



HAL
open science

Towards an integrated approach for the annotation of ICEs and IMEs in bacterial genomes

Julie Lao, Gérard Guédon, Thomas Lacroix, Nathalie N. Leblond-Bourget, Helene
H. Chiapello

► **To cite this version:**

Julie Lao, Gérard Guédon, Thomas Lacroix, Nathalie N. Leblond-Bourget, Helene H. Chiapello. Towards an integrated approach for the annotation of ICEs and IMEs in bacterial genomes. JOBIM 2018, Jul 2018, Marseille, France. <hal-01829606>

HAL Id: hal-01829606

<https://hal.univ-lorraine.fr/hal-01829606v1>

Submitted on 4 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Towards an integrated approach for the annotation of ICEs and IMEs in bacterial genomes

Julie Lao^{1,2}, Gérard Guédon¹, Thomas Lacroix², Nathalie Leblond-Bourget¹, Hélène Chiapello²

¹DynAMic, Université de Lorraine, INRA, Vandœuvre-lès-Nancy 54506, France

²MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France

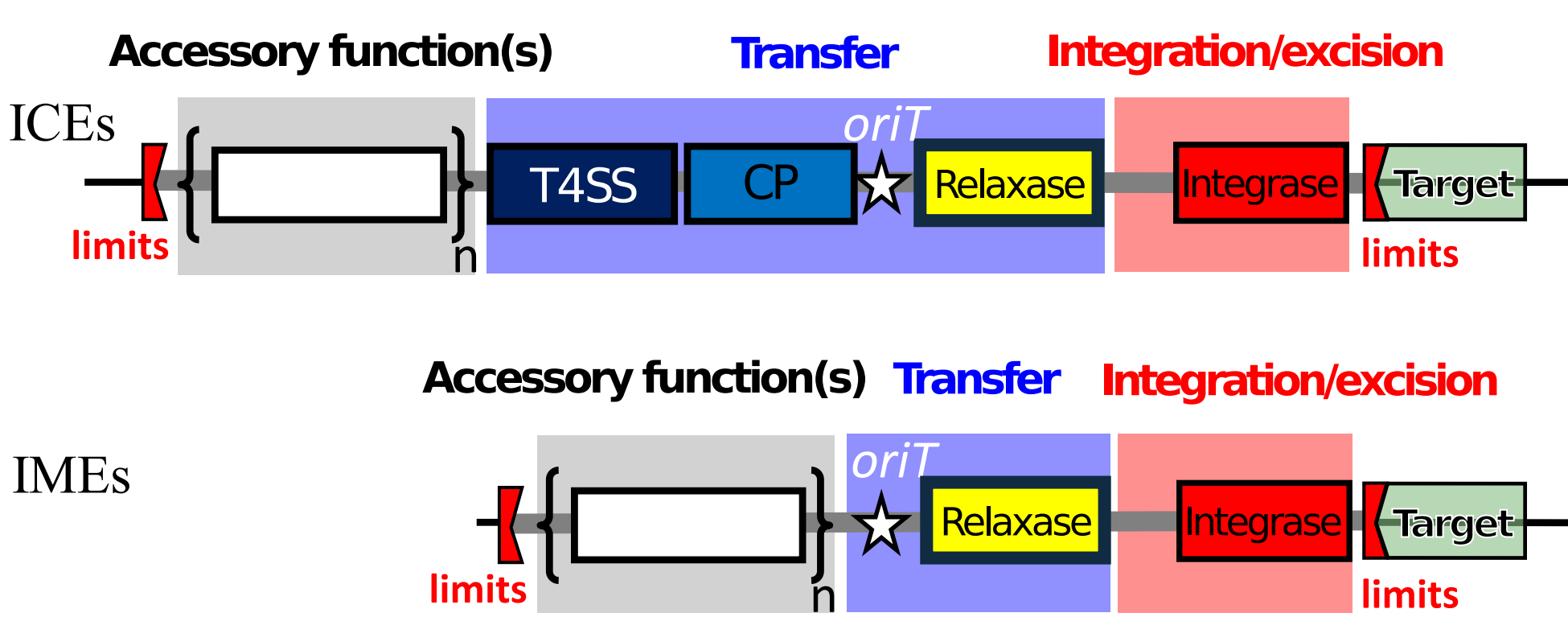
julie.lao@inra.fr

Abstract

- Integrative Conjugative Elements (ICEs) are mobile genetic elements (MGE) that are **integrated in bacterial chromosomes** which encode their **own transfer by conjugation**.
- Integrative Mobilizable Elements (IMEs) are another type of bacterial integrated MGE, they lack a full conjugation system and **hijack co-resident conjugative elements** to promote their own transfer.
- ICEs and IMEs are poorly known and annotated but **play a key role in bacterial genome evolution** and the very few systematic investigations of these MGEs in bacterial genomes suggest that they have a **very high prevalence**.
- **Dedicated bioinformatics approaches are needed** for their identification and precise delineation.

In this poster, we present the first results from the comparison of the two currently existing strategies for the detection and delineation of ICEs: "ICE Finder" strategy^a and "CONJscan" strategy^b

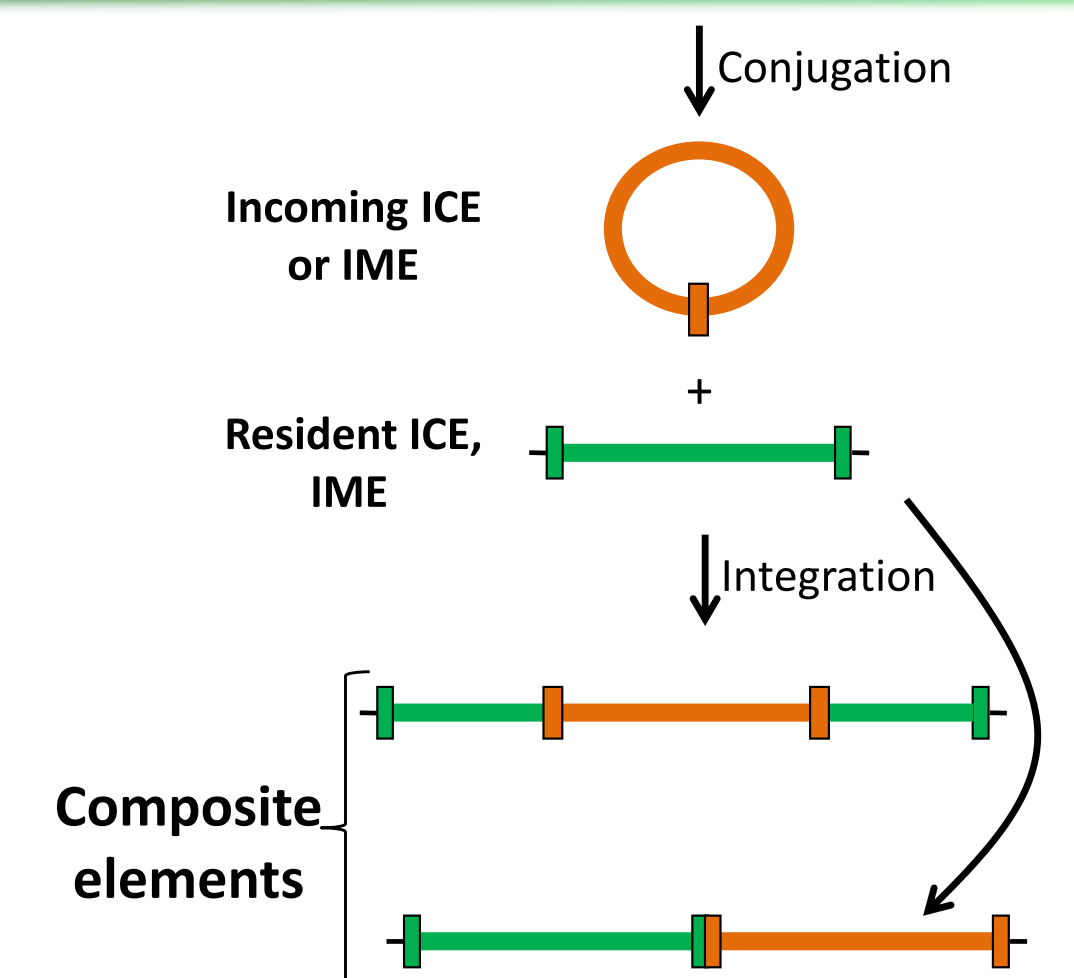
Modular structure



- Genes involved in the same biological function are grouped into modules
- ICE Finder searches 3 genes of the **transfer module** and 1 gene of the **integration module**
- CONJscan searches up to 6 genes of the **transfer module** and **don't search the integration module**

"ICE Finder"	Comparison of strategies	"CONJscan"
Detection		
✓ automatic	Transfer module	automatic ✓
✓ automatic	Integration module	don't search ✗
✓ manual	Target gene	don't search ✗
Delineation		
✓ manual	Simple element	automatic ✓
✓ manual	Composite element	no ✗
Streptococci only	Organism	All bacteria
1	Genomes	at least 4

Composite elements



Many composite elements result from the insertion of an incoming element into another or into one of its ends

"ICE Finder" strategy

- **Automatic** detection of "signature" proteins of the **transfer** and **integration** modules using **BlastP**. A **manually curated database** of known elements in **streptococci**^a is used.
- **Manually annotate and delineate simple and composite elements** at the **nucleotide level** through the research of the **target gene**.
- In many cases, the target gene contains one of the definite boundaries of the element.

"CONJscan" strategy

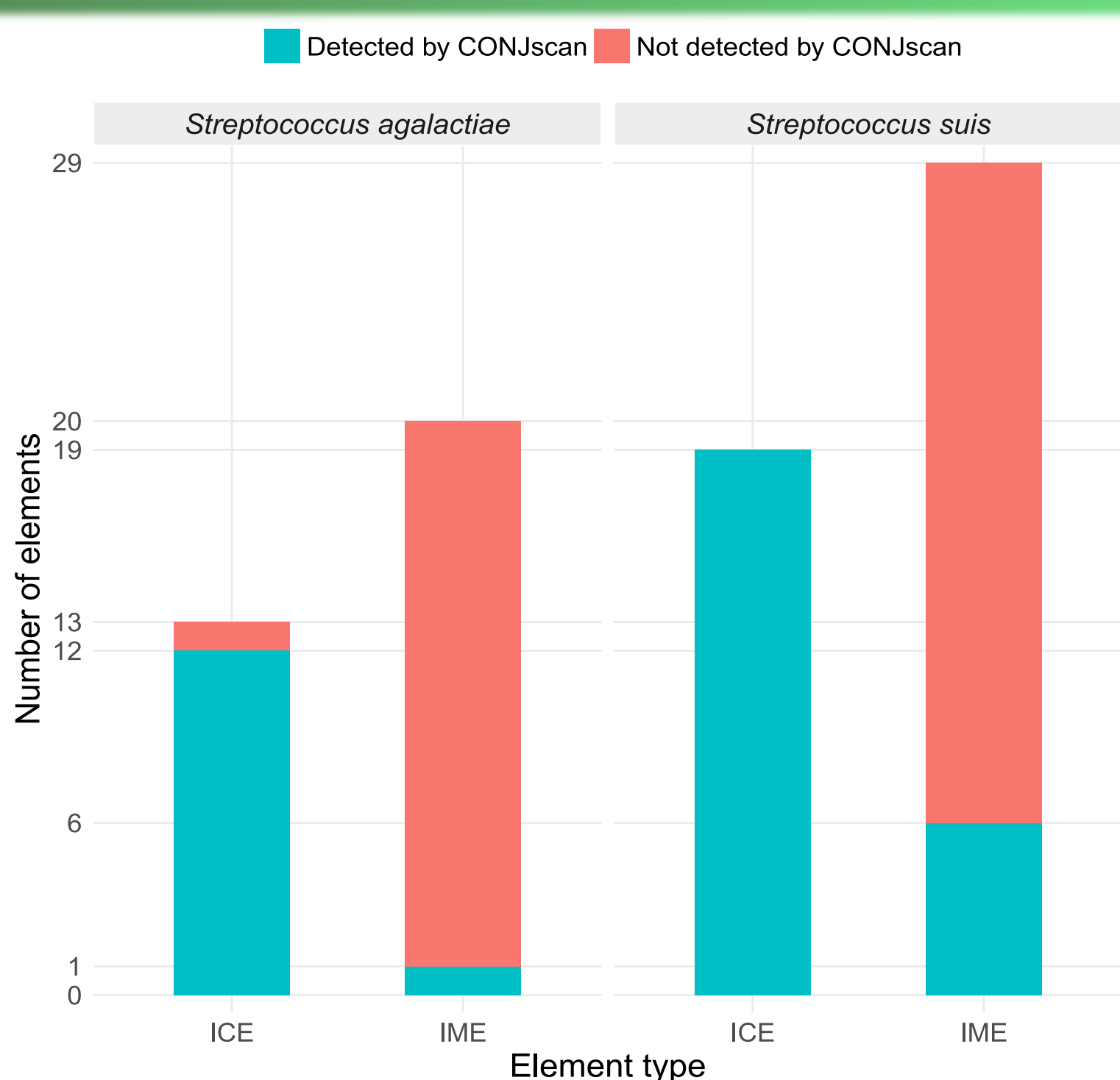
- It is a pipeline that delineates **ICEs only** in **all bacterial genomes** by using the **core genes** that surrounds them^{b,c}.
- This procedure is based on the detection of **one module only (transfer module)** using the CONJscan module of MacSyFinder software^{d,e}.
- It needs **at least 4 different closely-related** genomes to enable ICE annotation.

Comparison design

Results obtained with "ICE Finder" strategy are the reference

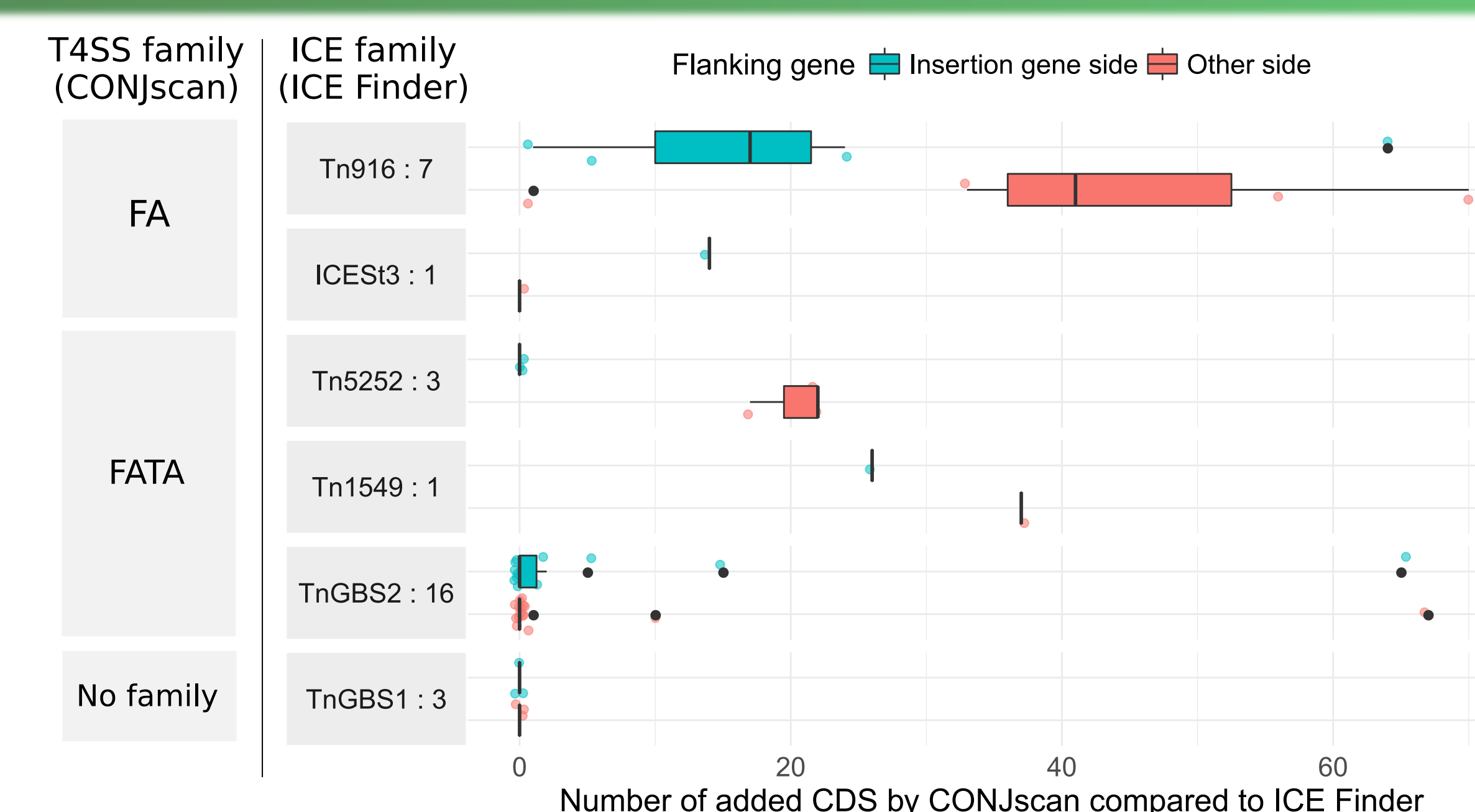
- **Dataset:** 15 genomes of *Streptococcus suis* and 7 genomes of *Streptococcus agalactiae*
- **Detection of transfer module by "CONJscan" strategy is assessed:** Is T4SS families of CONJscan consistent with ICE families of ICE Finder?
- **Is delineation of ICEs with "CONJscan" strategy and "ICE Finder" strategy similar?**

Detection of the transfer module



- ICE Finder detects the transfer module of all known ICEs and IMEs of *Streptococci*.
- CONJscan detects the transfer module of **most ICEs and of some IMEs**. It **does not have dedicated method for IME detection**.

Delineation results: CONJscan compared to ICE Finder



- ICE Finder delineates ICEs at the **nucleotide level**.
- CONJscan delineates **automatically** ICEs at the **gene level**.
- CONJscan **cannot delineate an element within another element and frequently failed to delineate elements that are integrated in tandem**.
- CONJscan **cannot delineate IMEs**.

Integration module and target gene

- "CONJscan" strategy cannot find the target gene because dedicated research of the insertion module is lacking.
- Unlike CONJscan, "ICE Finder" strategy search specifically the integration module (by BlastP of the integrase protein).

Conclusion

- CONJscan has a **more specific definition of the T4SS**.
- CONJscan does not specifically detect the integration module and does not detect most of transfer module from IMEs in *Streptococci*.
- Boundaries obtained with "CONJscan" strategy are not as precise as the ones obtained with "ICE Finder" strategy. But CONJscan always delineate elements with transfer module.

New strategy proposed:

- 1) Detection of **transfer module AND integration module** with MacSyFinder^d:
Development of a **mixed model "CONJscan + Integrase"** including more HMM profiles of transfer protein (of IMEs) that are **not presently searched by CONJscan**.
- 2) Get **first approximate delineation** with "CONJscan" strategy.
- 3) **Refine delineation** with the "ICE Finder" strategy.

References

- a. Ambroset, C., Coluzzi, C., Guédon, G., Devignes, M. D., Loux, V., Lacroix, T., ... & Leblond-Bourget, N. (2016). New insights into the classification and integration specificity of Streptococcus integrative conjugative elements through extensive genome exploration. *Frontiers in microbiology*, 6, 1483.
- b. Cury, J., Touchon, M., & Rocha, E. P. (2017). Integrative and conjugative elements and their hosts: composition, distribution and organization. *Nucleic acids research*, 45(15), 8943-8956.
- c. https://github.com/gem-pasteur/Macsfinder_models/blob/master/Data/Conjugation/Tutorial_ICE.ipynb
- d. Abby, S. S., Néron, B., Ménager, H., Touchon, M., & Rocha, E. P. (2014). MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS one*, 9(10), e110726.
- e. Abby, S. S., Cury, J., Guglielmini, J., Néron, B., Touchon, M., & Rocha, E. P. (2016). Identification of protein secretion systems in bacterial genomes. *Scientific reports*, 6, 23080.

