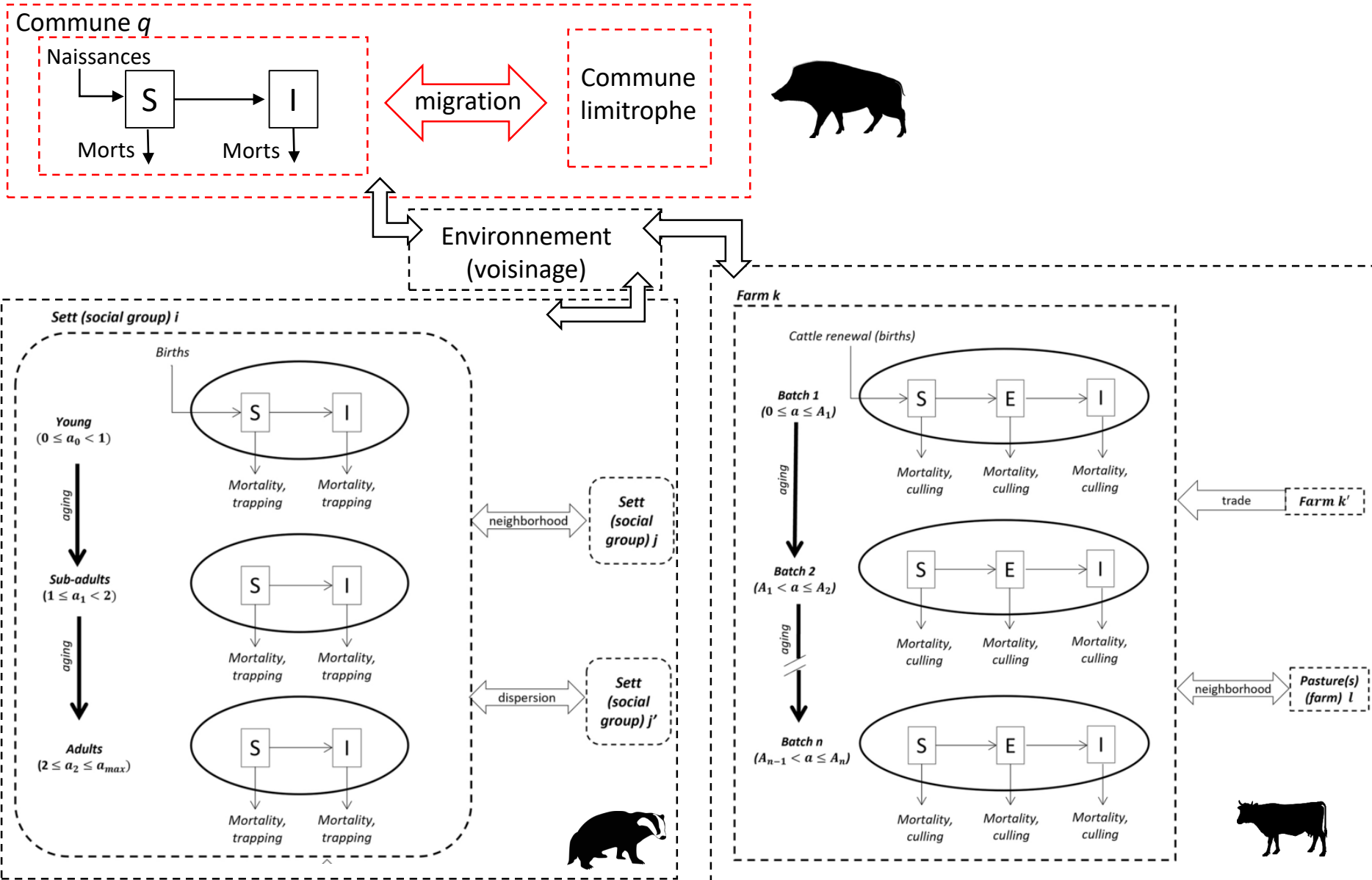
Stratégie

1. Simulation d'arbres de transmission
 - Permettant la construction d'arbres de référence où l'on connaît tous les individus infectés ainsi que les événements de transmission datés
2. Simulation de séquences génétiques
 - Nécessaires pour implémenter les méthodes de reconstruction
3. Schémas d'échantillonnage
 - Représentant des biais que l'on connaît de la surveillance de *M. bovis* en France
 - Appliqués aux arbres de référence
4. Reconstruction d'arbres de transmission
5. Comparaison entre les arbres de référence et les arbres reconstruits

1. Simulation d'arbres de transmission

- Modèle de simulation d'arbres de transmission existants (Bouchez-Zacria *et al.*, 2023) :
 - Durée de la simulation : 13 ans (2007-2020)
 - Stochastique/Compartimental/Pas de temps = 1 mois (cohérent avec bTB)
- Modifications:
 - Ajout d'une population de sangliers (modèle simple non validé).
 - Unité épidémiologique = individu
- 30 arbres de référence : $> 15/\text{espèce}$ & < 500 individus
- Bovins en cas index.







$$\lambda_{i,\text{pop}}(t) = \lambda_i^{\text{intra}}(t) + \lambda_i^{\text{voisins}}(t) + \lambda_i^{\text{env}}(t)$$

$$\lambda_i^{\text{intra}}(t) = \beta^{\text{intra}}_{\text{sp}} I/N$$

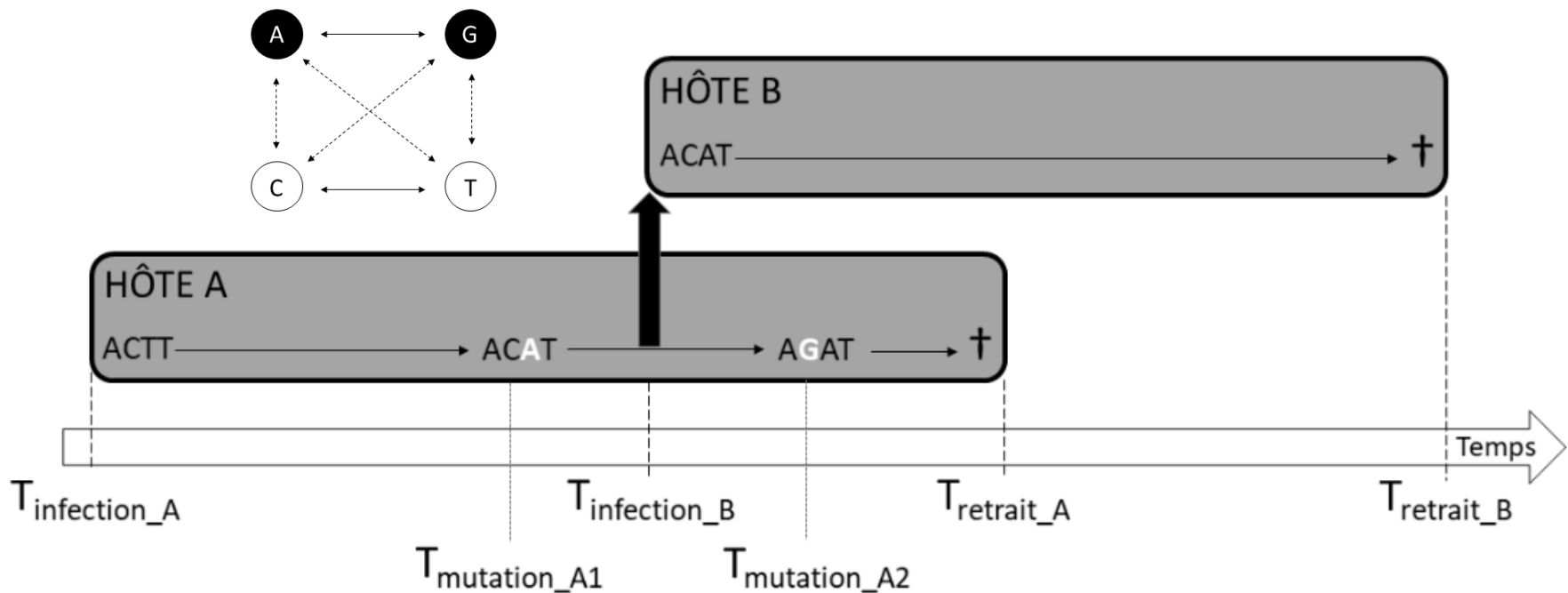
$$\lambda_i^{\text{voisins}}(t) = \sum_{(j, \text{sp}^*)/j \in V(\text{sp}, i)} \varepsilon_{\text{sp}} \lambda_j^{\text{intra}}(t)$$

$$\lambda_i^{\text{env}}(t) = \sum_{\substack{(j, \text{sp}^*)/j \in V(\text{sp}^*, i) \\ \text{sp}^* \neq \text{sp}}} \beta^{\text{env}}_{\text{sp}} \cdot 1^{\text{env}}_j$$



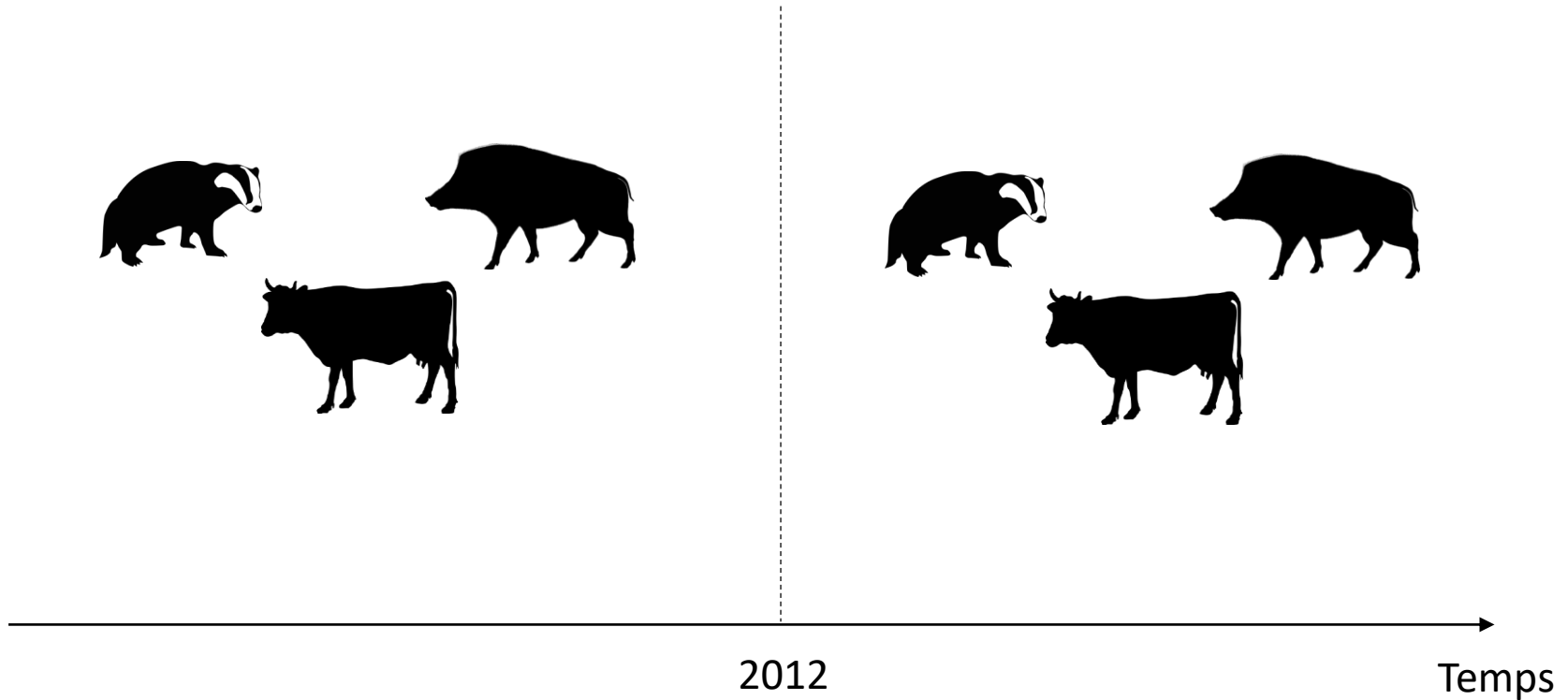
2. Simulation de séquences génétiques

- Pas de diversité intra-hôte.
- Modèle de substitution \rightarrow modèle HKY (Crispell *et al.*, 2017; Salvador *et al.*, 2019).
- Taux de mutation (μ) et ratio transition/transversion (κ) (Duault *et al.*, 2022).



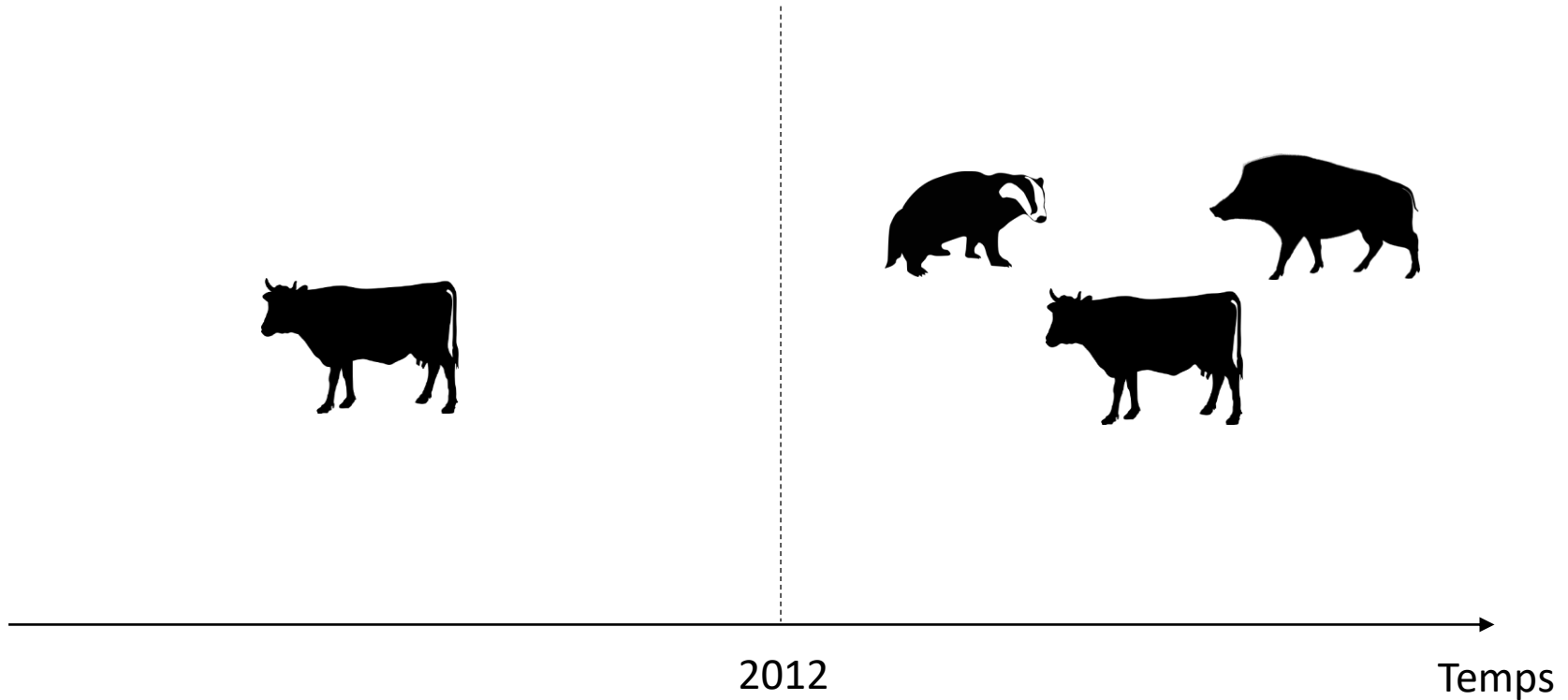
3. Schémas d'échantillonnage

- Schéma de **référence** = tous les hôtes sont échantillonnés.



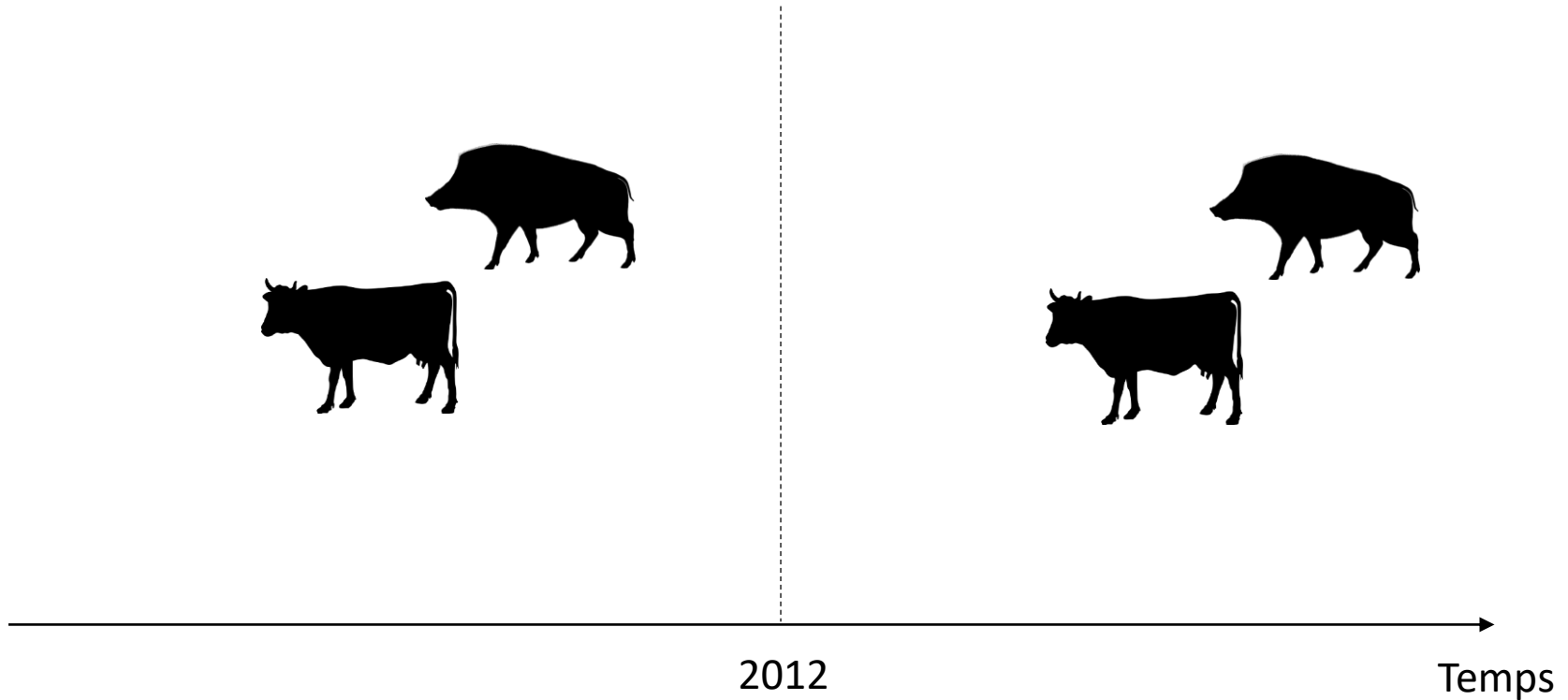
3. Schémas d'échantillonnage

- Biais temporel (T) = surveillance tardive de la faune sauvage.



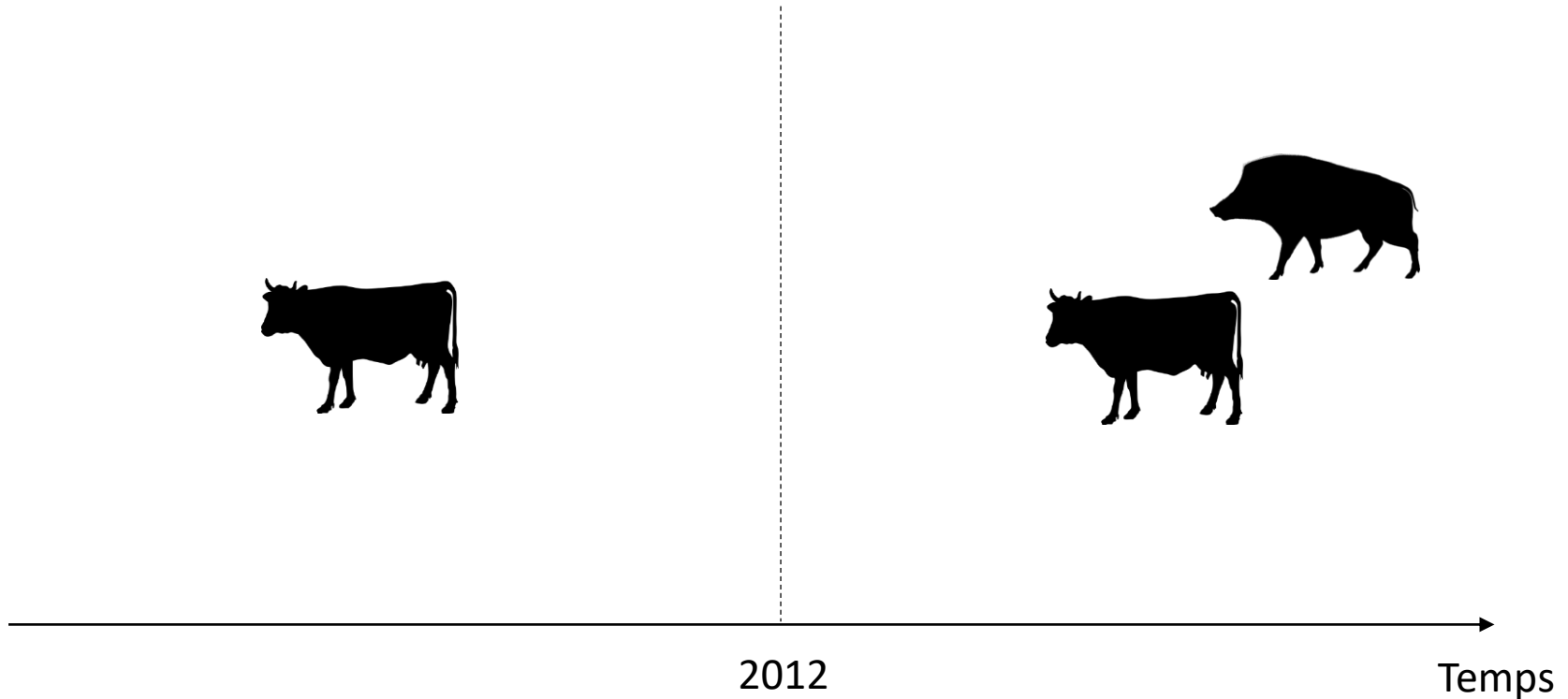
3. Schémas d'échantillonnage

- Biais d'espèce (B ou W) = surveillance d'une seule espèce sauvage.



3. Schémas d'échantillonnage

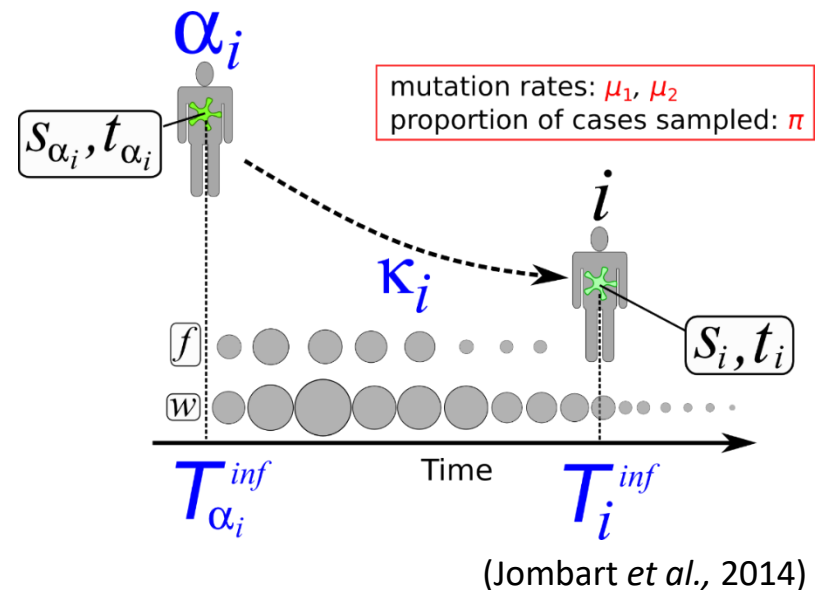
- Biais combiné (T+B ou T+W) = surveillance tardive + d'une seule espèce sauvage.



4. Reconstruction d'arbres de transmission

- seqTrack (Jombart *et al.*, 2011)
 - Ancêtre = temps d'échantillonnage plus ancien + distance génétique minimale
 - Algorithme d'Edmonds
 - Un seul arbre reconstruit
- *outbreaker2* (Campbell *et al.*, 2019)
 - Vraisemblance génétique = distance génétique
 - Vraisemblance temporelle = **modèle de transmission** (temps de génération + intervalle infection – échantillonnage)
 - **Méthode Bayésienne** = arbre reconstruit résume les arbres postérieurs en prenant l'ancêtre le plus fréquent pour chaque hôte infecté

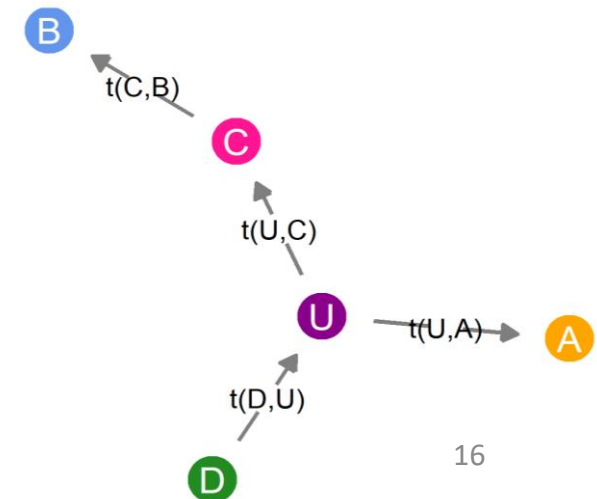
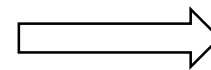
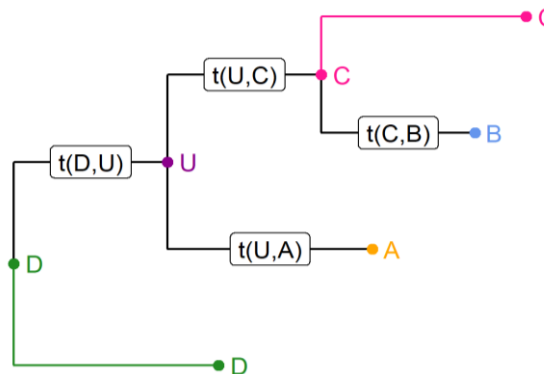
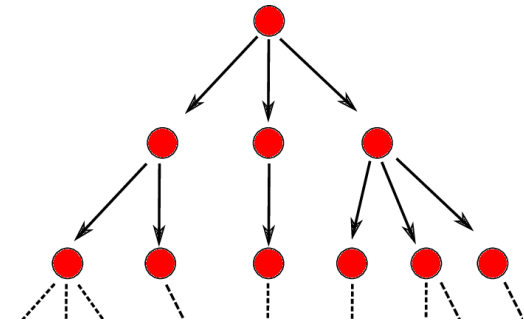
Méthodes non phylogénétiques



4. Reconstruction d'arbres de transmission

- Séquentielle = *TransPhylo* (Didelot *et al.*, 2017)
 - Evolution intra-hôte = processus coalescent neutre constant
 - Modèle épidémiologique (distribution du nombre de cas secondaires + temps de génération)
 - Méthode Bayésienne = sélection de l'arbre médoid (qui s'éloigne le moins de tous les autres) OU matrice de probabilités de transmission (probabilité la + haute)

Méthodes phylogénétiques



5. Indicateurs épidémiologiques

Exactitude

Présence de super-spreaders

outbreaker2
TransPhylo
seqTrack

Espèce-hôte du cas index

Taille de l'épidémie

Contribution des espèces-hôtes

outbreaker2
TransPhylo

Exactitude

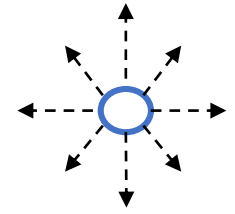
- Reconstruire liens de transmission dirigés = difficile (système impliquant la faune sauvage OU une proportion faible d'hôtes échantillonnés).



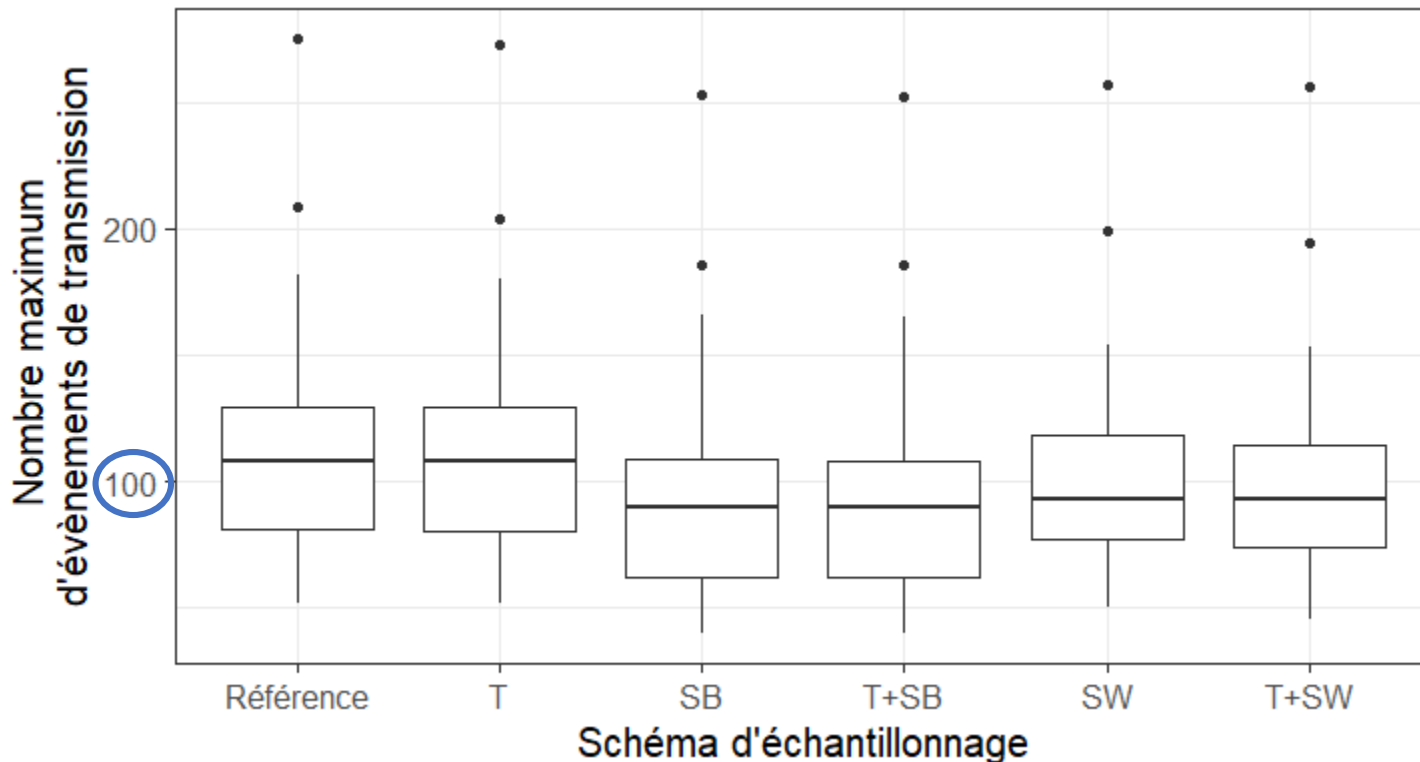
Maximum de 17% (*TransPhylo*)
seqTrack < *outbreaker2* < *TransPhylo*



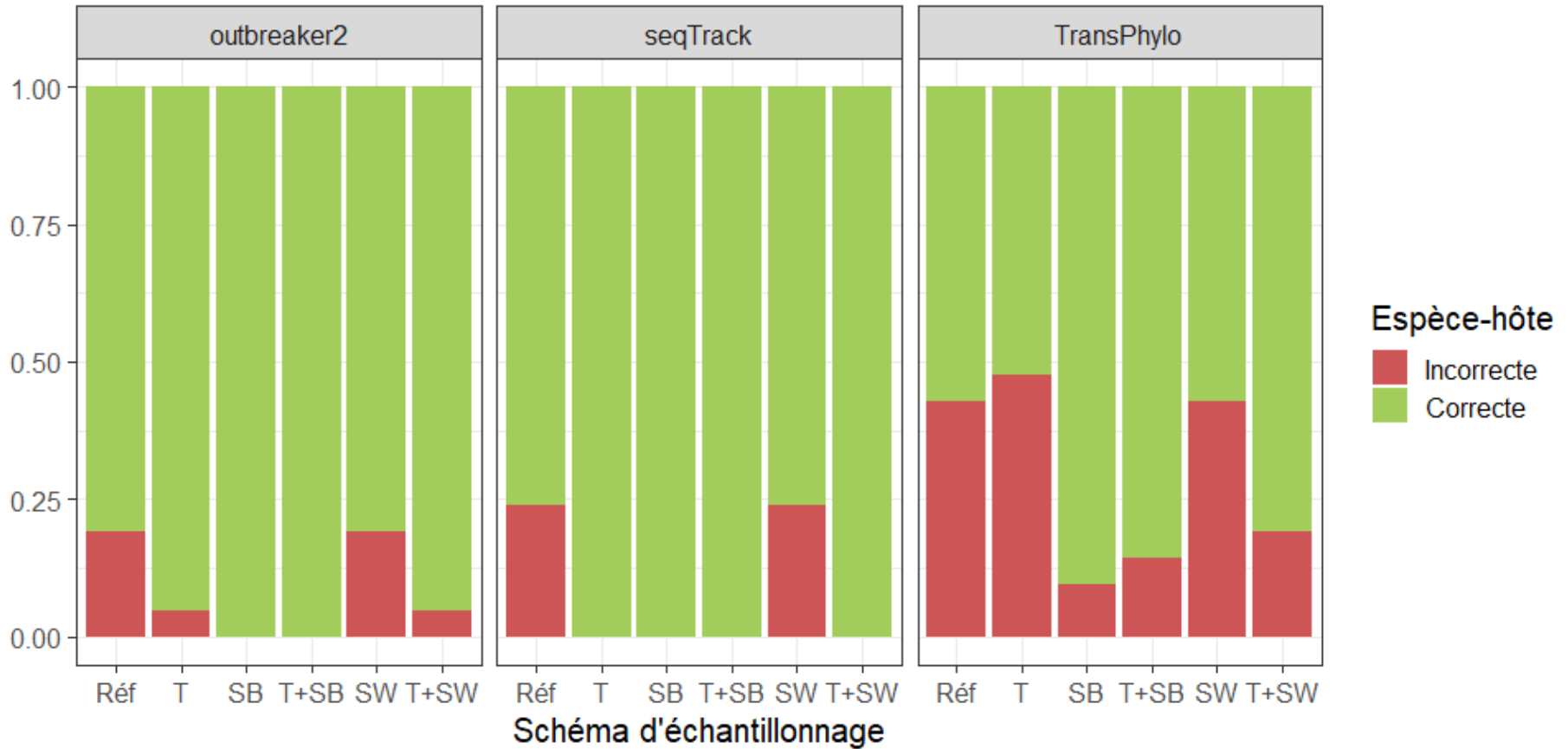
Présence de super-spreaders



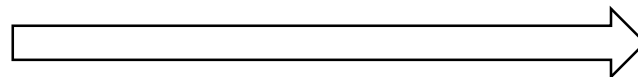
- Concerne uniquement seqTrack
 - Absent dans les arbres de référence
 - Moins de 10% des hôtes infectés sont responsables de plus de 80% des événements de transmission



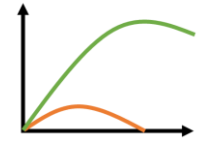
Espèce-hôte du cas index



Meilleure performance avec du biais



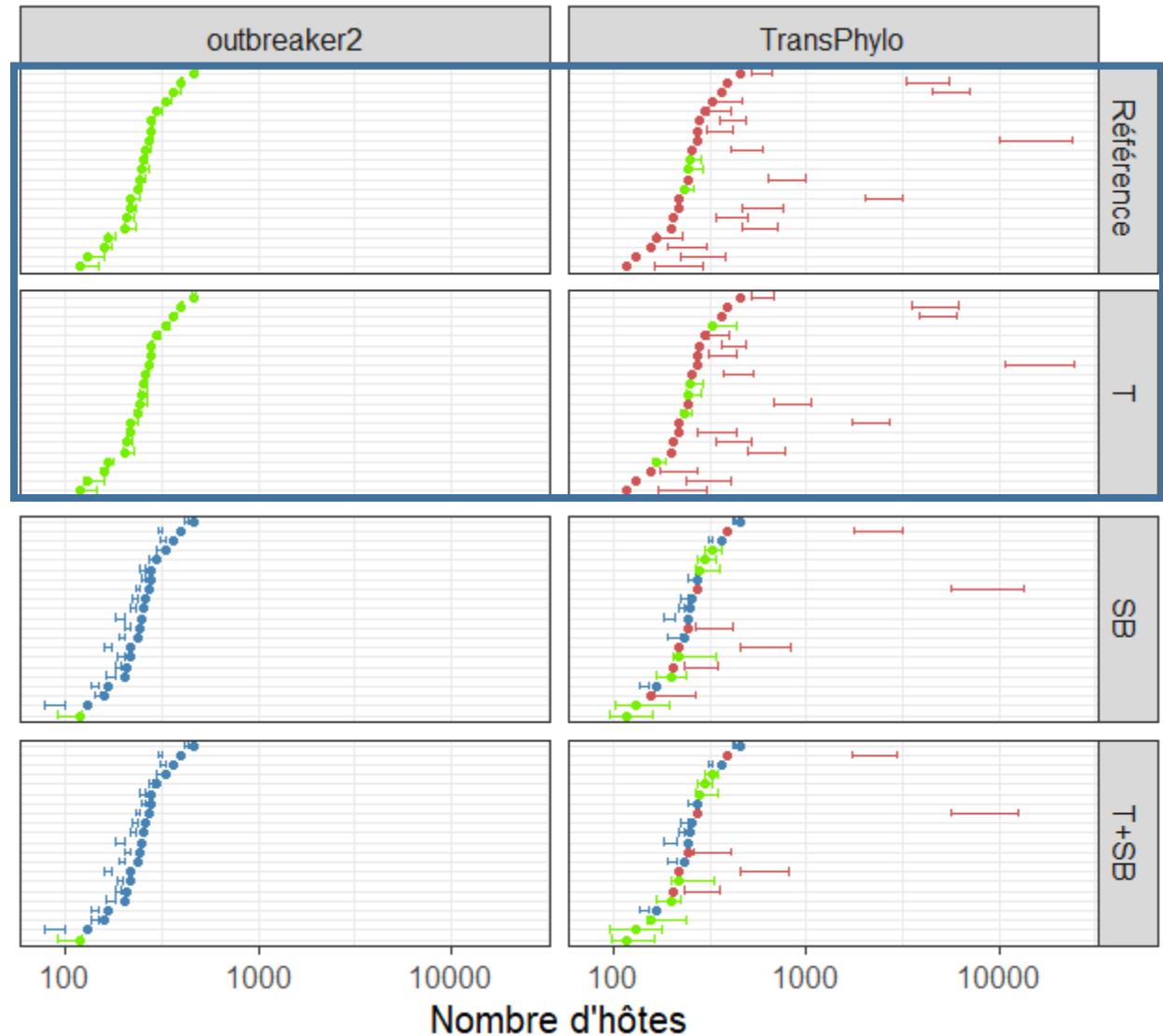
Taille de l'épidémie



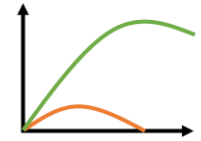
outbreaker2
 =
 Taille simulée dans
 l'intervalle
 ≠
TransPhylo
 =
 Surestimation de la
 taille de l'épidémie

Estimation

- Sous-estimation
- Correcte
- Surestimation



Taille de l'épidémie



outbreaker2
=
Moins bonne
performance avec un
biais d'espèce

≠

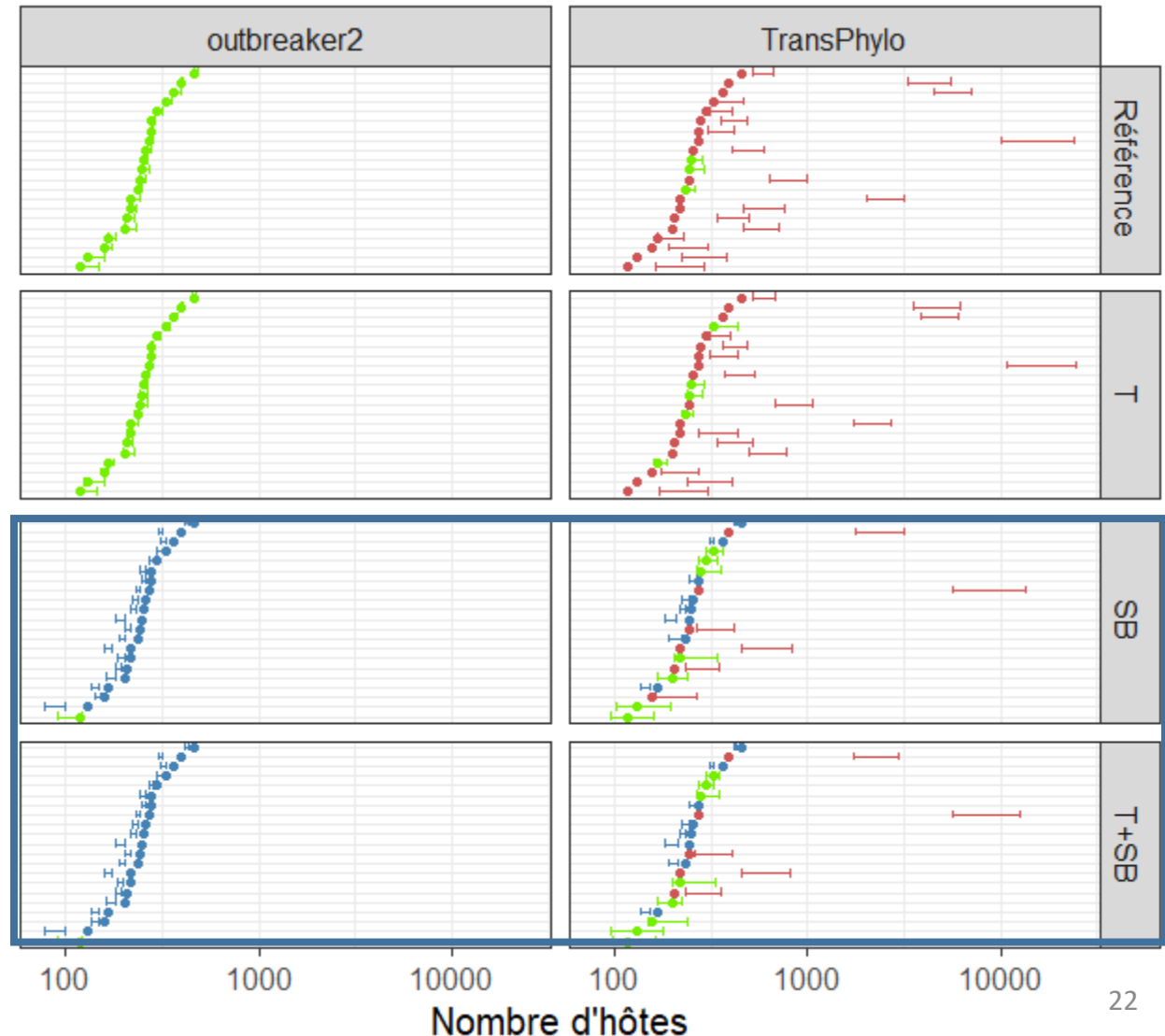
TransPhylo

=

Meilleure
performance avec un
biais d'espèce

Estimation

- Sous-estimation
- Correcte
- Surestimation



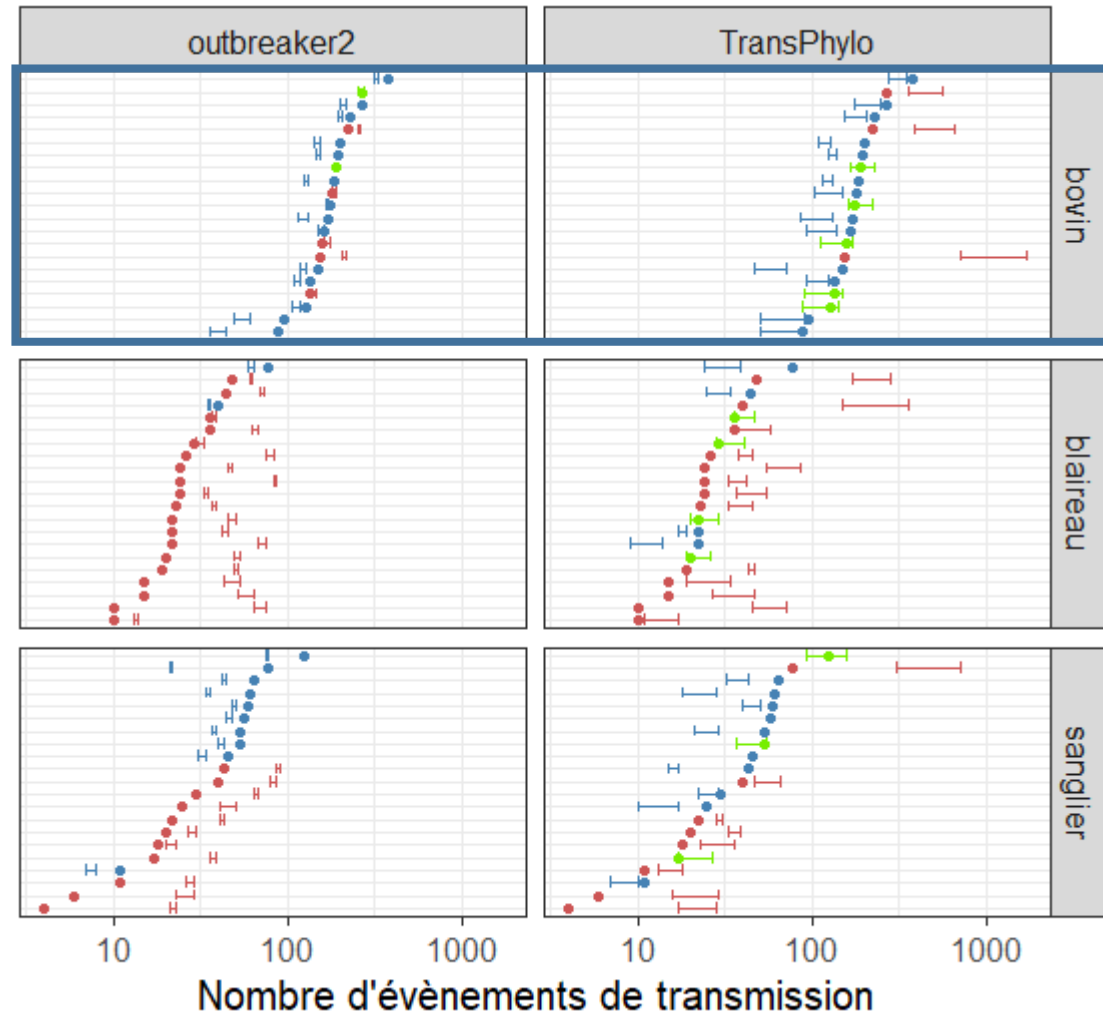
Contribution des espèces-hôtes



Bovins

=

Sous-estimation du
nombre
d'évènements de
transmission



Estimation

- Sous-estimation
- Correcte
- Surestimation

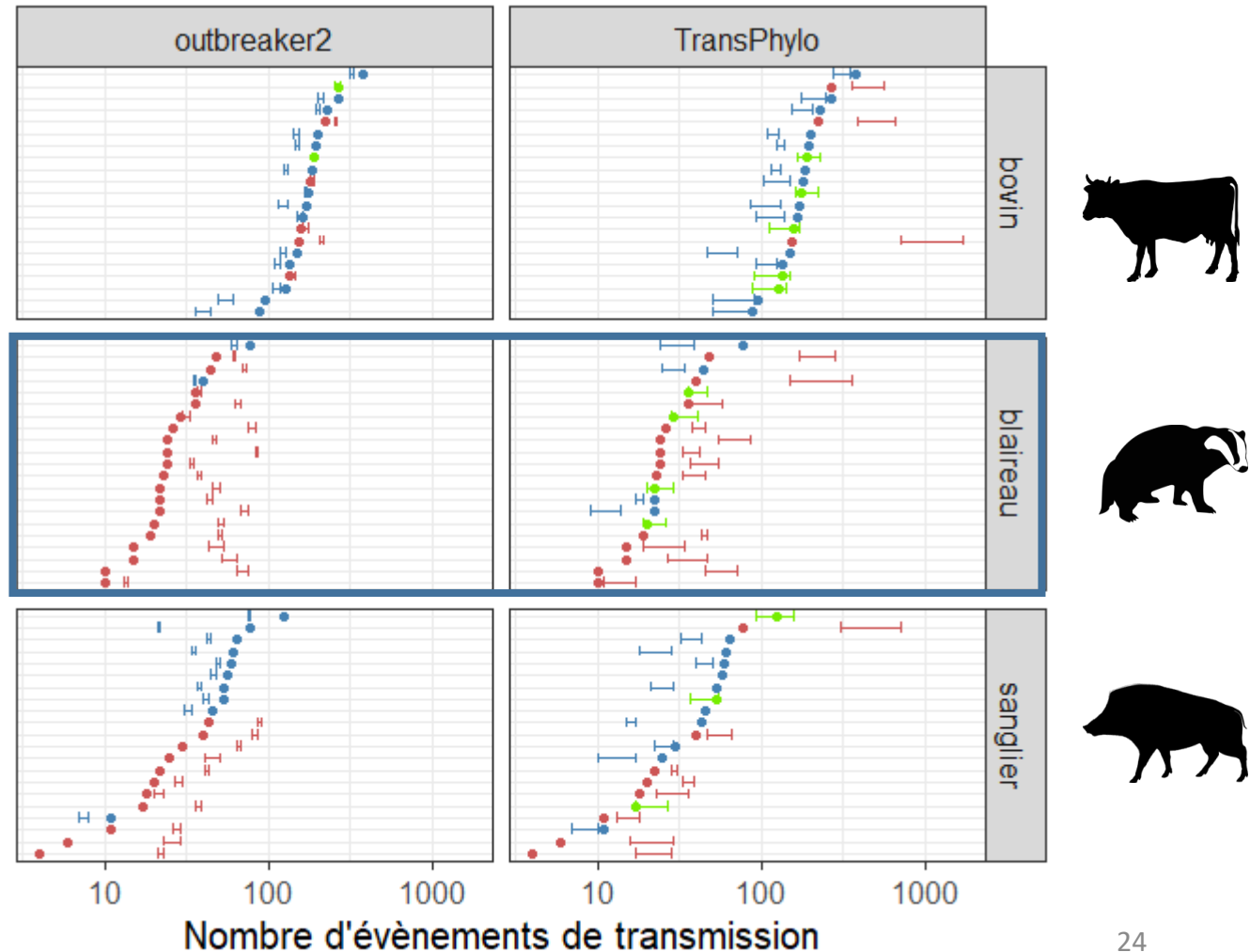
Contribution des espèces-hôtes



Blaireaux
=
Surestimation du
nombre
d'évènements de
transmission

Estimation

- Sous-estimation
- Correcte
- Surestimation



Contribution des espèces-hôtes

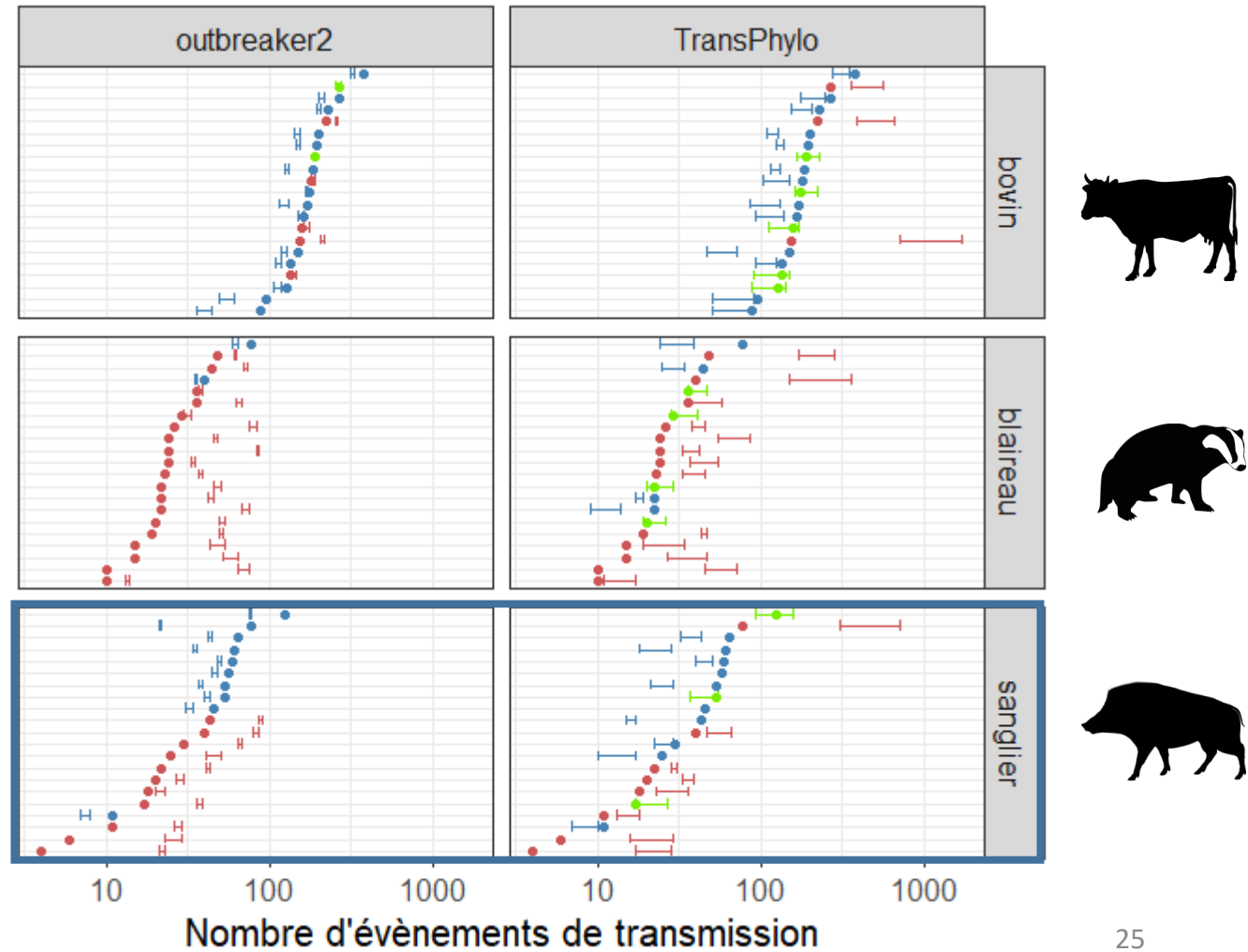


Sanglier

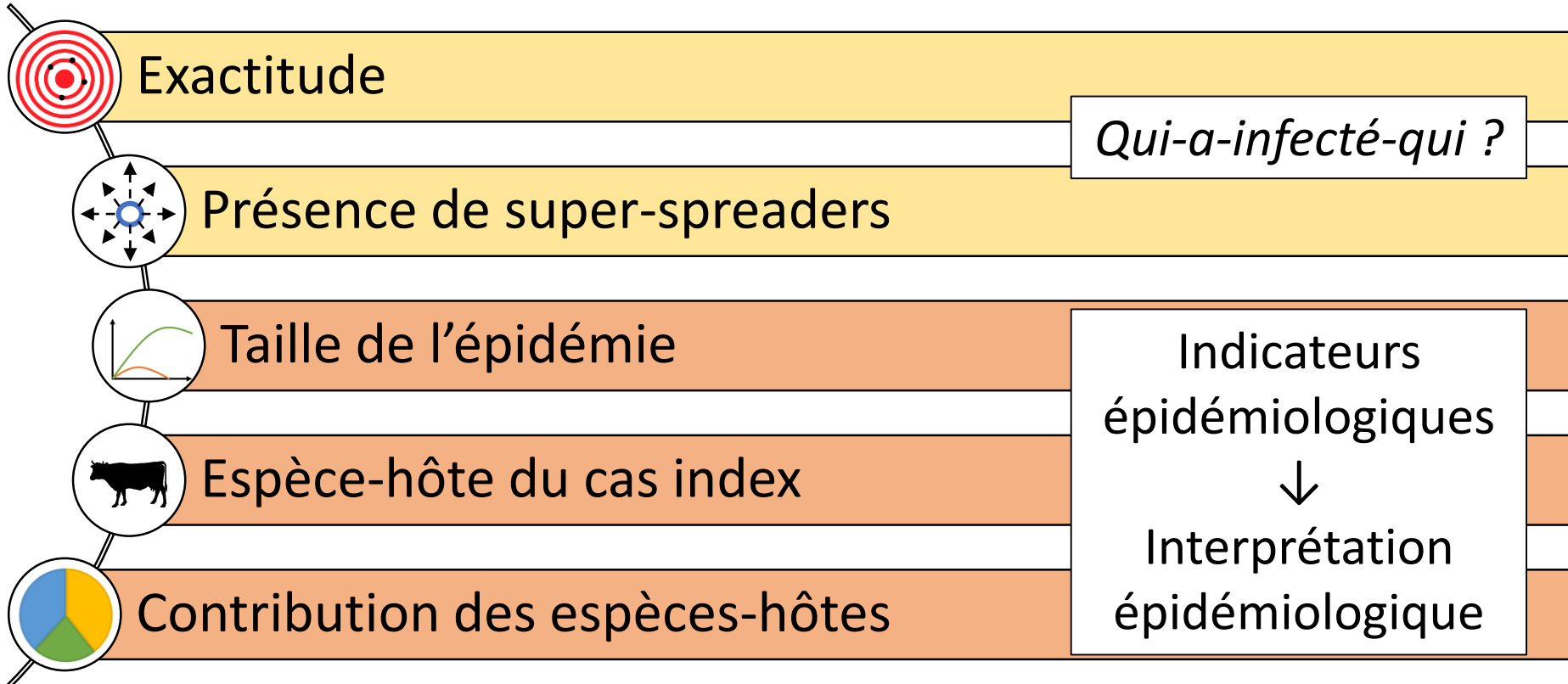
=

Mauvaise estimation


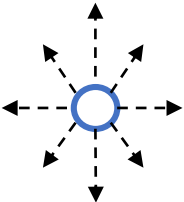
Aucune tendance



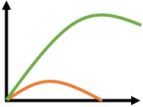


Deux types d'indicateurs



Qui-a-infecté-qui ?

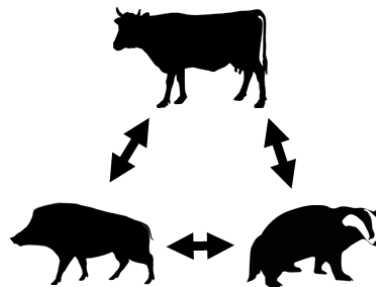
Indicateur	Performance
<p data-bbox="208 522 444 565">Exactitude</p> 	<ul style="list-style-type: none"><li data-bbox="542 544 1721 715">■ Mauvaise performance globale attendue<ul style="list-style-type: none"><li data-bbox="639 605 1721 715">■ Similaire aux résultats de Sobkowiak <i>et al.</i>, 2022 (<i>M. tuberculosis</i> = taux de mutation faible)<li data-bbox="542 733 1315 776">■ Toujours plus faible pour seqTrack
<p data-bbox="150 831 502 939">Présence de super-spreaders</p> 	<ul style="list-style-type: none"><li data-bbox="542 853 1779 1153">■ seqTrack<ul style="list-style-type: none"><li data-bbox="639 915 1460 958">■ Absence de modèle de transmission<li data-bbox="639 976 1779 1090">■ Fortement dépendant de l'ordre d'échantillonnage (Nigsch <i>et al.</i>, 2021)<li data-bbox="639 1109 1692 1152">■ Super-spreaders très prolifiques (peu réalistes)

Interprétation épidémiologique

Indicateur	Performance
Taille de l'épidémie 	<ul style="list-style-type: none"> ■ Surestimée par <i>TransPhylo</i> et non par <i>outbreaker2</i> ■ Hypothèses de <i>TransPhylo</i>: faible échantillonnage + diversité intra-hôte
Espèce-hôte du cas index 	<ul style="list-style-type: none"> ■ Meilleure performance pour seqTrack et <i>outbreaker2</i> ■ <i>TransPhylo</i>: hôtes non échantillonnés = cas index
Contribution des espèces-hôtes 	<ul style="list-style-type: none"> ■ Globalement mauvaise performance → méthodes inadaptées ■ Sous-estimation de la contribution des BV ■ Surestimation de la contribution de la FS

A retenir

- Limites de l'étude :
 - Contraintes pratiques (**temps de calcul très long**)
 - Modèle de simulation des séquences → pas de variation du taux de mutation entre espèces-hôtes
 - Données épidémiologiques considérées (comparaison entre méthodes)
 - Seules trois méthodes (*phybreak*)
- Conclusion sur l'étude d'un système multi-hôtes :
 - Nécessité de combiner différents types de méthodes
 - Besoin de développer de nouvelles méthodes
 - Evaluation de ces méthodes sur des systèmes complexes

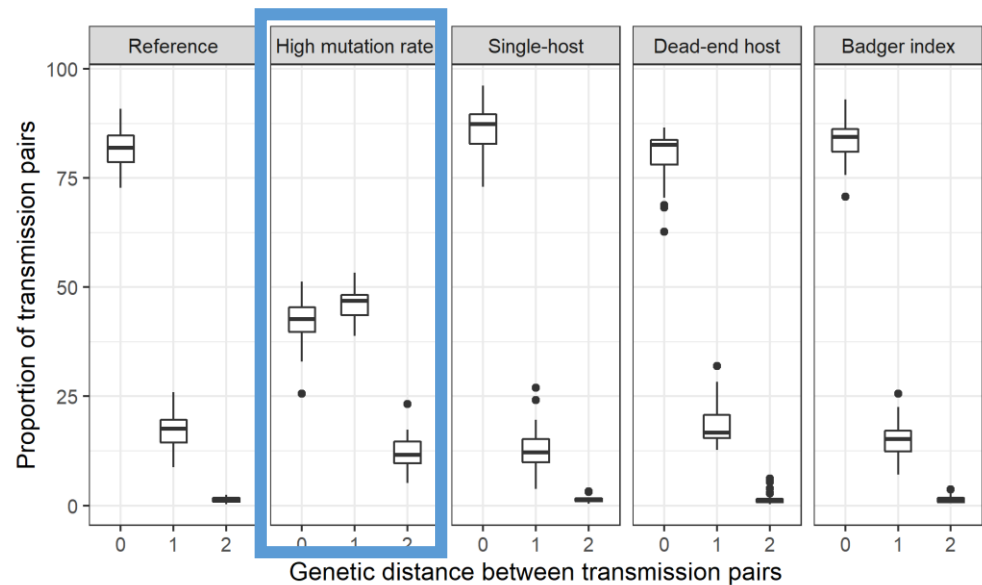


Remerciements

- Benoit Durand et Laetitia Canini
- Maria-Laura Boschioli et Lorraine Michelet
- Comité de thèse :
 - Gaël Beaunée
 - Anne Cori
 - Patrick Hoscheit
 - Elisabeta Vergu
- Toute l'équipe EPI :
 - Gina Zanella
 - Maud Marsot
 - Viviane Domarin
 - Clémence Nadal
 - Alex Drouin
 - Hélène Ladreyt
 - Amalia Rataud
- Amaias Avalos
- Lenin Vinueza
- Erika Ornelas Eusebio
- Toute l'équipe EPIMAI :
 - Julie Rivière
 - Barbara Dufour
 - Valentine Guétin-Poirier
 - Guillaume Crozet
 - Maxime Delsart
- Université Paris-Saclay

Résultats: convergence

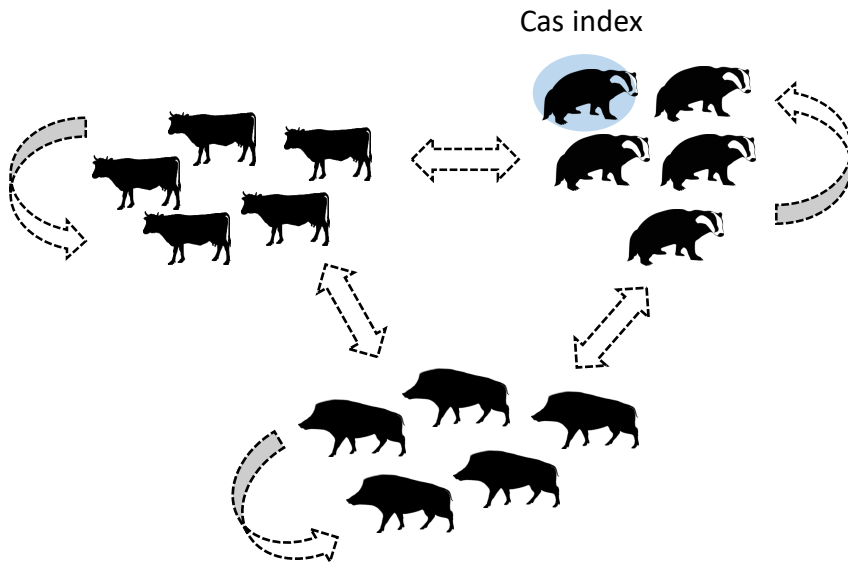
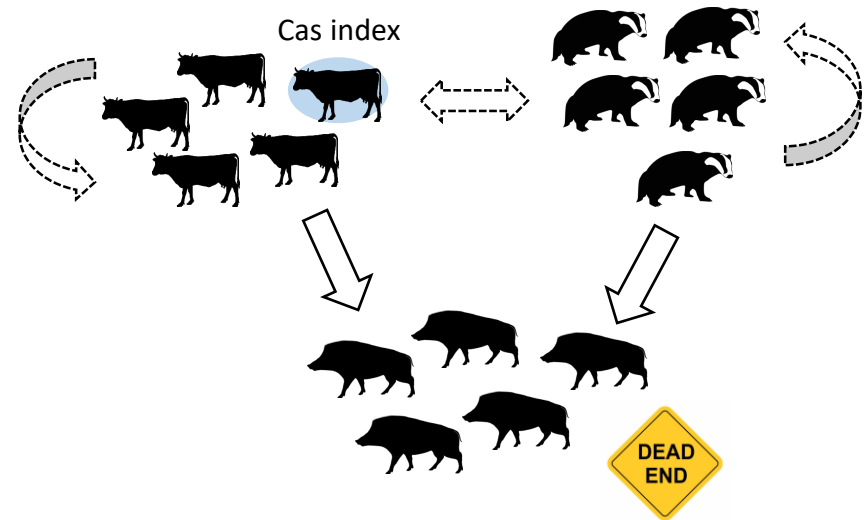
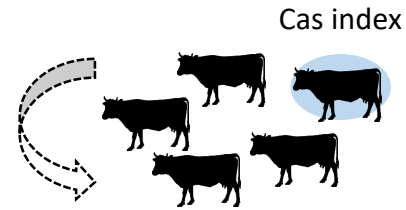
- Facteur limitant pour *TransPhylo*
- Distance génétique entre infecteur et infecté:
 - Majorité < 1
 - ↗ taux de mutation



Paramètre	Arbres ayant convergé	Arbres n'ayant pas convergé
Nombre d'arbres	21	9
Nombre d'hôtes infectés	245 (118-458)	336 (114-376)
Proportion de séquences uniques	6,1 (1,8-8,7)	5,8 (3,4-8,1)
Moyenne de l'écart après la transmission (en nombre de SNP)	0,19 (0,11-0,28)	0,18 (0,09-0,25)

6. Scénarios alternatifs

- Taux de mutation plus élevé (10μ)
- Système composé uniquement de bovins
- Cul-de-sac épidémiologique (SG)
- Blaireau en cas index



Résultats: scénarios alternatifs

Scénario	Résultats par rapport à la référence
Taux de mutation plus élevée	↗ Diversité génétique ↗ Exactitude ↘ Nombre d'évènements dus à un super-spreaders
Système composé uniquement de bovins	↘ Diversité génétique Super-spreaders pour les 3 méthodes
Cul-de-sac épidémiologique (SG)	Surestimation de la contribution des SG
Blaireau en cas index	Moins bonne performance