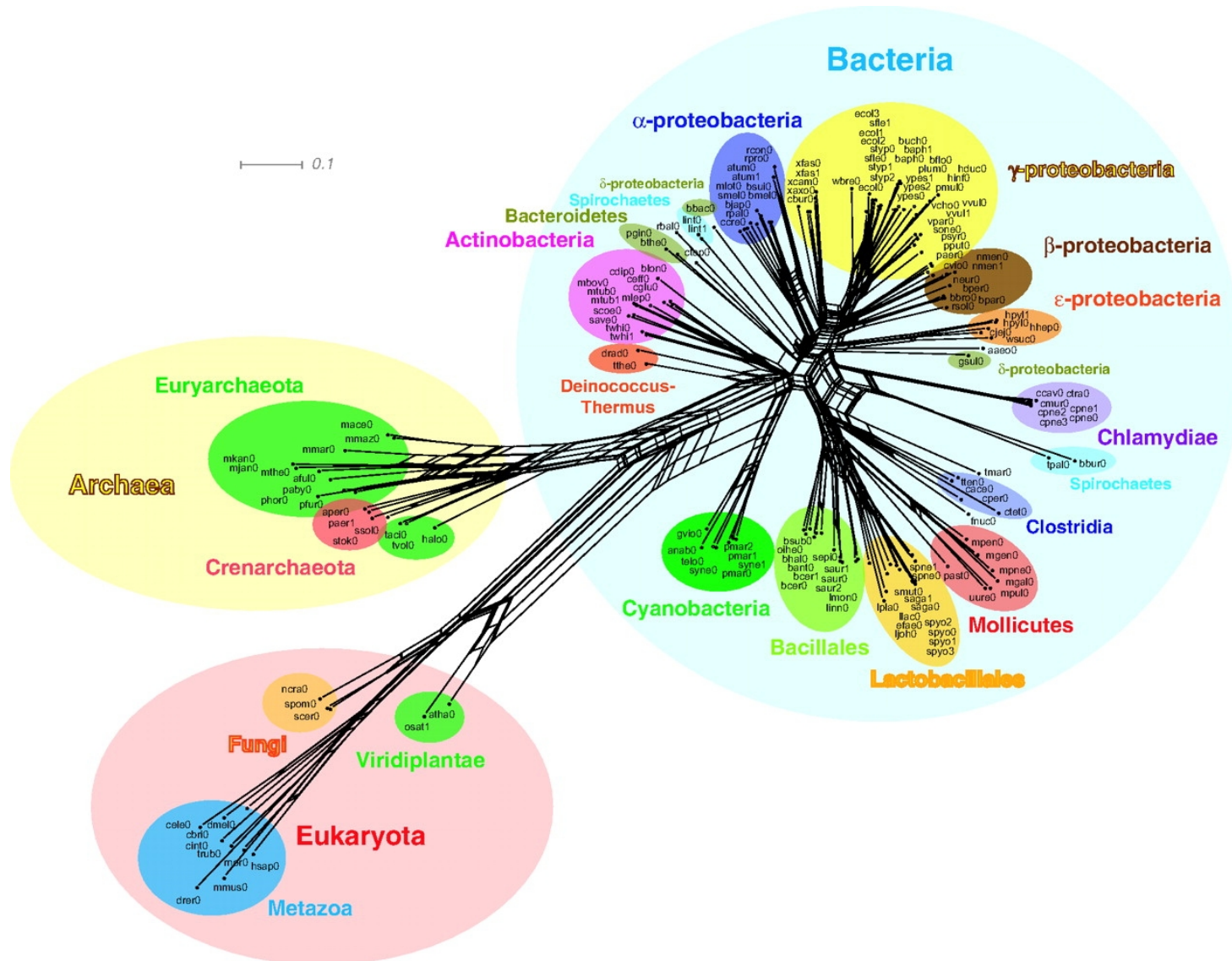


# ARNnc chez les Archae

## Nature des ARNnc connus & méthodes de prédiction



# ARNnc chez les Archae : Rnomics & Bioinformatique

tRNA, rRNA, RNaseP

1999-2000

ARNnc à boîtes C/D chez *Pyrococcus*

- Gaspin et al. J.Mol. Biol., 2000
- Omer et al. Science, 2000

# ARNnc chez les Archae : Rnomics & Bioinformatique

ARNt, ARNr, RNaseP

1999-2000

ARNnc à boîtes C/D chez *Pyrococcus*

- Gaspin et al. J.Mol. Biol., 2000
- Omer et al. Science, 2000

2001-2002

Approches à l'échelle du génome

- Bioinformatique

Klein et al., PNAS, 2002

Schattner, NAR, 2002

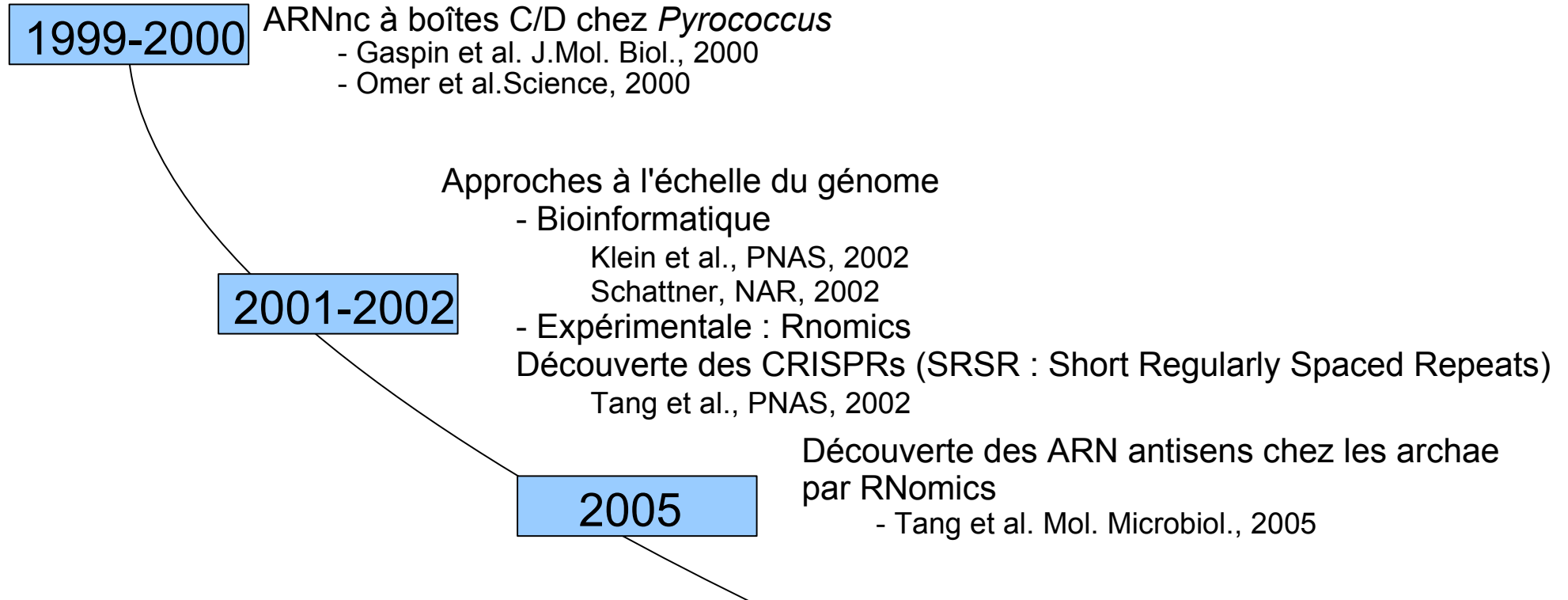
- Expérimentale : Rnomics

Découverte des CRISPRs (SRSR : Short Regularly Spaced Repeats)

Tang et al., PNAS, 2002

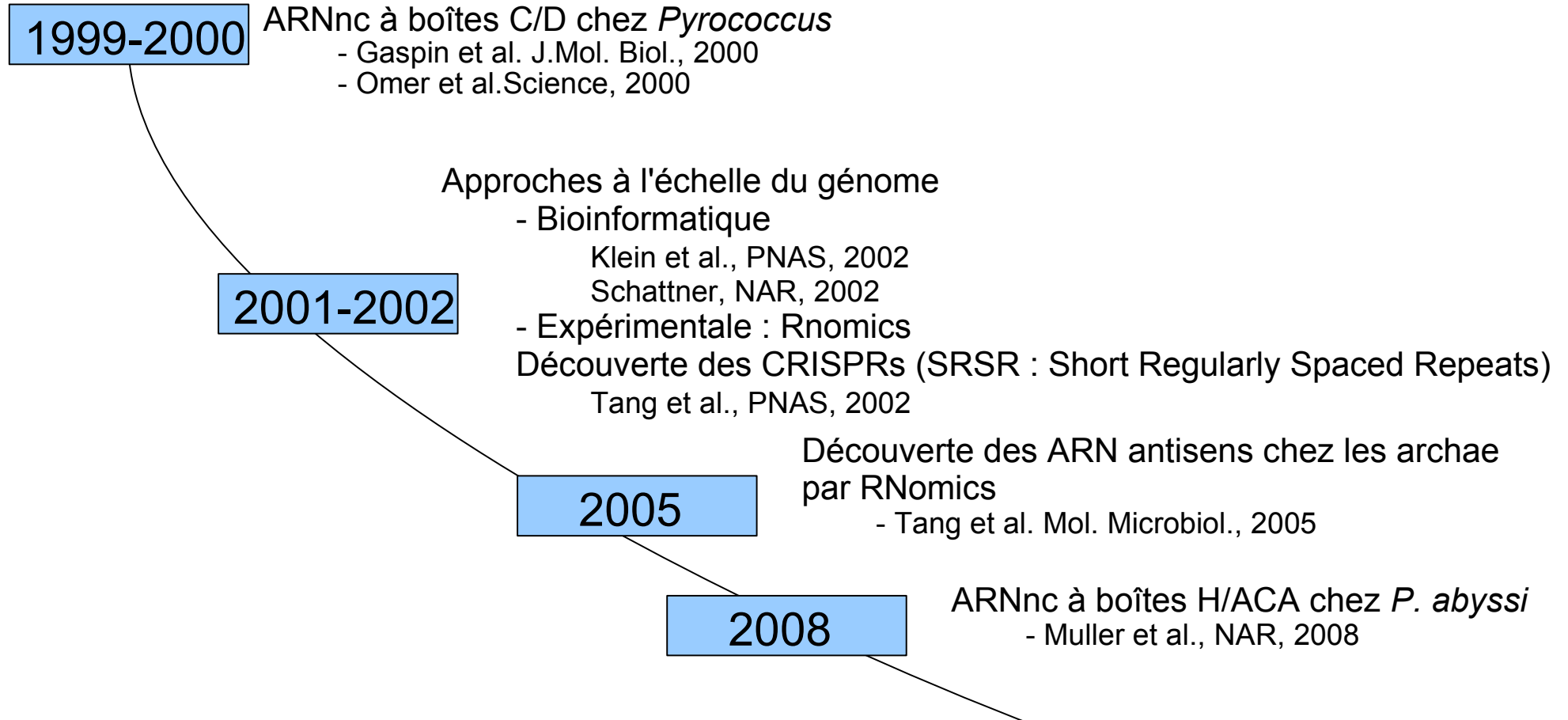
# ARNnc chez les Archae : Rnomics & Bioinformatique

ARNt, ARNr, RNaseP



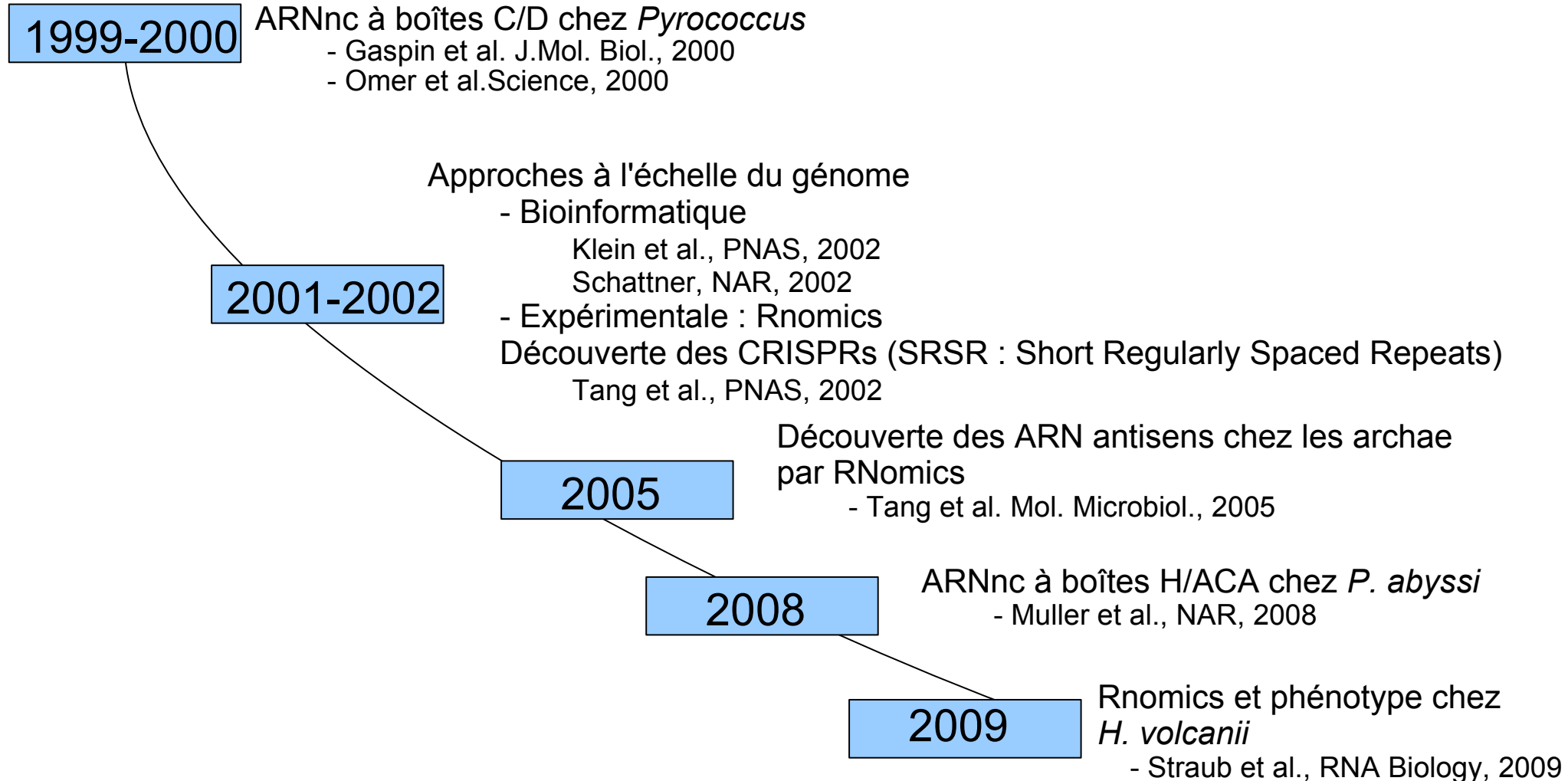
# ARNnc chez les Archae : Rnomics & Bioinformatique

ARNt, ARNr, RNaseP



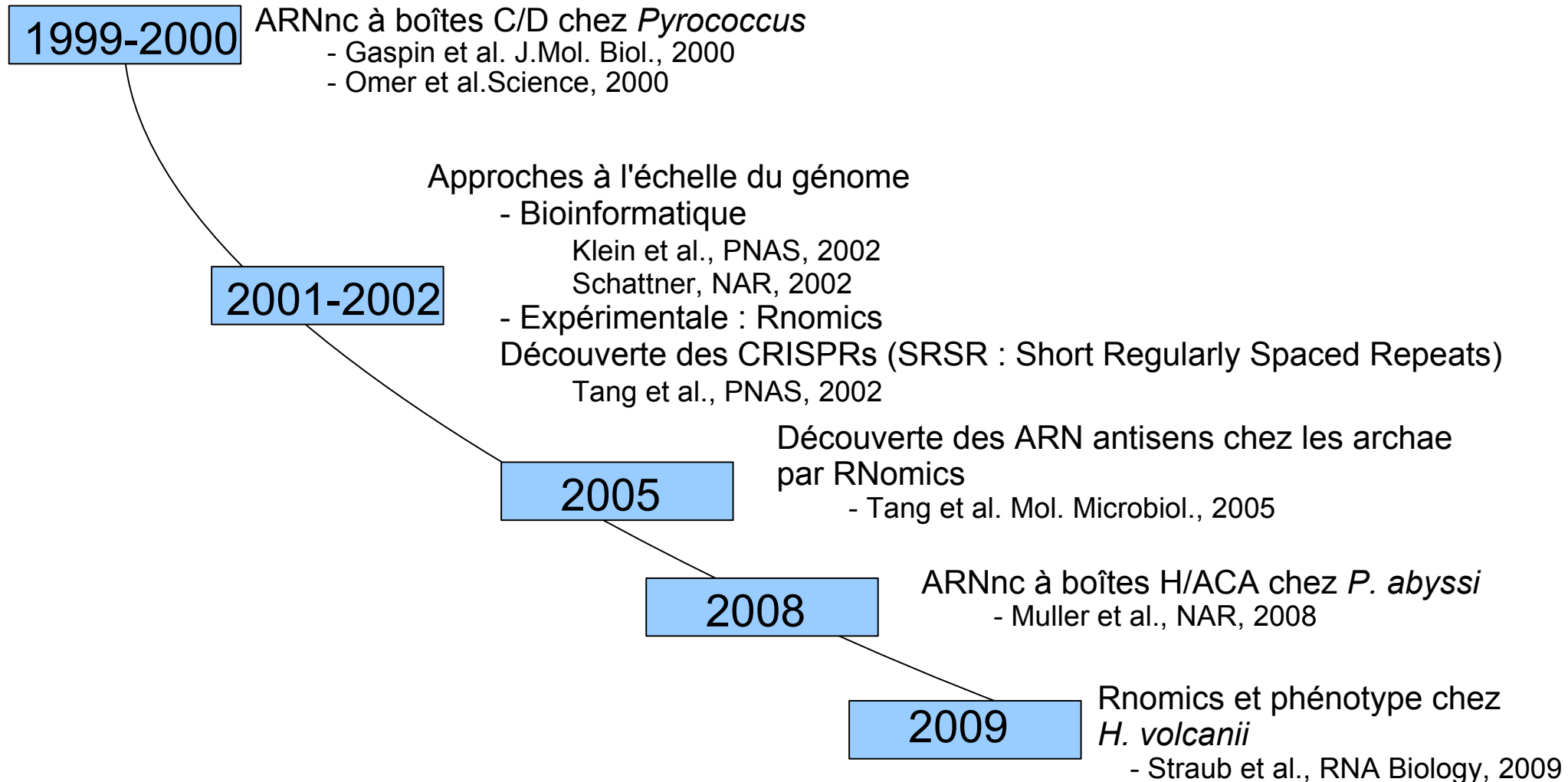
# ARNnc chez les Archae : Rnomics & Bioinformatique

ARNt, ARNr, RNaseP



# ARNnc chez les Archae : Rnomics & Bioinformatique

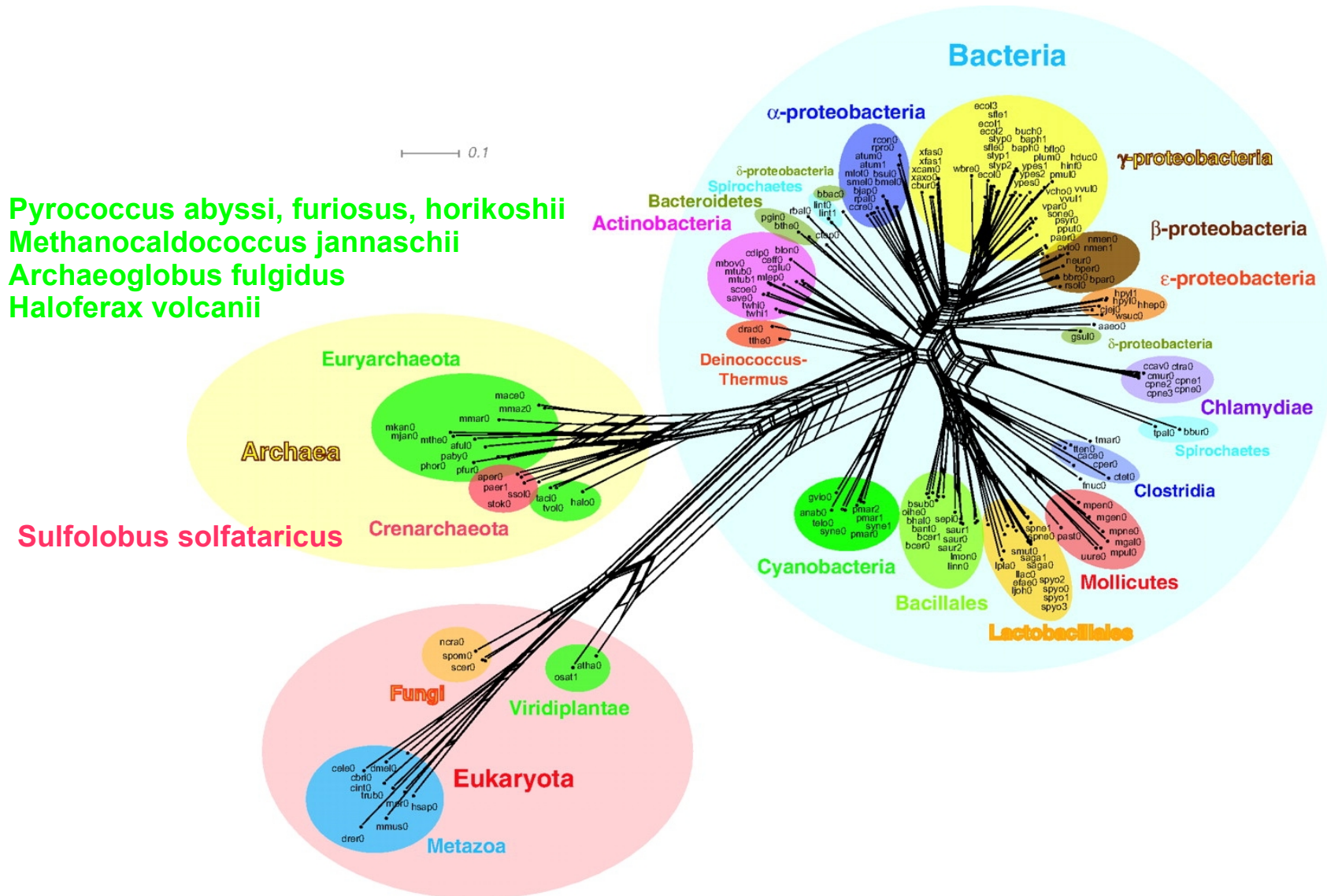
ARNt, ARNr, RNaseP



**Haut débit ?**

# ARNnc chez les Archae : quels organismes étudiés ?

*Pyrococcus abyssi*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*  
*Methanocaldococcus jannaschii*  
*Archaeoglobus fulgidus*  
*Haloferax volcanii*



Fukami-Kobayashi, K. et al. Mol Biol Evol 2007 24:1181-1189



# **Approches expérimentales à l'échelle du génomome complet : RNomics**

# Caractéristiques des ARNnc candidats trouvés par RNomics

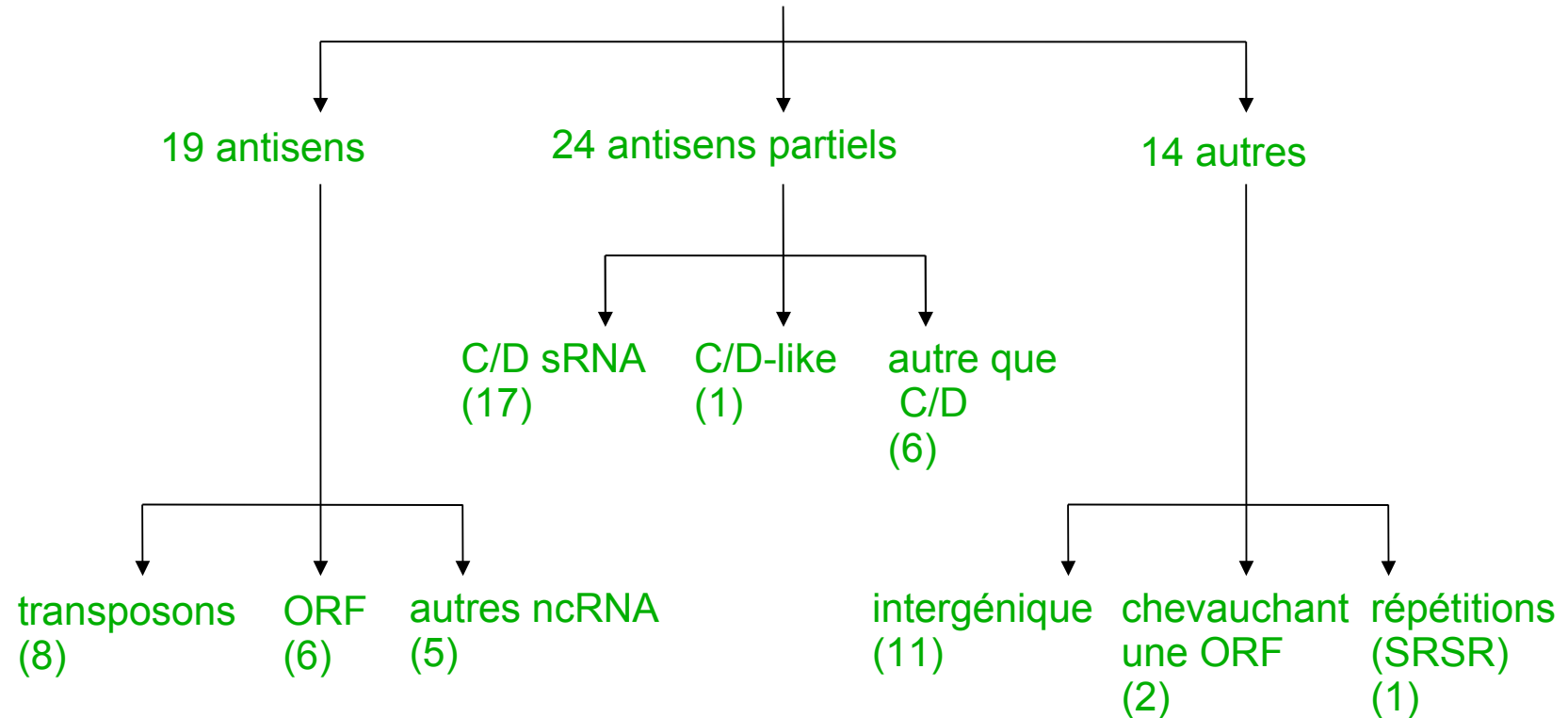
Taille : 50-500nt

86 candidats Tang TH et al. 2002 PNAS, *A. fulgidus*

45 candidats Zago et al. Mol. Microbiol., 2005, *S. solfataricus*

**57 candidats** Tang et al. Mol. Microbiol, 2005, *S. solfataricus*

39 candidats, Straub et al., RNA Biology, 2009, *H. volcanii*



# Caractéristiques des ARNnc candidats trouvés par RNomics

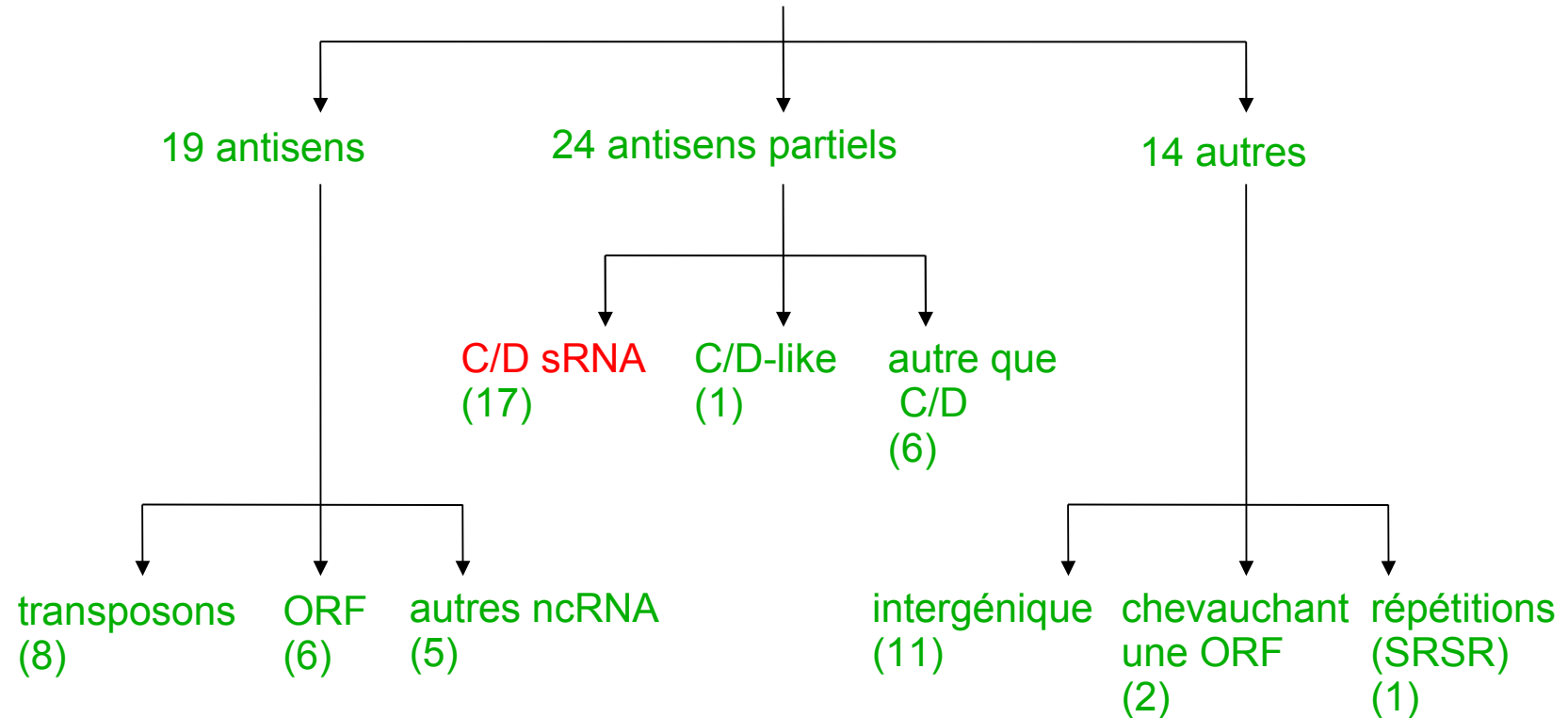
Taille : 50-500nt

86 candidats Tang TH et al. 2002 PNAS, *A. fulgidus*

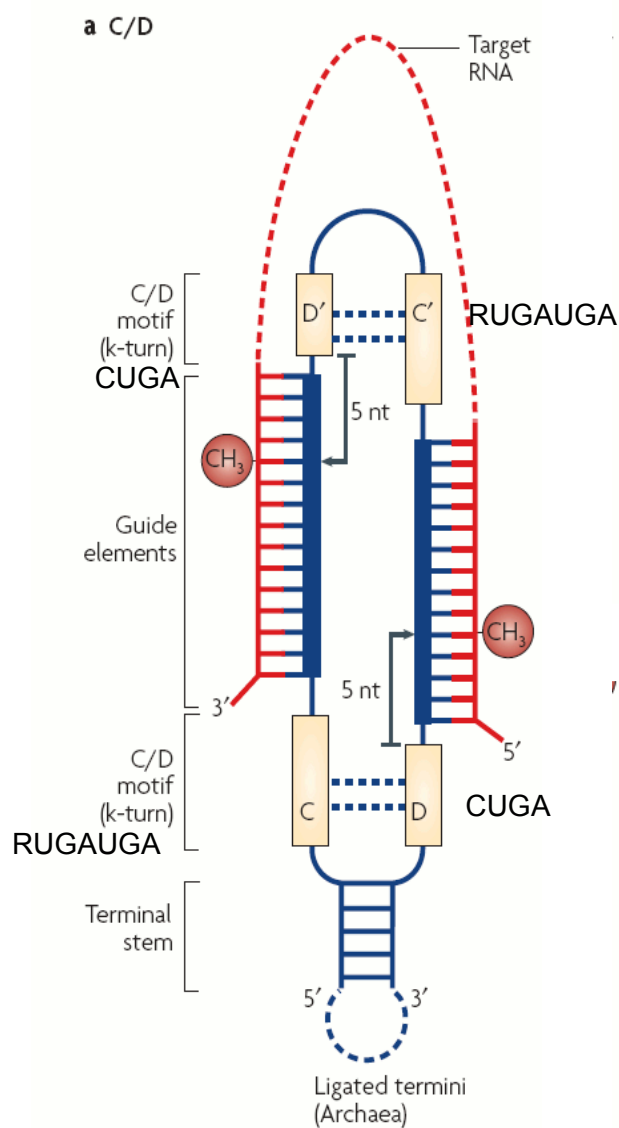
45 candidats Zago et al. Mol. Microbiol., 2005, *S. solfataricus*

**57 candidats** Tang et al. Mol. Microbiol, 2005, *S. solfataricus*

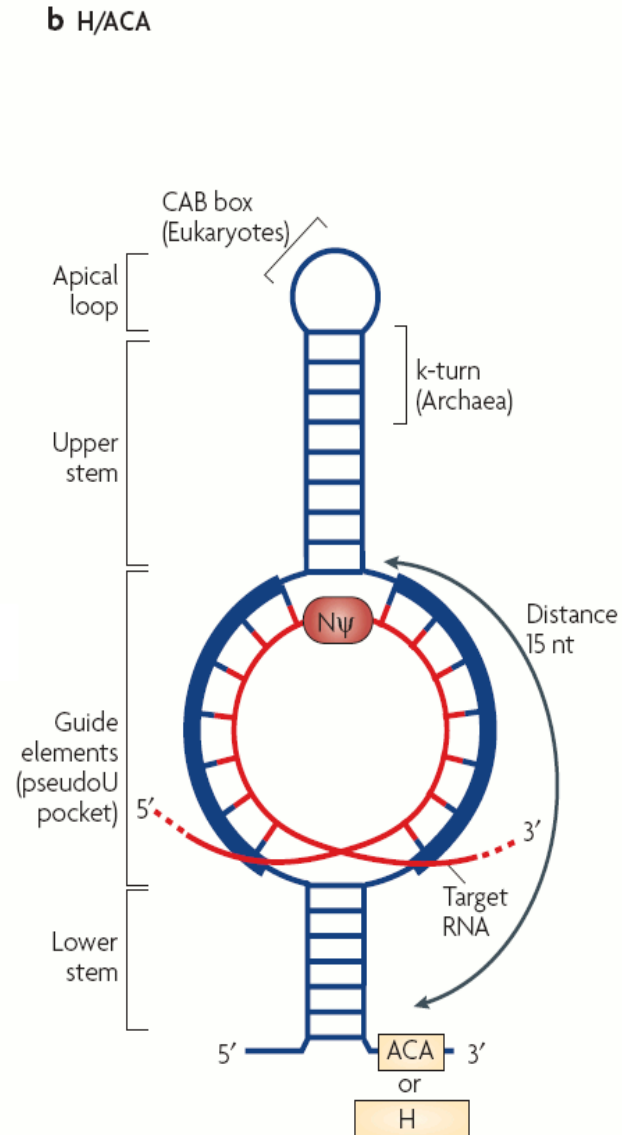
39 candidats, Straub et al., RNA Biology, 2009, *H. volcanii*



# Exemples d'ARN antisens partiels de fonction connue « Motifs » caractéristiques des familles



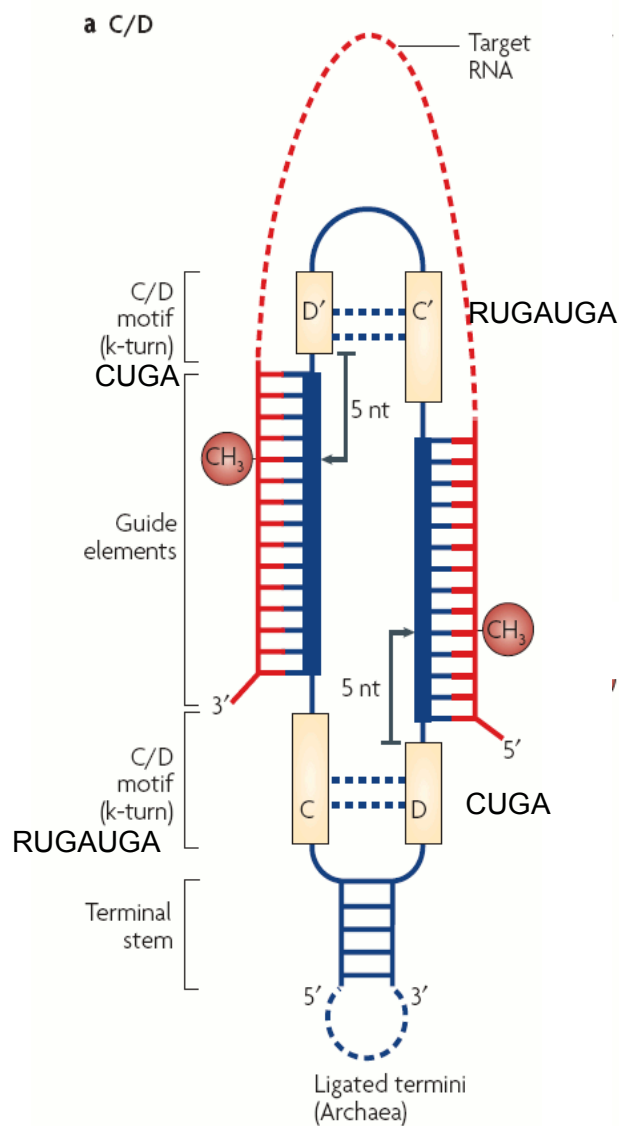
ARNnc à boîtes C/D



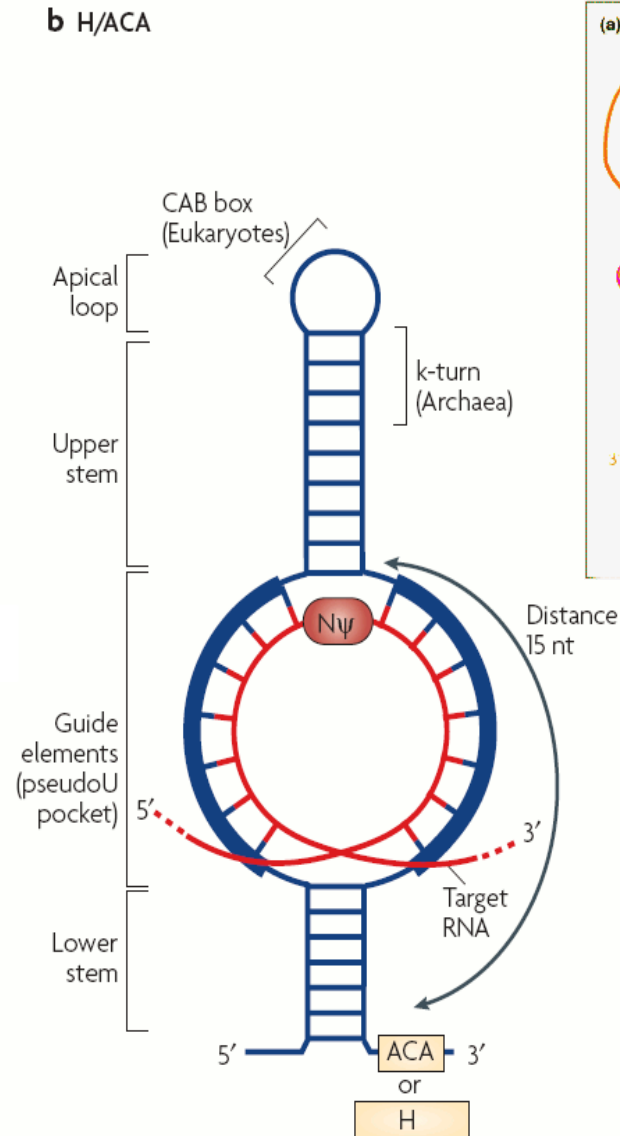
ARNnc à boîte H/ACA

# Exemples d'ARN antisens partiels de fonction connue

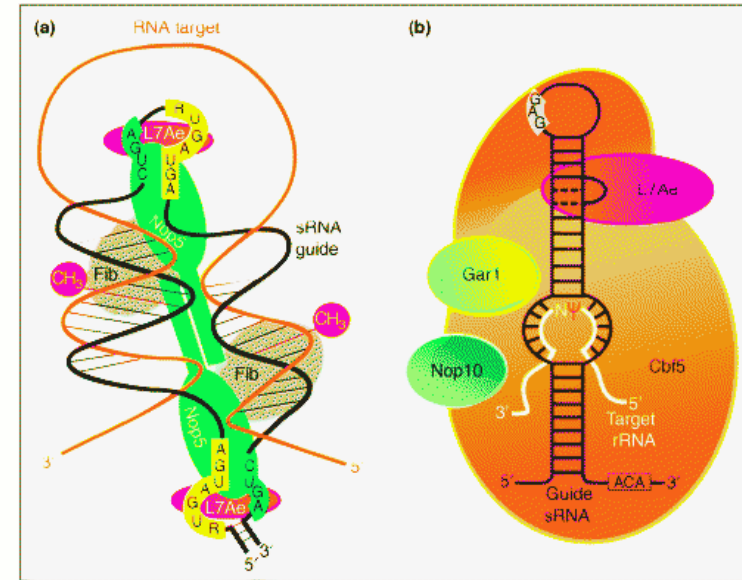
## « Motifs » caractéristiques des familles



ARNnc à boîtes C/D



ARNnc à boîte H/ACA



# Caractéristiques des ARNnc candidats trouvés par RNomics

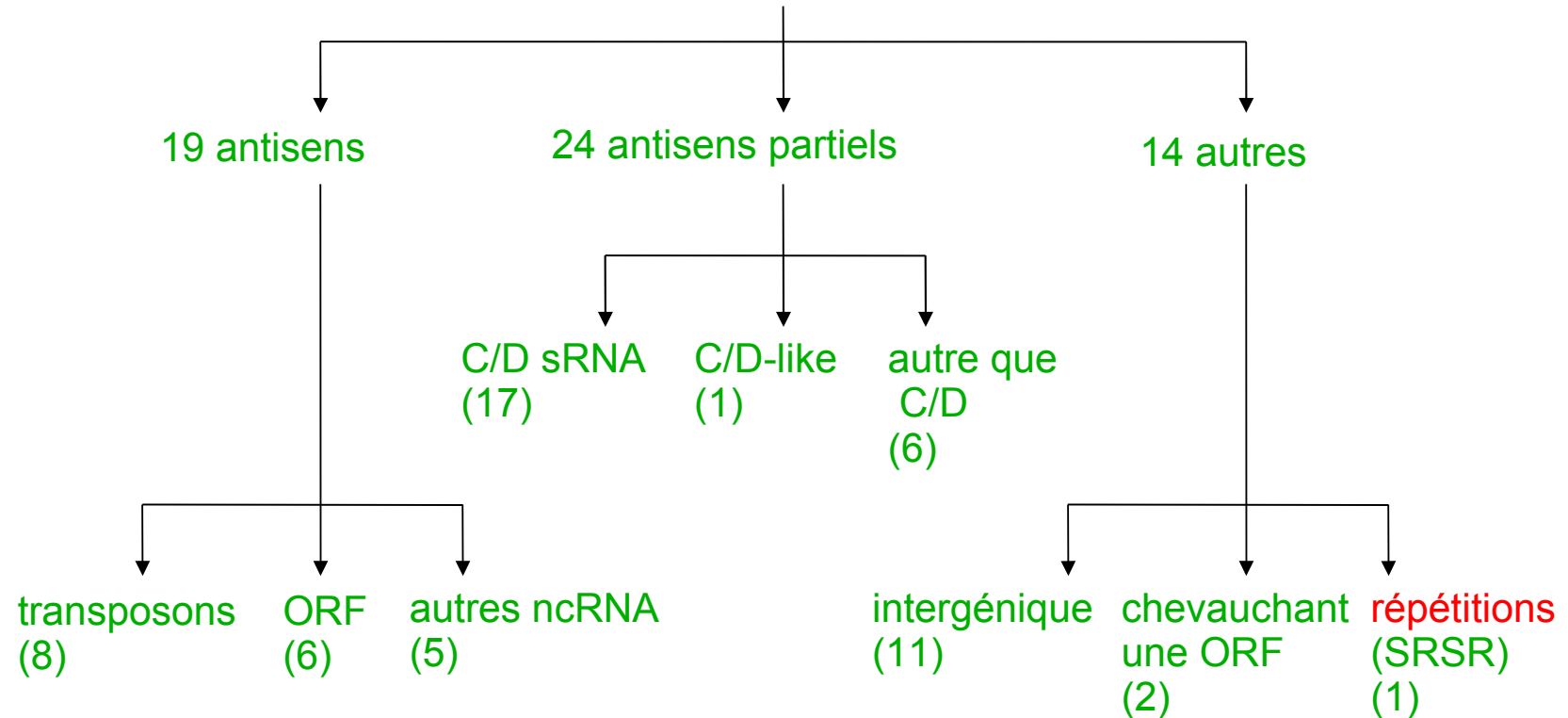
Taille : 50-500nt

86 candidats Tang TH et al. 2002 PNAS, *A. fulgidus*

45 candidats Zago et al. Mol. Microbiol., 2005, *S. solfataricus*

**57 candidats** Tang et al. Mol. Microbiol, 2005, *S. solfataricus*

39 ncandidates, Straub et al., RNA Biology, 2009, *H. volcanii*



# CRISPRs : Clustered Regularly Interspaced Short Palindromic Repeats ou SRSR : Short Regularly Spaced Repeats

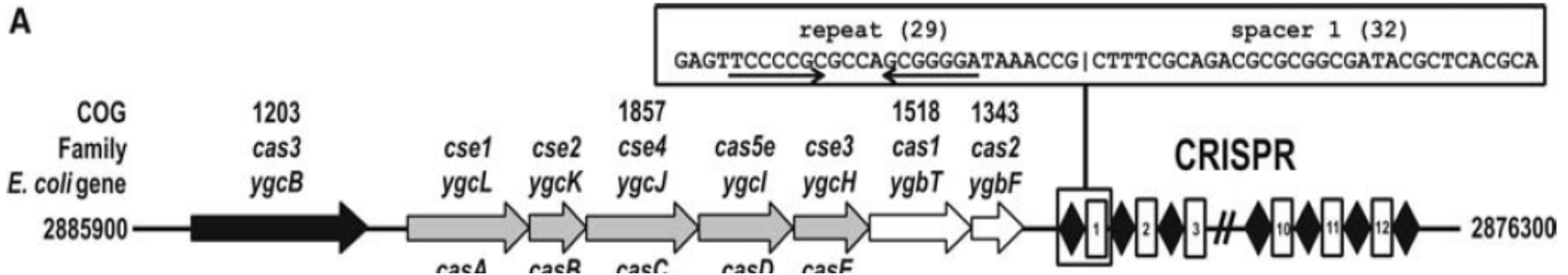


Figure tirée de Brouns et al., Science, 2008

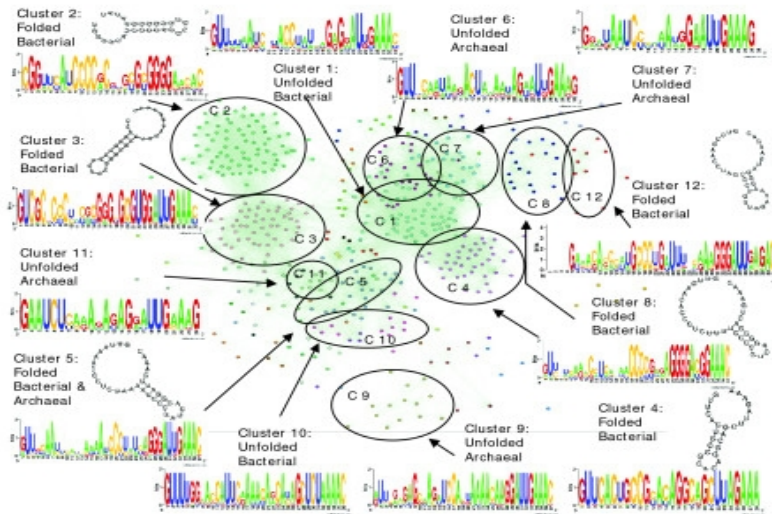


Figure tirée de Kunin et al., Genome Biology, 2007

**Protéines CAS ou  
Cluster de séquences répétées ?**

# Caractéristiques des sRNA candidats trouvés par RNomics

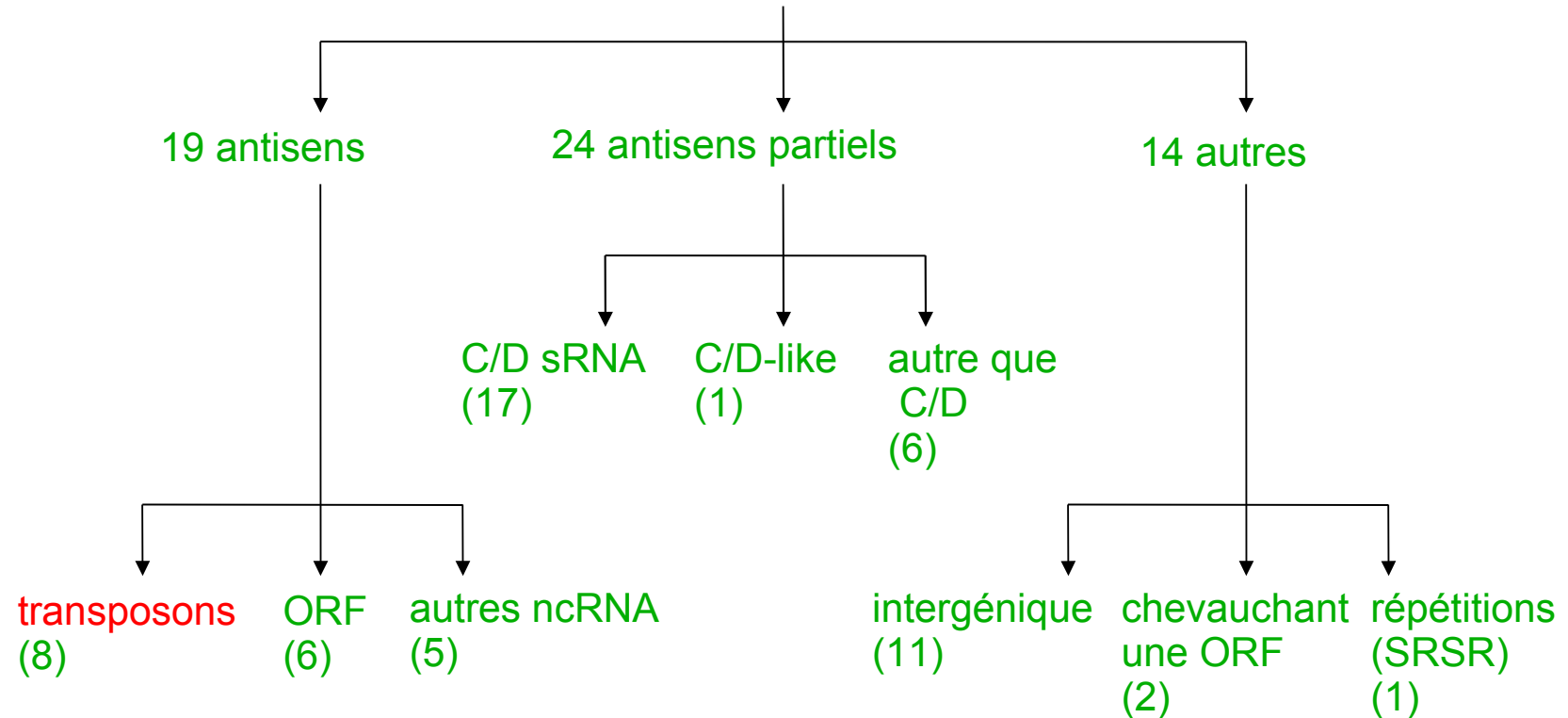
Taille : 50-500nt

86 ncRNA candidats Tang TH et al. 2002 PNAS, *A. fulgidus*

45 ncRNA candidats Zago et al. Mol. Microbiol., 2005, *S. solfataricus*

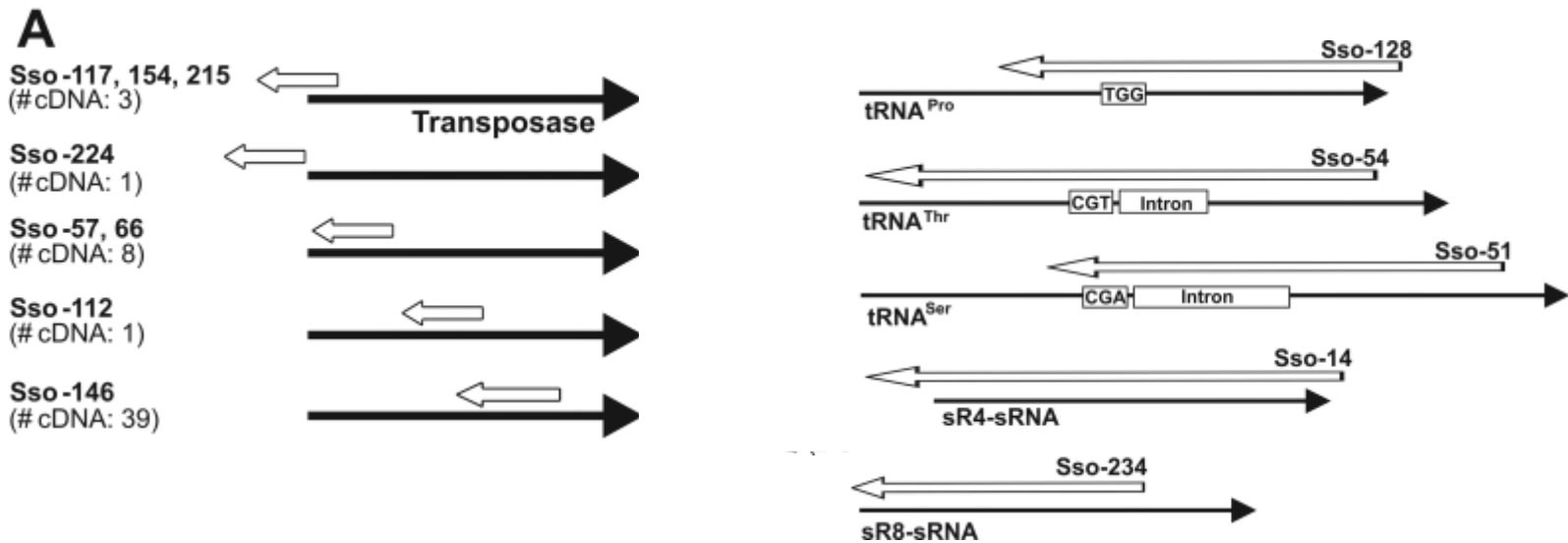
**57 ncRNA candidats** Tang et al. Mol. Microbiol, 2005, *S. solfataricus*

39 ncRNA candidates, Straub et al., RNA Biology, 2009, *H. volcanii*





# Représentation schématique des ARN antisens



# Caractéristiques des sRNA candidats trouvés par RNomics

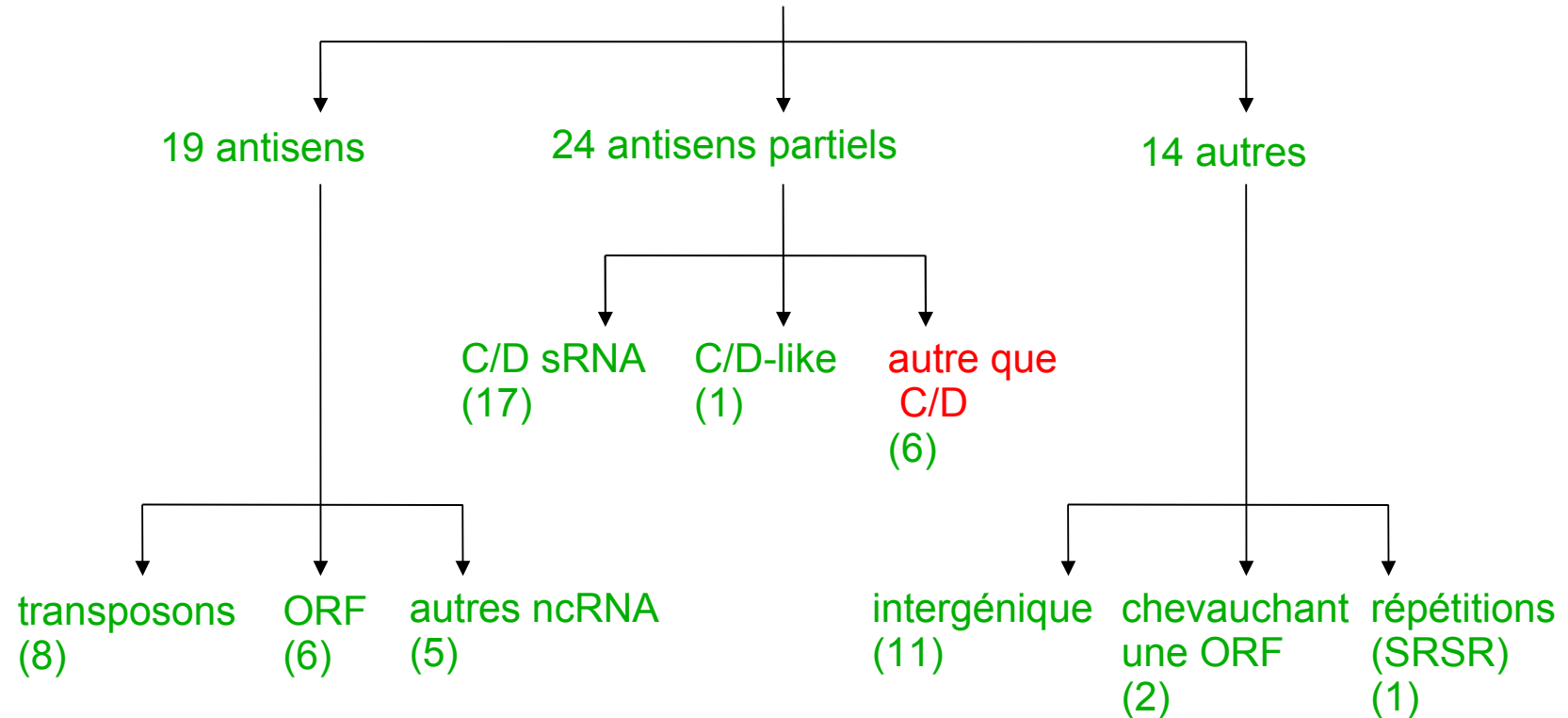
Taille : 50-500nt

86 ncRNA candidats Tang TH et al. 2002 PNAS, *A. fulgidus*

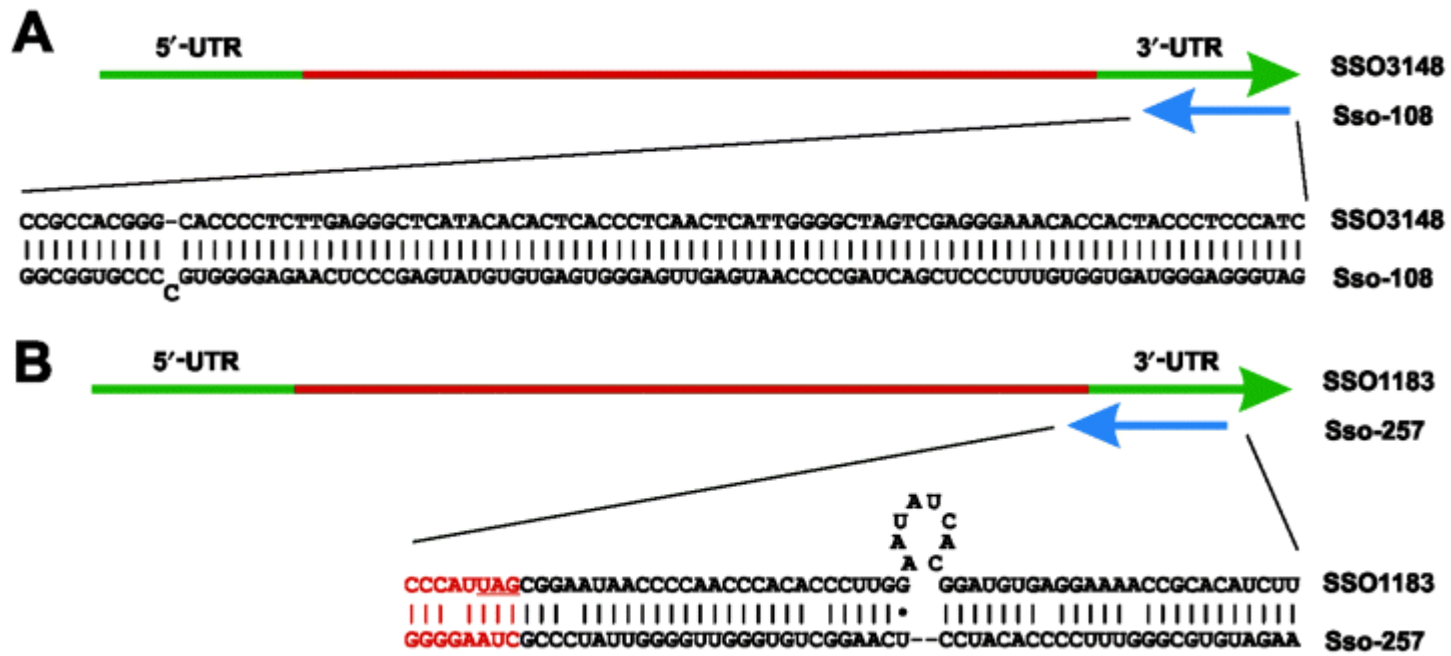
45 ncRNA candidats Zago et al. Mol. Microbiol., 2005, *S. solfataricus*

**57 ncRNA candidats** Tang et al. Mol. Microbiol, 2005, *S. solfataricus*

39 ncRNA candidates, Straub et al., RNA Biology, 2009, *H. volcanii*



# Représentation schématique des ARN antisens partiels



# Caractéristiques des sRNA candidats trouvés par RNomics

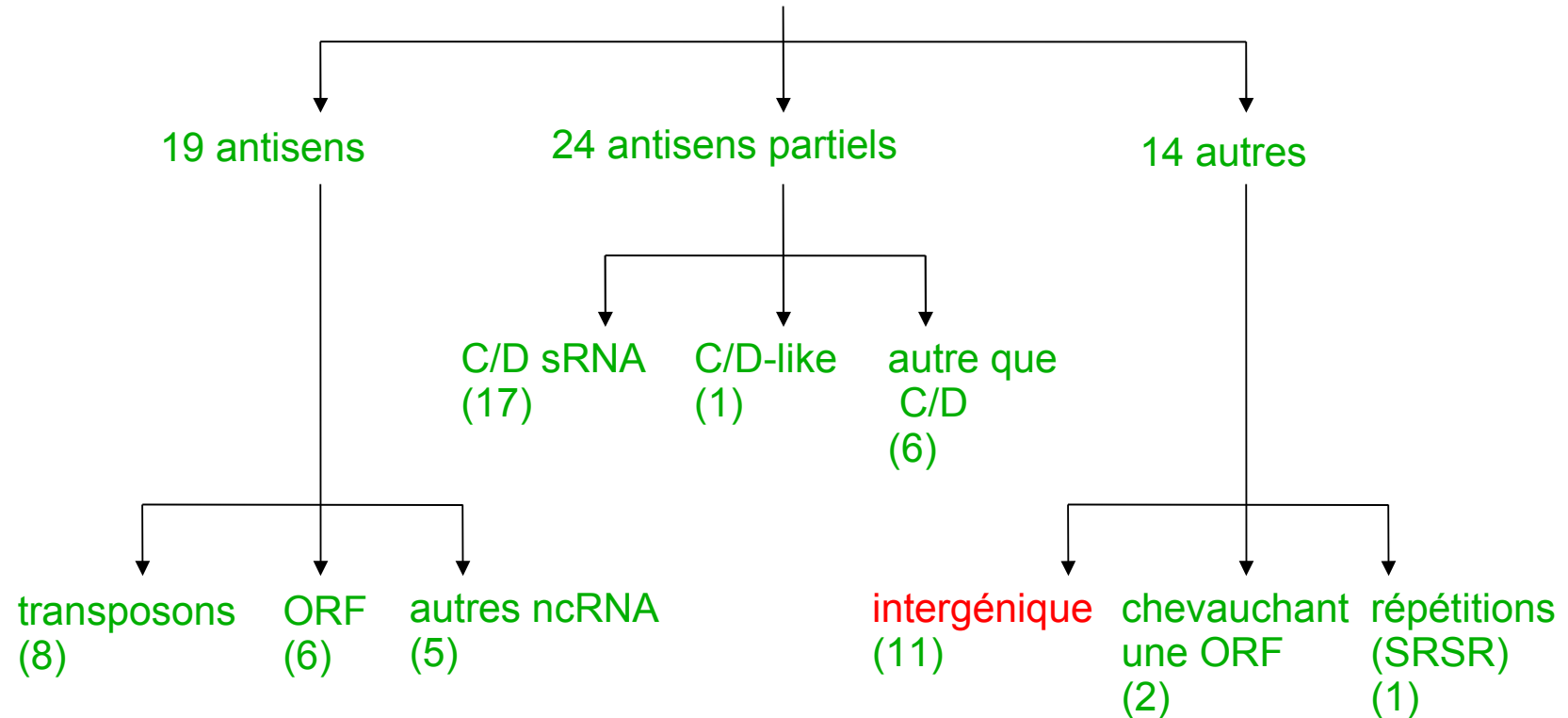
Taille : 50-500nt

86 ncRNA candidats Tang TH et al. 2002 PNAS, *A. fulgidus*

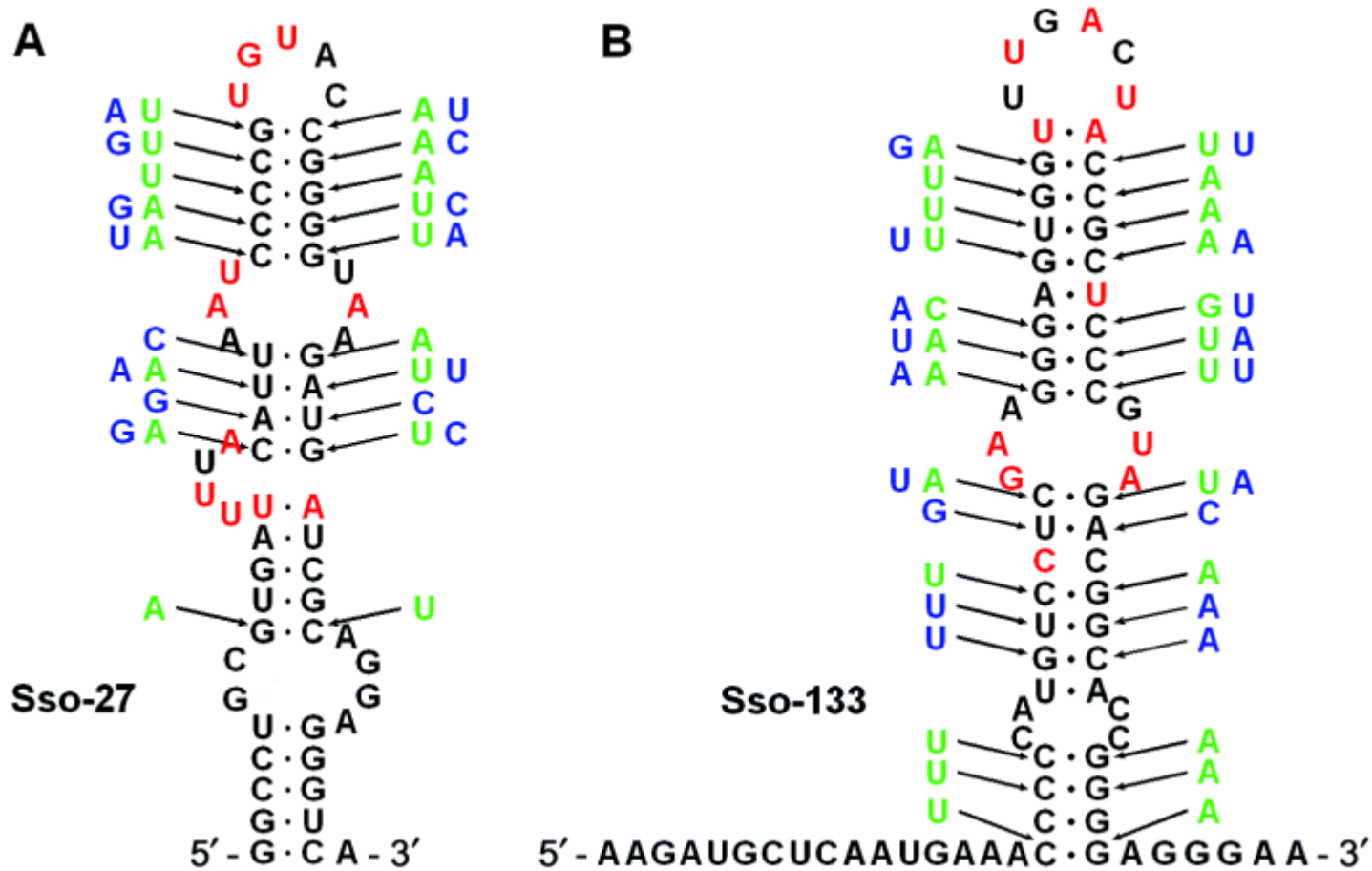
45 ncRNA candidats Zago et al. Mol. Microbiol., 2005, *S. solfataricus*

**57 ncRNA candidats** Tang et al. Mol. Microbiol, 2005, *S. solfataricus*

39 ncRNA candidates, Straub et al., RNA Biology, 2009, *H. volcanii*



# Exemples de conservation d'ARN localisé dans une région intergénique



**Vert** : conservé dans *S. tokodaii*  
**Bleu** : conservé dans *S. acidocaldarius*  
**Rouge** : conservé dans les 3 organismes

# **Approches bioinformatique à l'échelle du génomome complet**

# Essentiellement 3 approches complémentaires

- Recherche de motifs d'ARNnc sur la base de familles connues : du spécifique au généraliste
  - Famille répertoriée d'ARNnc, nécessite la conservation d'éléments de séquences et/ou de structure, autorise des « déviations »
- Analyse comparative de génomes
  - Existence d'un ou plusieurs génomes apparentés, hypothèse de l'existence d'ARNnc communs aux organismes analysés
- Approche *ab initio* : biais de composition, promoteurs & terminateurs orphelins, motifs (K-turn,...) ...
  - Pas d'organisme apparenté, accès à la spécificité, exploite les propriétés des ARN et du génome étudié

# Recherche d'ARNnc candidats

## Existence de familles connues

- Exploite les conservations connues ou « apprises » de la famille considérée
  - Via un outil spécifique :
    - tRNAscan-SE pour la recherche d'ARNt  
Lowe & Eddy, NAR, 1997
  - Via des outils exploitant directement les alignements
    - Infernal, Erpin (Lambert & Gautheret, Muller et al., NAR, 2008)
  - Via des outils généralistes
    - MilPat (Thébault et al., Bioinformatics, 2006), DARN! (Zytnicki et al., Constraints, 2008)...



# Recherche d'ARNnc candidats

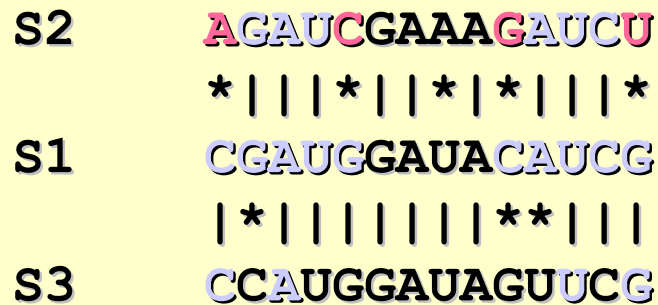
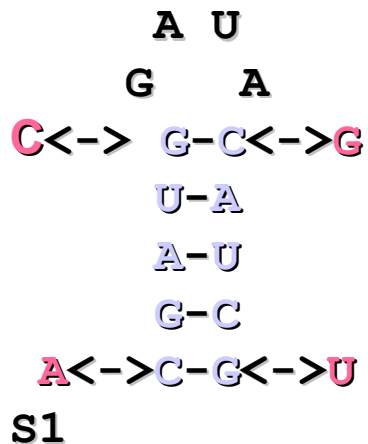
## Analyse comparative

- Exploite les conservations en séquence & en structure :  
QRNA
  - Recherche les conservations en séquences
  - Sélectionne les alignements 2 à 2 sur la base de 3 modèles :  
codant, RNA, autres
  - Utilisation de QRNA chez les *Pyrococcus*
    - 17 candidats trouvés, 4 validés par Northern, dont 3 ARNnc à boîtes  
H/ACA (Klein & Eddy, PNAS, 2002)

# Recherche d'ARNnc candidats

## Analyse comparative

- Exploite les conservations en séquence & en structure : QRNA
  - Recherche les conservations en séquences
  - Sélectionne les alignements 2 à 2 sur la base de 3 modèles



# Recherche d'ARNnc candidats

## Analyse comparative

Codant

<b>G</b>	<b>Q</b>	<b>K</b>	<b>V</b>	<b>L</b>								
GGT	CAG	AAAG	TACTT									
		*				*			*			*
GGAC	CAG	AAG	GGTT	CTC								

ARN

TT	G	T	C	G	A	A	G	A	A	C	G	
			*	*					*	*		
TT	G	A	C	C	G	A	A	G	G	T	C	G

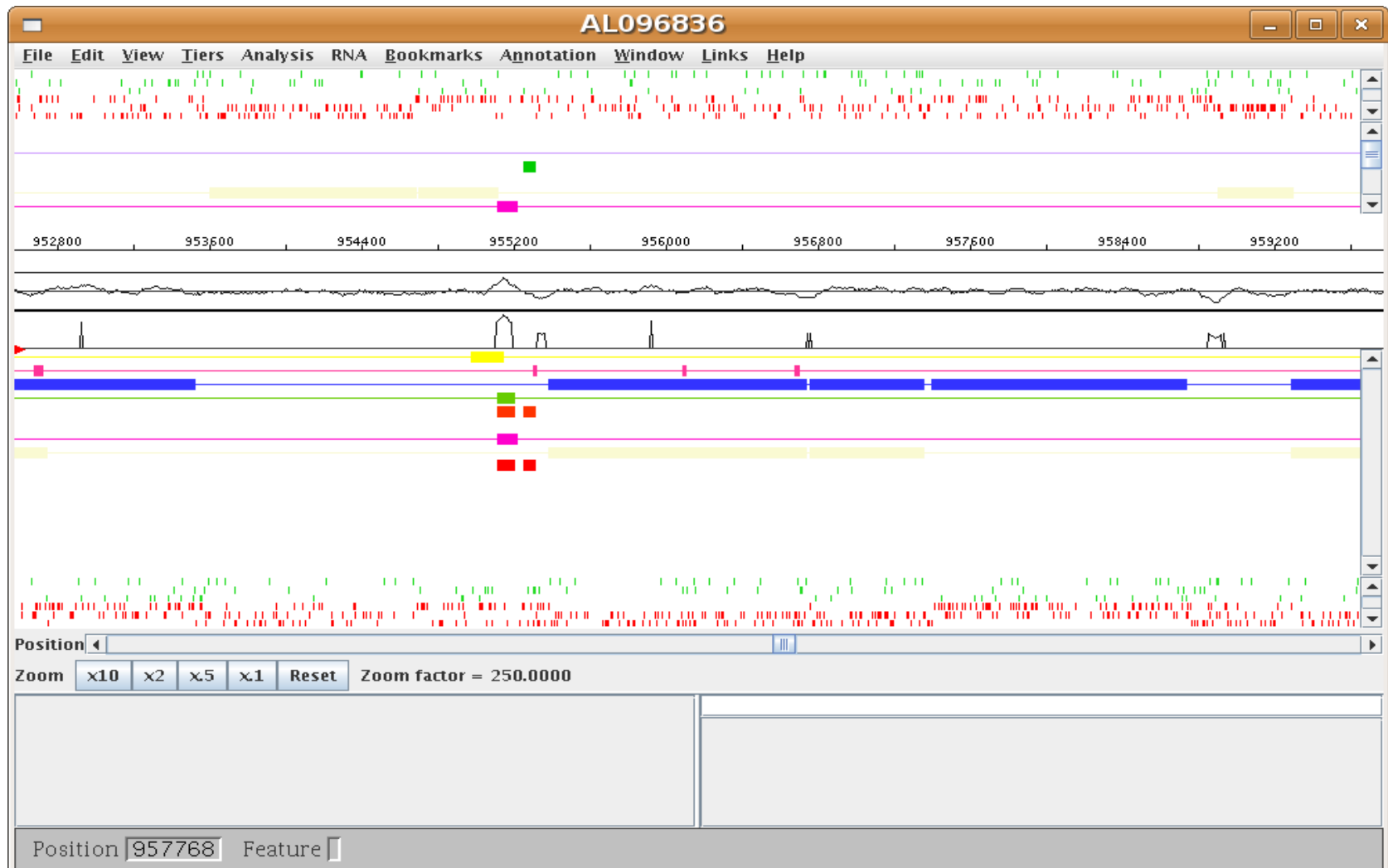
Autre

GT	T	A	A	C	T	G	A	G	T	A	A	C	G	
	*	*		*						*				
G	C	A	A	G	C	T	G	A	G	T	T	A	C	G

# Recherche d'ARNnc candidats

## Analyse « ab initio »

- Exploitation du biais de composition dans les séquences



# Recherche d'ARNnc candidats

## Analyse « ab initio »

- Exploitation du biais de composition dans les séquences

Ex : une archae AT-riche hyperthermophile, *M. Jannaschi*

GC% Génome : 32%

GC% ARN : 64%

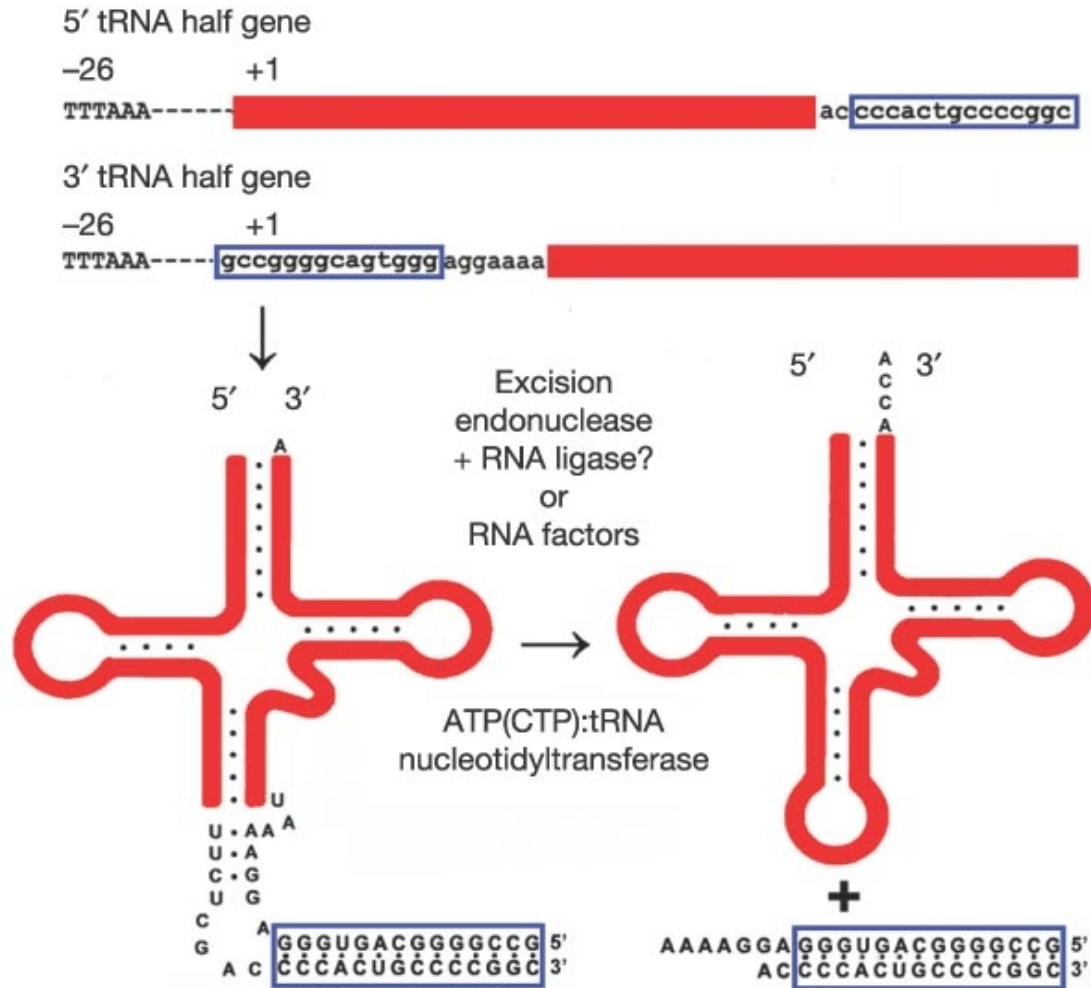
CMC*	LgPred <sup>□</sup>	LgRéelle	GC%	Northern	**%GC
Mj1	84	?	64		
Mj2	101	129	63	Oui	cnr1
Mj3	68	127	65	Oui	cnr3
Mj4	54	?	72		cnr4
Mj5	81	?	63		
Mj6/Mj6a	117	129	64	Oui	cnr12
Mj7	69	127	70	Oui	cnr14
Mj8	61	?	67		
Mj9	70	?	73		cnr19
	142	?	61		Cnr5
	119	?	56.1		cnr7
	67	?	53		Cnr8

□Candidats de longueur prédite > 50

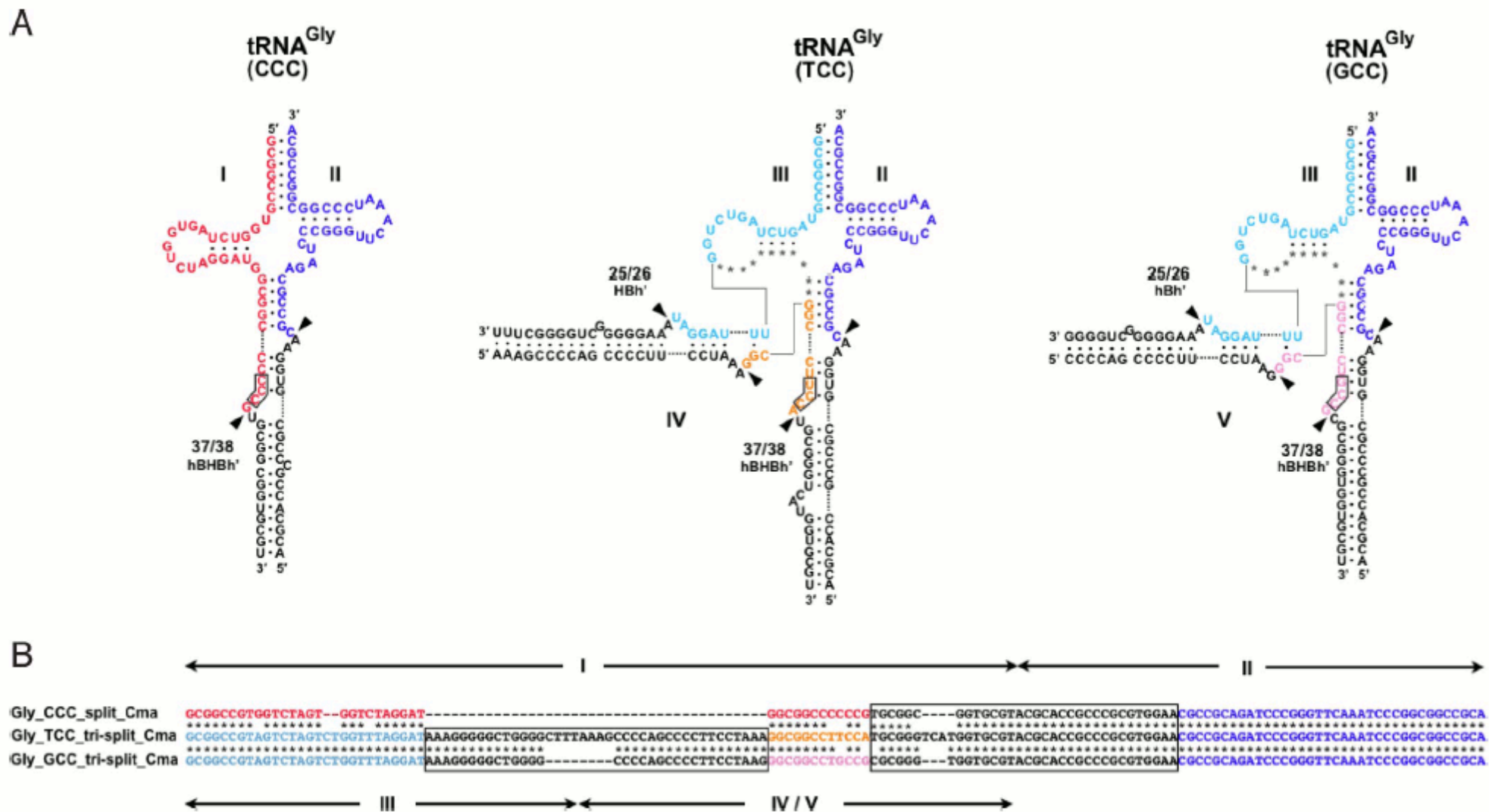
\*Klein & Eddy, PNAS, 2002

\*\*Schattner, NAR, 2002

# Pour conclure...



- Randau L et al. [2005](#) *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5'- and 3'-halves. *Nature*, 433(7025):537-41.



**Fig. 1.** Alternative splicing of 5 split tRNA<sup>Gly</sup> transcripts in *C. maquilingensis*. (A) Schematic diagrams of the 3 novel tRNA<sup>Gly</sup> assembly models. Three potential mature joined tRNAs formed by alternative assembly of 5 split tRNA<sup>Gly</sup> transcripts (I–V) are shown: I + II, tRNA<sup>Gly</sup> (CCC); III + IV + II, tRNA<sup>Gly</sup> (TCC); and III + V + II, tRNA<sup>Gly</sup> (GCC). Anticodons are boxed. The locations of bulge–helix–bulge (BHB) splicing motifs represent each leader–exon junction (arrowhead). (B) Nucleotide sequence alignment of the 3 pretRNA<sup>Gly</sup> candidates with the inclusion of gaps (–). Identical nucleotide sequences are indicated by asterisks. Leader sequences are boxed. Exon sequences are colored by transcript (I, red; II, blue; III, light blue; IV, orange; and V, pink).