

Genome validation

Clément Birbes

Andreea Dreau

Christophe Klopp

Claire Kuchly

BUSCO score

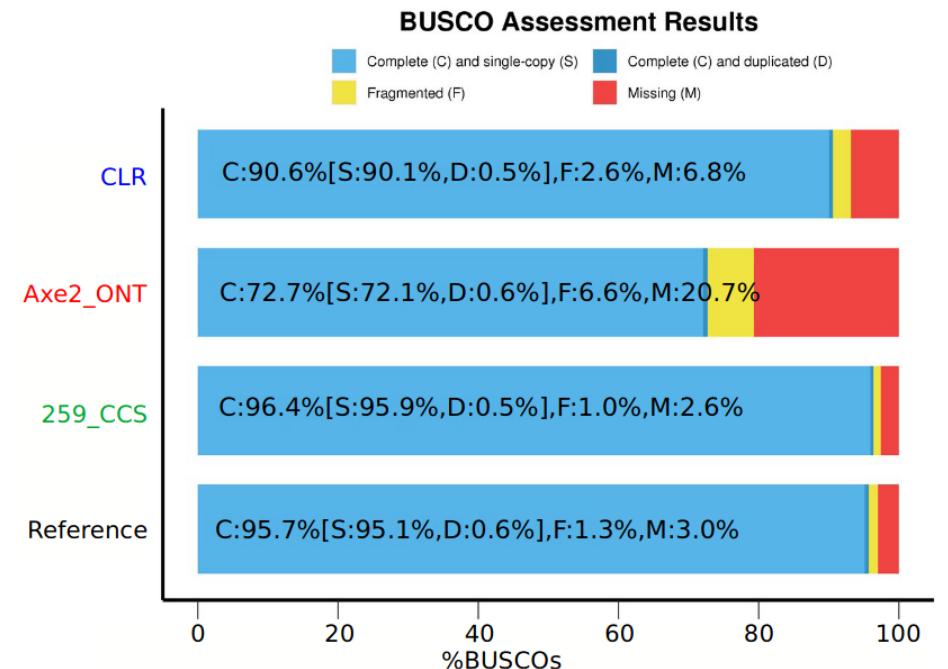


Pros :

- + Easy to understand
- + No additional analysis
- + Explicit graphs

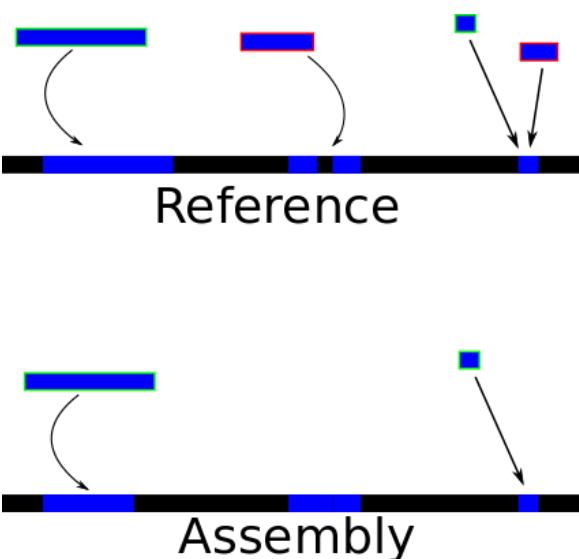
Cons :

- Only targets genes from the database
- Imprecise databases
- Difficult to assess the quality of the assembly after a score of 95%

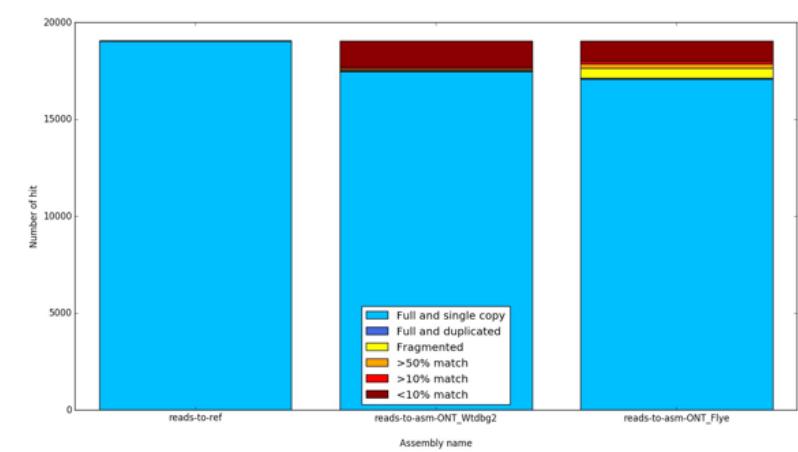


ASMGene Analysis

- Align all annotated transcripts to a reference genome and select best matches.
- For each transcript, record a hit if the transcript is mapped with >99 % identity.
- Align transcript to assembly and extract information for precedent « hit transcript ».



H Metric	ref.cdna	asm.cdna
X full_sgl	36426	36389
X full_dup	0	18
X frag	0	4
X part50+	0	5
X part10+	0	0
X part10-	0	10

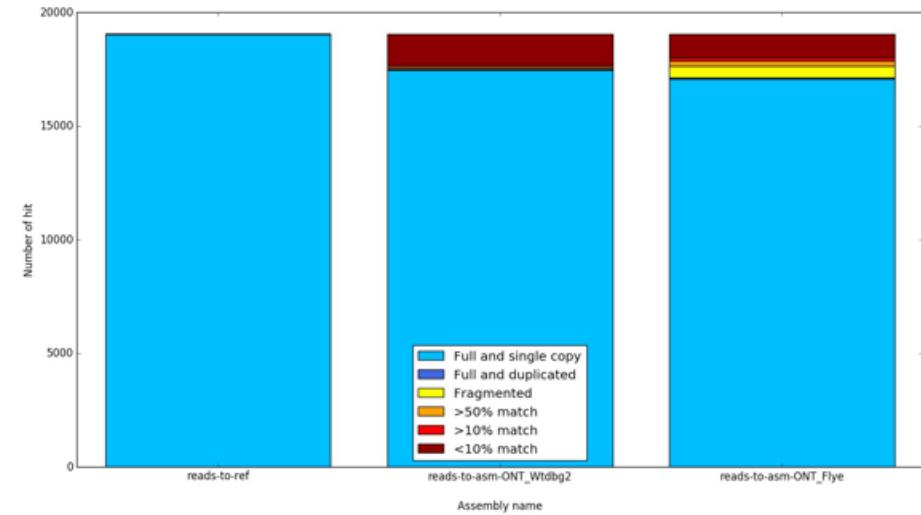


ASMGene Analysis



Pros :

- + Can produce BUSCO-like graphs
- + Analyses all reference transcripts

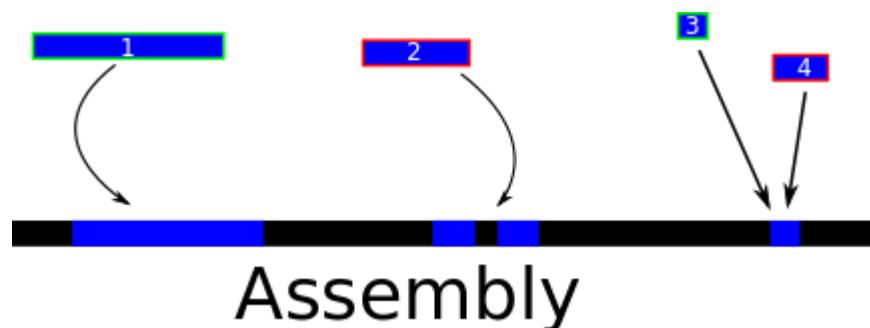


Cons :

- Have to fix a « hit » threshold on the reference
- Needs a reference assembly AND transcripts
- Needs several jobs to industrialize the analysis

ASMGene Homemade Analysis

- Align all annotated transcripts to an assembly
- Filter matches :
 - > 90 % identity → Good
 - > 0 % identity → Present
 - Missing
- Compare multiple assemblies on the same graph



Good : 1 – 3

Present : 2

Missing : 4

ASMGene Homemade Analysis



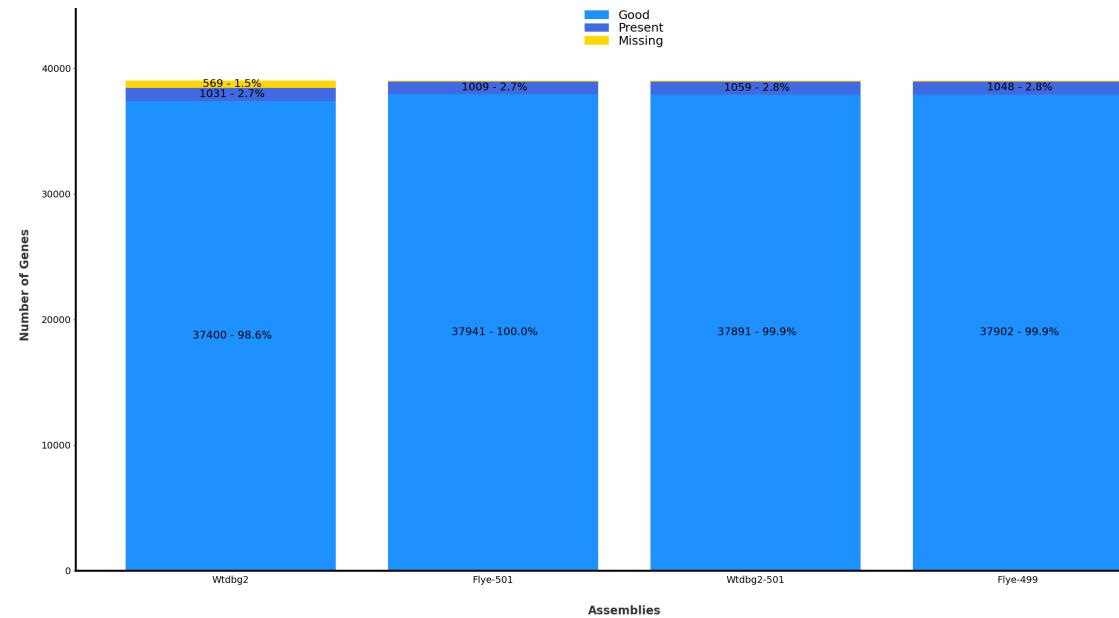
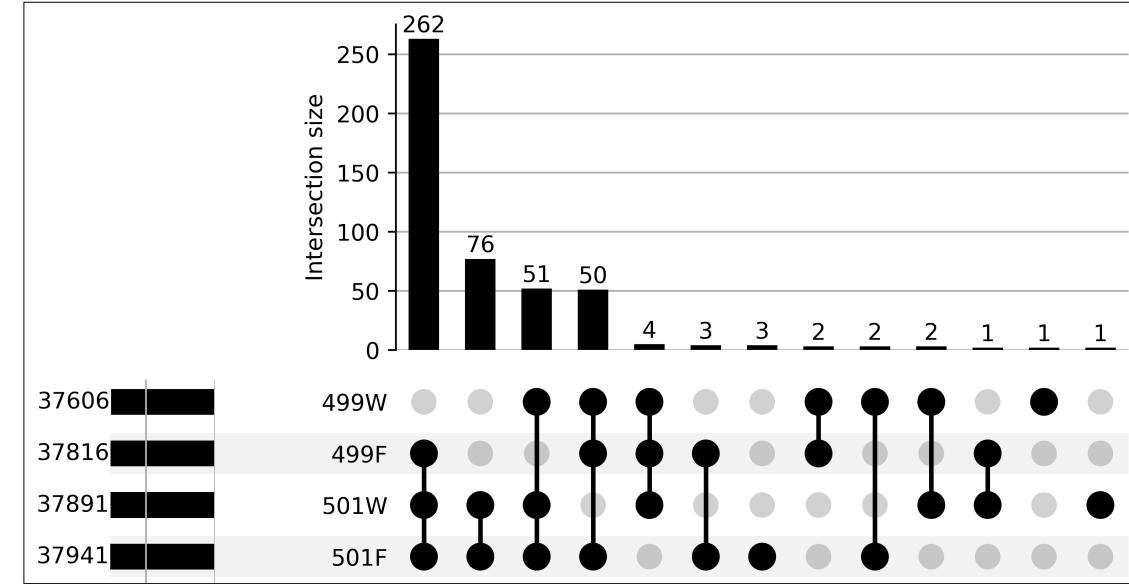
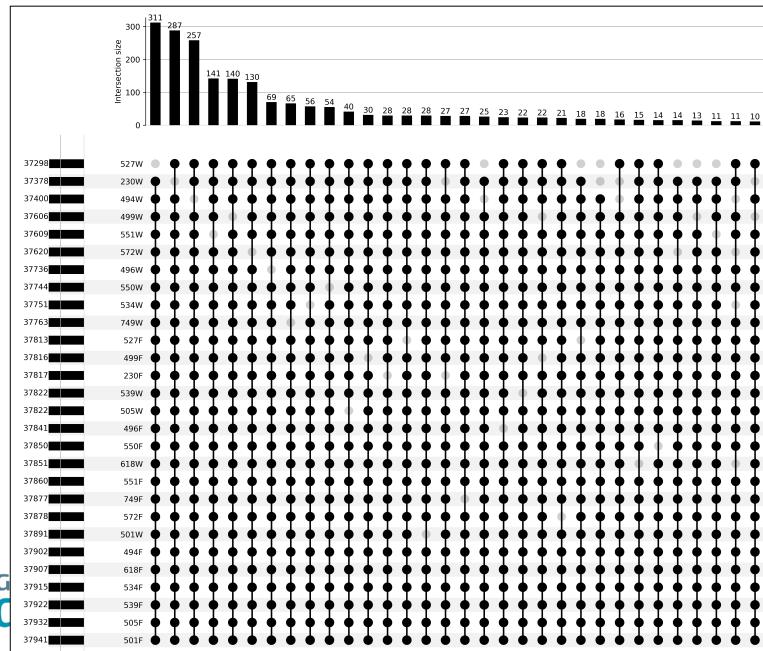
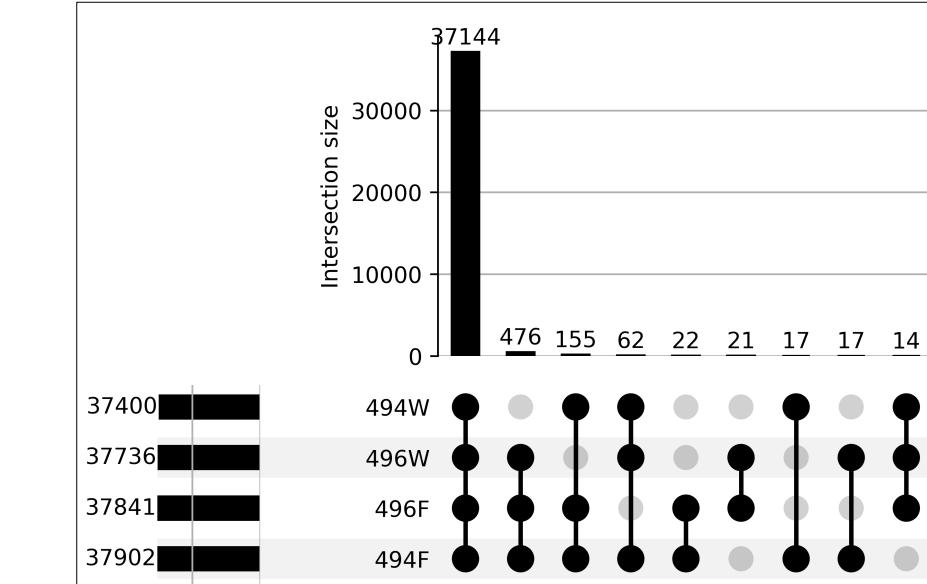
Pros :

- + Produce upset-plot, easy to compare assemblies
- + Analyse all reference transcripts
- + Not dependent on a reference assembly
- + Clearer graph than busco, but less informative (can be improved)

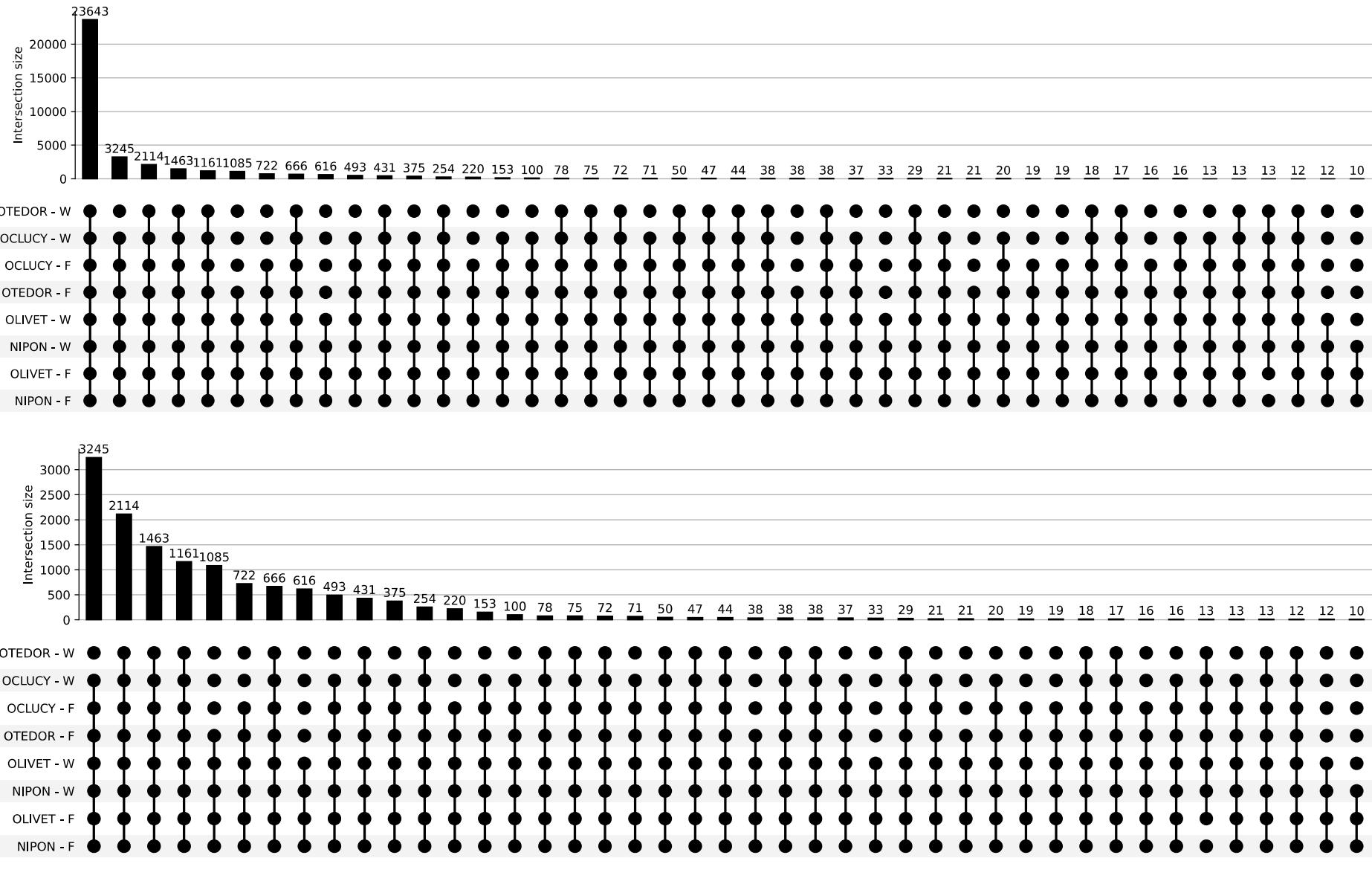
Cons :

- Need a reference transcripts set
- Need several jobs to industrialize the analysis

ASMGene Homemade Analysis - Exemple plot



ASMGene Homemade Analysis - Exemple plot



Conclusion

- BUSCO score is the simplest method
- BUSCO score is imprecise
- Homemade method is the most informative method
- Homemade method can be improved and adapted to each case

Partenaires du projet



Coordination

Cécile Donnadieu
Christine Gaspin
Carole lampietro
Denis Milan



Axe1 Génomique

Clément Birbes
Arnaud Di-Franco
Andreea Dreau
Camille Eché
Thomas Faraut
Carole lampietro
Christophe Klopp
Claire Kuchly
Camille Marcuzzo
Amandine Suin
Matthias Zytnicki

Axe2 Epigénétique

Remy Félix Serres
Paul Terzian
Celine Vandecasteele
Christophe Klopp

Axe3 Métagénomique

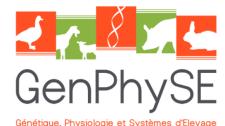
Adrien Castinel
Jean Mainguy
Olivier Bouchez
Géraldine Pascal
Claire Hoede

Axe 4 Transversal

Amandine Broha
Abdias-Archimede Towe-Patipe
Erwan Denis
Romain Therville
Didier Laborie
Céline Noirot
Gérald Salin
Marie-Stéphane Trotard

GenPhySE

Julie Demars
Cédric Cabau
Sylvie Combes
Patrice Dehais
Thomas Faraut
Katia Feve
Nathalie Iannucelli
Sophie Leroux
Géraldine Pascal
Frédérique Pitel



MIAT

Matthias Zytnicki

GABI

Didier Boichard
Mekki Boussaha
Sébastien Fritz
Cécile Grohs

Le Moulon

Alain Charcosset
Johann Joets
Delphine Madur
Stéphane Nicolas
Rémi Séraphin
Clémentine Vitte



Remerciements



Coordination

Cécile Donnadieu
Christine Gaspin
Carole lampietro
Denis Milan



Plus spécifiquement impliqués dans les travaux présentés aujourd'hui

Auteur 1

Auteur 2

Auteur 3

Auteur 4

Auteur 5

Auteur 6

