Accurate Identification of Transcription Regulatory Sequences and Genes in Coronaviruses

Chuanyi Zhang,[†,1] Palash Sashittal,[†‡,2] Michael Xiang,[2] Yichi Zhang,[2] Ayesha Kazi,[2] and Mohammed El-Kebir [*,2]

[1]Department of Electrical & Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

[2]Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA

[†]These authors share joint first authorship.

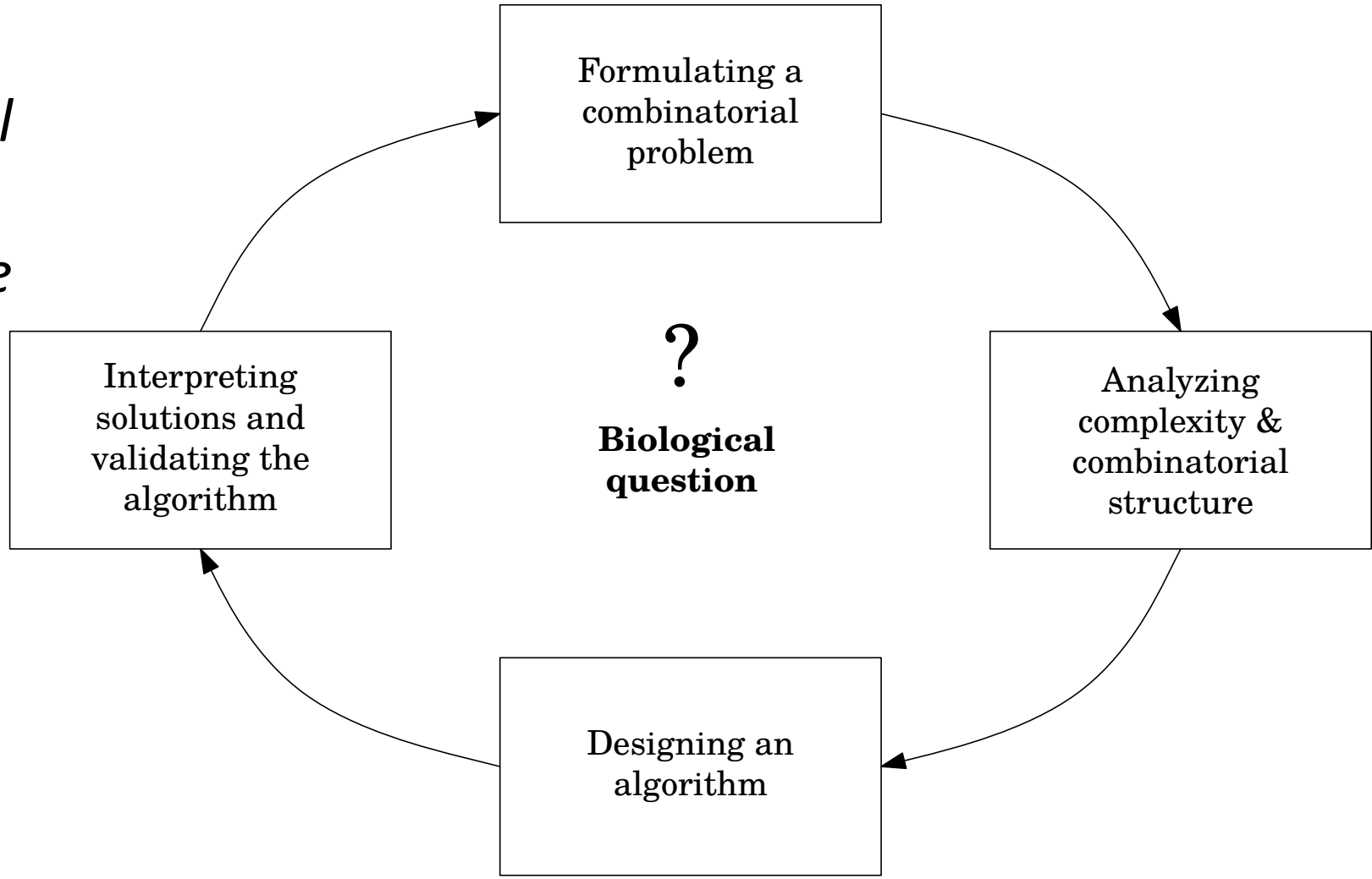
[‡]Present address: Department of Computer Science, Princeton University, Princeton, NJ, USA

RECOMB 2022 / Molecular Biology and Evolution (2022)

Research Statement & Approach

Lab focus:

Application of combinatorial optimization techniques to answers questions and solve problems in biology.



Problem != Algorithm

Accurate Identification of Transcription Regulatory Sequences and Genes in Coronaviruses

Chuanyi Zhang,[†,1] Palash Sashittal,[†‡,2] Michael Xiang,[2] Yichi Zhang,[2] Ayesha Kazi,[2] and Mohammed El-Kebir [*,2]

[1]Department of Electrical & Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

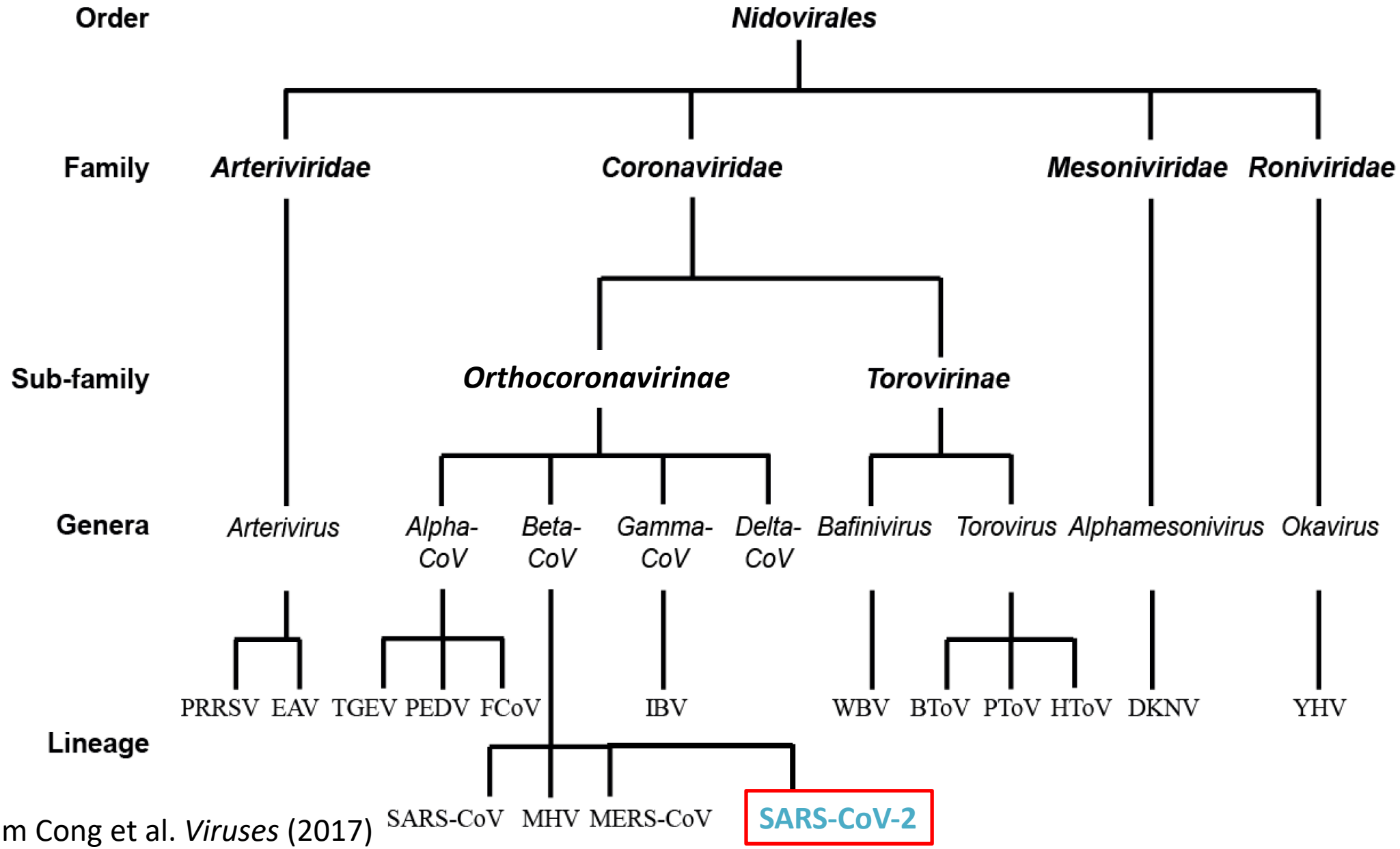
[2]Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA

[†]These authors share joint first authorship.

[‡]Present address: Department of Computer Science, Princeton University, Princeton, NJ, USA

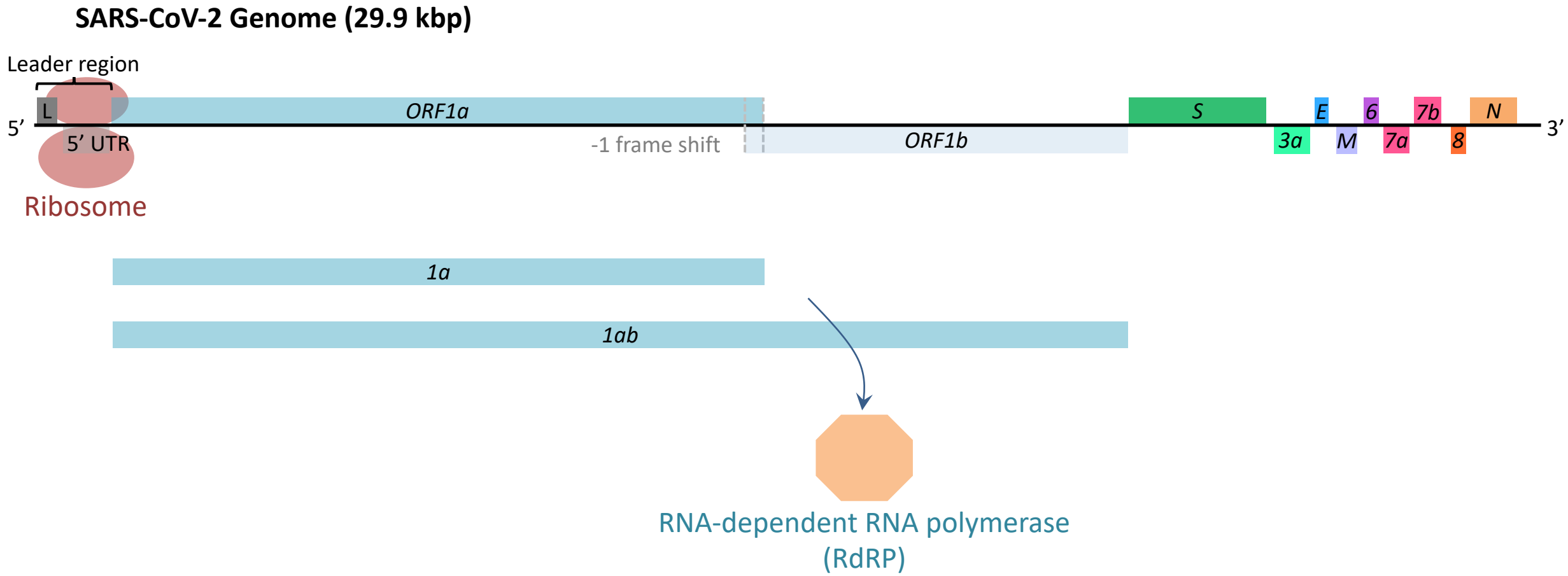
RECOMB 2022 / Molecular Biology and Evolution (2022)

Motivation – Background

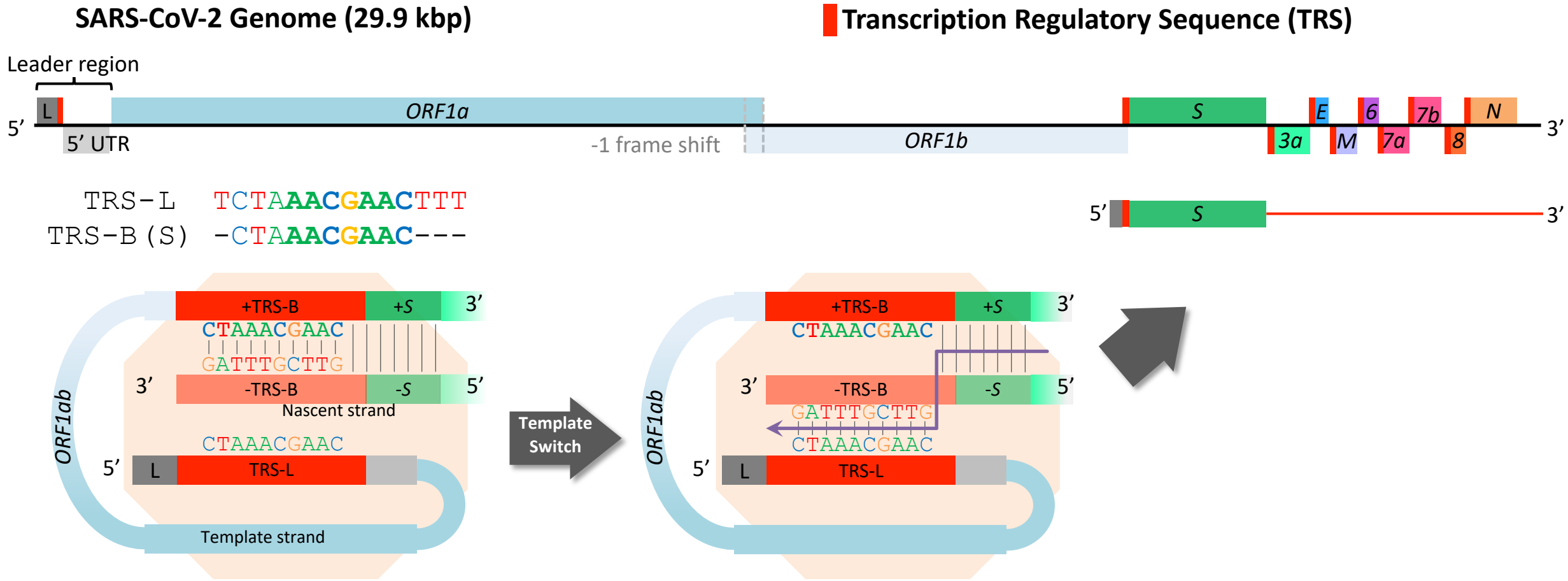


Adapted from Cong et al. *Viruses* (2017)

Motivation – Background



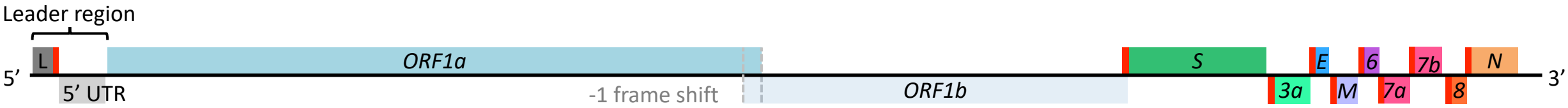
Motivation – Background



Motivation – Background

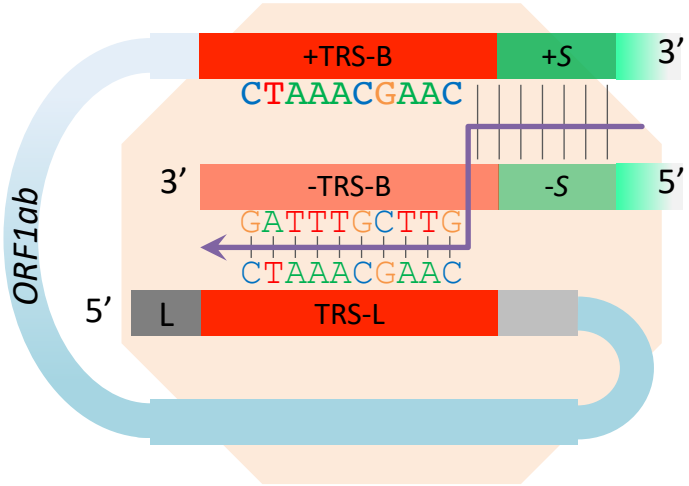
SARS-CoV-2 Genome (29.9 kbp)

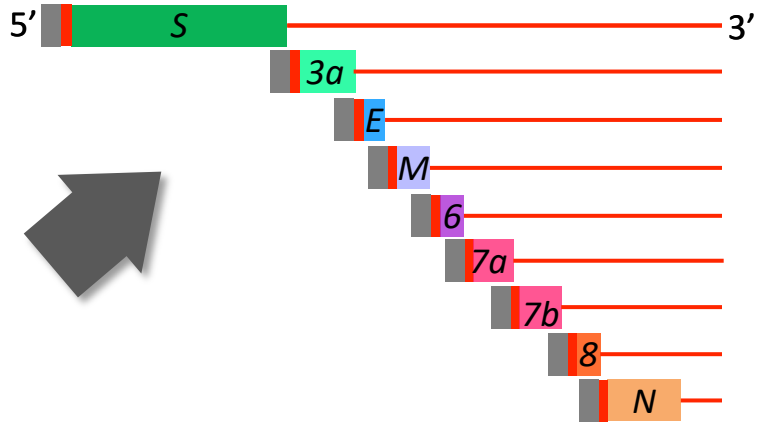
Transcription Regulatory Sequence (TRS)




```
L    TCTAACGAACTTT
S    -CTAACGAAC---
3a   --TAACGAACTT-
E    ----TACGAACTT-
M    TCTAACGAACT--
6    ----CACGAAC---
7a   --TAACGAAC---
7b   --TAAAGAACCTT
8    -CTAACGAAC---
N    TCTAACGAAC---
```

TRS-L and TRS-Bs





Subgenomic messenger RNAs (sgmRNAs)

Discontinuous transcription due to template switching of RdRP at transcription regulatory sequences

Sola et al., *Annual Review of Virology,* (2015)

Motivation – Two Key Questions

Question 1. Can we identify TRS-L and TRS-Bs in **annotated** genomes?

Annotated genome $\mathbf{v}$ with genes $\mathbf{x}_0, \ldots, \mathbf{x}_n$

TRS-L $\mathbf{a}_0$ and TRS-Bs $\mathbf{a}_1, \ldots, \mathbf{a}_n$

Question 2. Can we identify TRS-L, TRS-Bs and their **corresponding genes** in **unannotated** genomes?

Unannotated genome $\mathbf{v}$

TRS-L $\mathbf{a}_0$, TRS-Bs $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and associated genes $\mathbf{x}_0, \ldots, \mathbf{x}_n$

Motivation – Current Methods

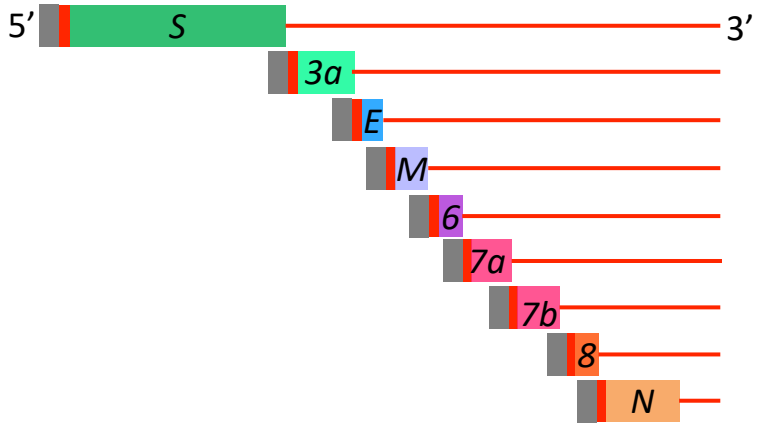
- **Motif finding (with annotation):**
  - MEME [Bailey et al. 2009] – general purpose
  - SuPER [Yang et al. 2021] – needs prior knowledge
- **Gene finding (without annotation):**
  - Prodigal [Hyatt et al. 2010, 2021]
  - Glimmer3 [Salzberg et al. 1998, Delcher et al. 2007]
  - VADR [Schäffer et al. 2020]

Methods	TRS	Gene
CORSID	✓	✓
CORSID-A	✓	✗
SuPER	✓	✗
MEME	✓	✗
Glimmer3	✗	✓
Prodigal	✗	✓
VADR	✗	✓

There exists no method that combines motif + gene finding for coronaviruses

Outline

- TRS Identification
  - Problem Statement
  - Method: CORSID-A
- TRS + Gene Identification
  - Problem Statement
  - Method: CORSID
- Results
- Conclusion




```
L   TCTAAACGAACTTT
S   -CTAAACGAAC---
3a  --TAAACGAACTT-
E   ----TACGAACTT-
M   TCTAAACGAACT--
6   ----CACGAAC---
7a  --TAAACGAAC---
7b  --TAAAAGAACCTT
8   -CTAAACGAAC---
N   TCTAAACGAAC---
```

Problem Statement – Definitions

Annotated genome $\mathbf{v}$ with genes $\mathbf{x}_0, \ldots, \mathbf{x}_n$

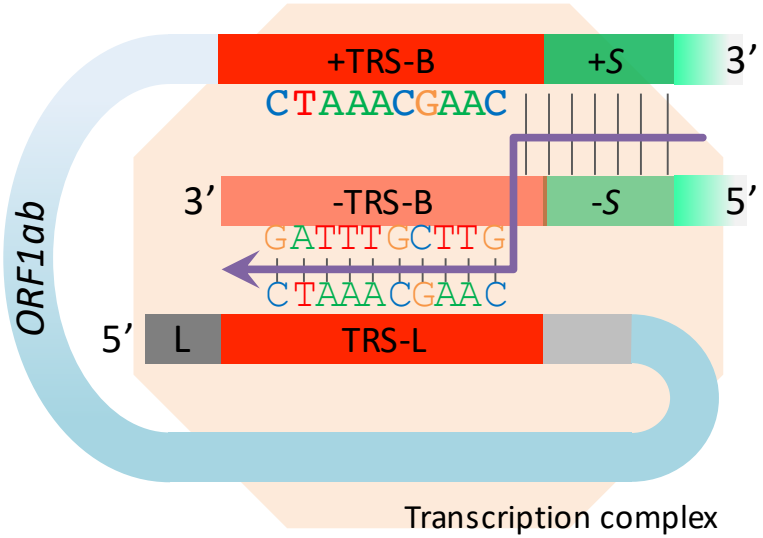
TRS-L $\mathbf{a}_0$ and TRS-Bs $\mathbf{a}_1, \ldots, \mathbf{a}_n$

TRS alignment is a constrained multiple sequence alignment (MSA):
(1) No gaps in $\mathbf{a}_0$; (2) No internal gaps in $\mathbf{a}_i$ $(i > 0)$;

TRS alignment A

$\mathbf{a}_0 =$	L	TCTAACGAACTTT
$\mathbf{a}_1 =$	S	-CTAACGAAC---
$\mathbf{a}_2 =$	3a	--TAACGAACTT-
$\mathbf{a}_3 =$	E	----TACGAACTT-
$\mathbf{a}_4 =$	M	TCTAACGAACT--
$\mathbf{a}_5 =$	6	----CACGAAC---
$\mathbf{a}_6 =$	7a	--TAACGAAC---
$\mathbf{a}_7 =$	7b	--TAAAGAACCTT
$\mathbf{a}_8 =$	8	-CTAACGAAC---
$\mathbf{a}_9 =$	N	TCTAACGAAC---

Core sequence: $|\mathbf{c}(A)| = 7$

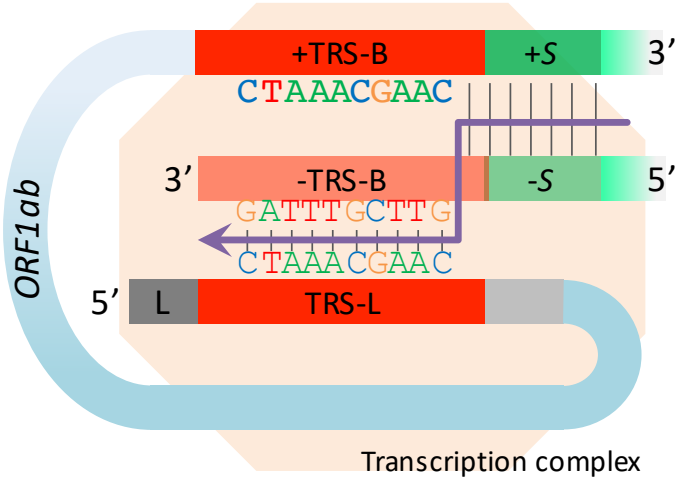


Problem Statement – Definitions

TRS alignment is a constrained multiple sequence alignment (MSA):
(1) No gaps in $\mathbf{a}_0$; (2) No internal gaps in $\mathbf{a}_i$ $(i > 0)$;
(3) While MSA are typically scored via sum-of-pairs, we score only pairs $(\mathbf{a}_0, \mathbf{a}_i)$

$\mathbf{a}_0$	L	TRS-L TCTAAACGAACTTT TRS-B	#matches (+1)	#mismatches (-2)	$s(\mathbf{a}_0, \mathbf{a}_i)$
$\mathbf{a}_1$	S	–CTAAACGAAC–––	10	0	10
$\mathbf{a}_2$	3a	––TAAACGAACTT–	11	0	11
$\mathbf{a}_3$	E	––––TACGAACTT–	8	1	6
$\mathbf{a}_4$	M	TCTAAACGAACT––	12	0	12
$\mathbf{a}_5$	6	––––CACGAAC–––	6	1	4
$\mathbf{a}_6$	7a	––TAAACGAAC–––	9	0	9
$\mathbf{a}_7$	7b	––TAAAAGAACCTT	10	2	6
$\mathbf{a}_8$	8	–CTAAACGAAC–––	10	0	10
$\mathbf{a}_9$	N	TCTAAACGAAC–––	11	0	11

$$\sum_{i=1}^{9} s(\mathbf{a}_0, \mathbf{a}_i) = s(A) = 79$$



Problem Statement – **TRS Identification**

TRS Identification Problem: Given non-overlapping sequences $\mathbf{w}_0, \ldots, \mathbf{w}_n$, core sequence length $\omega > 0$, find a TRS alignment $A = [\mathbf{a}_0, \ldots, \mathbf{a}_n]^\top$ such that (i) $\mathbf{a}_i$ corresponds to a subsequence in $\mathbf{w}_i$ for all $i \in \{0, \ldots, n\}$, (ii) the core sequence $\mathbf{c}(A)$ has length at least ω and (iii) the alignment has maximum score $s(A)$

$s(A) = 79$

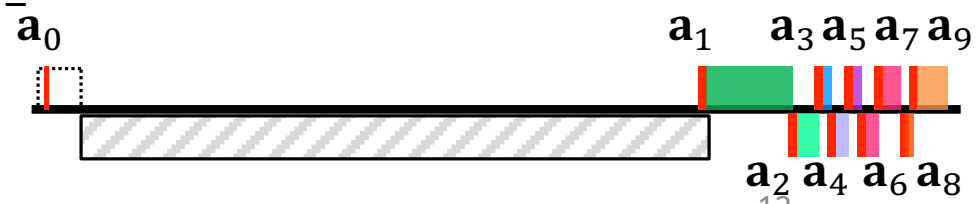
$\mathbf{c}(A) = \text{AACGAAC}$

Candidate regions: $\mathbf{w}_0, \ldots, \mathbf{w}_9$

Core sequence $\omega = 7$

```
a0 =  L TCTAAACGAACTTT
a1 =  S -CTAAACGAAC---
a2 = 3a --TAAACGAACTT-
a3 =  E ----TACGAACTT-
a4 =  M TCTAAACGAACT--
a5 =  6 ----CACGAAC---
a6 = 7a --TAAACGAAC---
a7 = 7b --TAAAGAACCTT
a8 =  8 -CTAAACGAAC---
a9 =  N TCTAAACGAAC---
```

TRS-ID



Problem Statement – **TRS Identification**

TRS Identification Problem: Given non-overlapping sequences $\mathbf{w}_0, \ldots, \mathbf{w}_n$, core sequence length $\omega > 0$, find a TRS alignment $A = [\mathbf{a}_0, \ldots, \mathbf{a}_n]^\top$ such that (i) $\mathbf{a}_i$ corresponds to a subsequence in $\mathbf{w}_i$ for all $i \in \{0, \ldots, n\}$, (ii) the core sequence $\mathbf{c}(A)$ has length at least ω and (iii) the alignment has maximum score $s(A)$

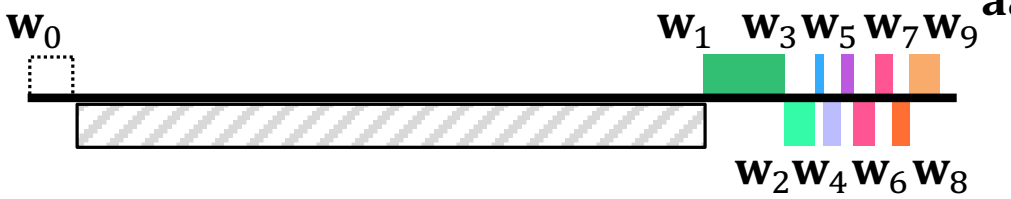
$s(A) = 79$

$\mathbf{c}(A) = \text{AACGAAC}$

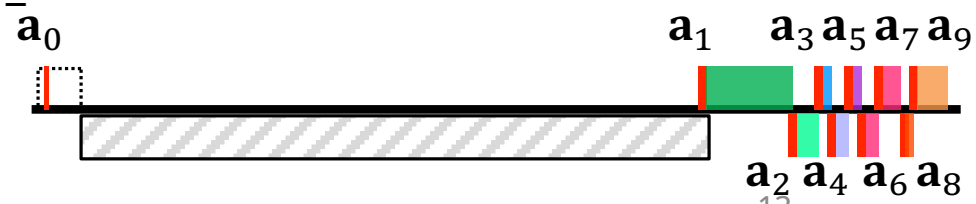
Candidate regions: $\mathbf{w}_0, \ldots, \mathbf{w}_9$

Core sequence $\omega = 7$

```
a0 =  L TCTAAACGAACTTT
a1 =  S -CTAAACGAAC---
a2 = 3a --TAAACGAACTT-
a3 =  E ----TACGAACTT-
a4 =  M TCTAAACGAACT--
a5 =  6 ----CACGAAC---
a6 = 7a --TAAACGAAC---
a7 = 7b --TAAAGAACCTT
a8 =  8 -CTAAACGAAC---
a9 =  N TCTAAACGAAC---
```



TRS-ID



Key observation:
TRS alignment A must have induced core sequence $c(A)$ with length at least ω. Thus, input sequences $\mathbf{w}_1, \ldots, \mathbf{w}_n$ depend on one another and we cannot consider them in isolation.

Method – TRS Identification (I)

Key idea 1:
Break dependency using a subsequence $\mathbf{u}$ of $\mathbf{w}_0$ requiring that $\mathbf{c}(A)$ contains $\mathbf{u}$.

Key idea 2:
Solution to TRS-ID corresponds to window $\mathbf{u}^*$ that induces TRS alignment with maximum score.

Constrained TRS Identification Problem:
Given candidate regions $\mathbf{w}_0, \mathbf{w}_1, \ldots \mathbf{w}_n$ and subsequence $\mathbf{u}$ of $\mathbf{w}_0$, find the optimal TRS alignment such that $\mathbf{u}$ is contained in the core sequence.

Method – TRS Identification (I)

Key idea 1:
Break dependency using a subsequence $\mathbf{u}$ of $\mathbf{w}_0$ requiring that $\mathbf{c}(A)$ contains $\mathbf{u}$.

Key idea 2:
Solution to TRS-ID corresponds to window $\mathbf{u}^*$ that induces TRS alignment with maximum score.

Constrained TRS Identification Problem:
Given candidate regions $\mathbf{w}_0, \mathbf{w}_1, \ldots \mathbf{w}_n$ and subsequence $\mathbf{u}$ of $\mathbf{w}_0$, find the optimal TRS alignment such that $\mathbf{u}$ is contained in the core sequence.

Key idea 3:
Each problem decomposes into n subproblems of aligning $\mathbf{w}_0$ to $\mathbf{w}_i$ spanning $\mathbf{u}$.

Key idea 4:
Pairwise Constrained TRS-ID is a variant of local alignment with three differences, (i) alignment $A = [\mathbf{a}_0, \mathbf{a}_i]^T$ may not contain gaps, (ii) $\mathbf{a}_0$ must contain $\mathbf{u}$.

Intermezzo – Global vs. Local Alignment

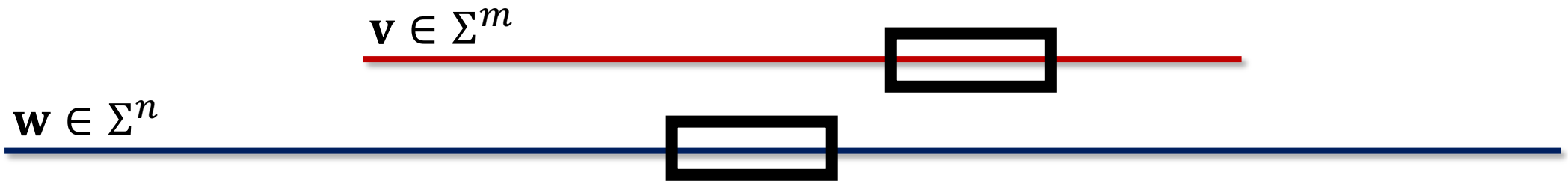
Global Alignment problem: Given strings $\mathbf{v} \in \Sigma^m$ and $\mathbf{w} \in \Sigma^n$ and scoring function δ, find alignment of $\mathbf{v}$ and $\mathbf{w}$ with maximum score.

$\mathbf{v} \in \Sigma^m$

$\mathbf{w} \in \Sigma^n$

Intermezzo – Global vs. Local Alignment

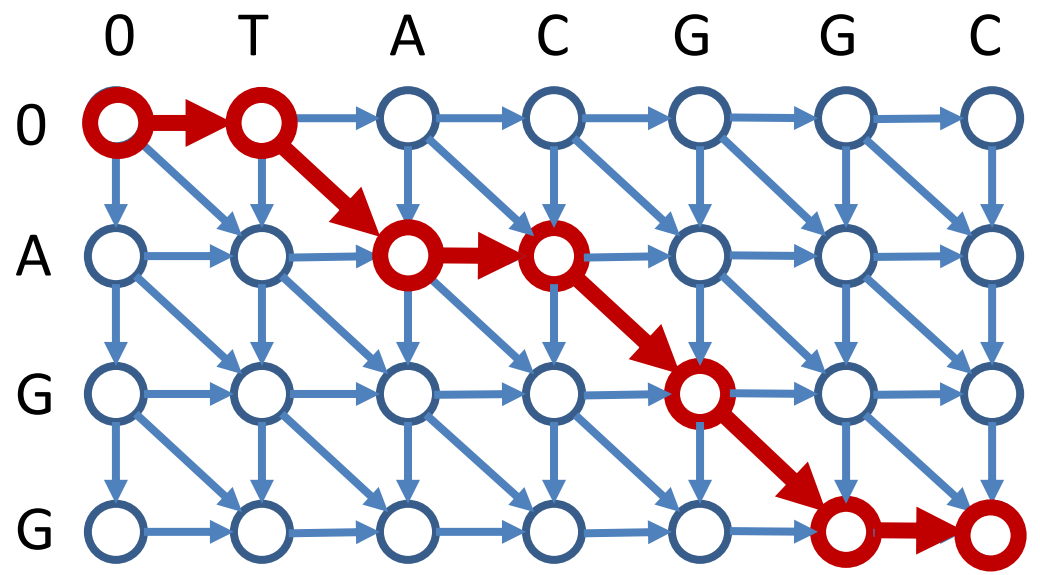
Global Alignment problem: Given strings $\mathbf{v} \in \Sigma^m$ and $\mathbf{w} \in \Sigma^n$ and scoring function δ, find alignment of $\mathbf{v}$ and $\mathbf{w}$ with maximum score.



Local Alignment problem: Given strings $\mathbf{v} \in \Sigma^m$ and $\mathbf{w} \in \Sigma^n$ and scoring function δ, find a substring of $\mathbf{v}$ and a substring of $\mathbf{w}$ whose alignment has maximum global alignment score s^* among *all* global alignments of *all* substrings of $\mathbf{v}$ and $\mathbf{w}$.

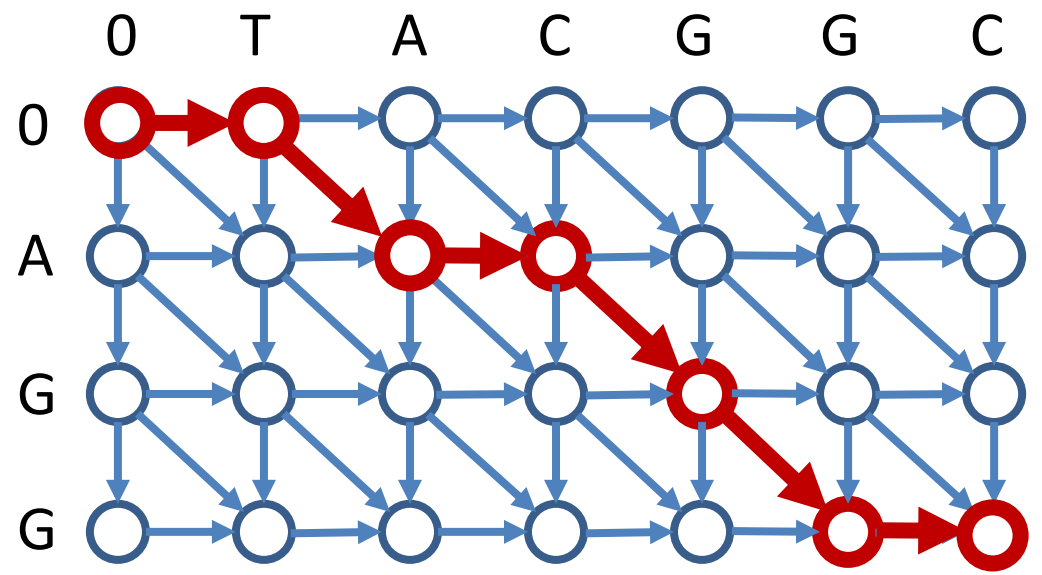
Intermezzo – Global vs. Local Alignment

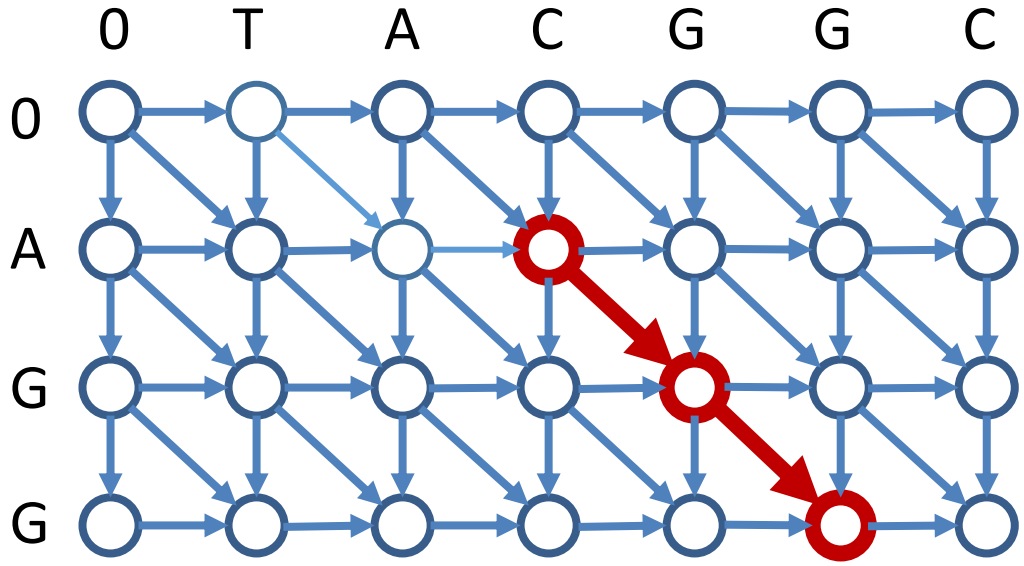
$$s[i,j] = \max \begin{cases} 0, & \text{if } i = 0 \text{ and } j = 0, \\ s[i-1,j] + \delta(v_i, -), & \text{if } i > 0, \\ s[i,j-1] + \delta(-, w_j), & \text{if } j > 0, \\ s[i-1,j-1] + \delta(v_i, w_j), & \text{if } i > 0 \text{ and } j > 0. \end{cases}$$



Intermezzo – Global vs. Local Alignment

$$s[i,j] = \max \begin{cases} 0, & \text{if } i = 0 \text{ and } j = 0, \\ s[i-1,j] + \delta(v_i, -), & \text{if } i > 0, \\ s[i,j-1] + \delta(-, w_j), & \text{if } j > 0, \\ s[i-1,j-1] + \delta(v_i, w_j), & \text{if } i > 0 \text{ and } j > 0. \end{cases}$$





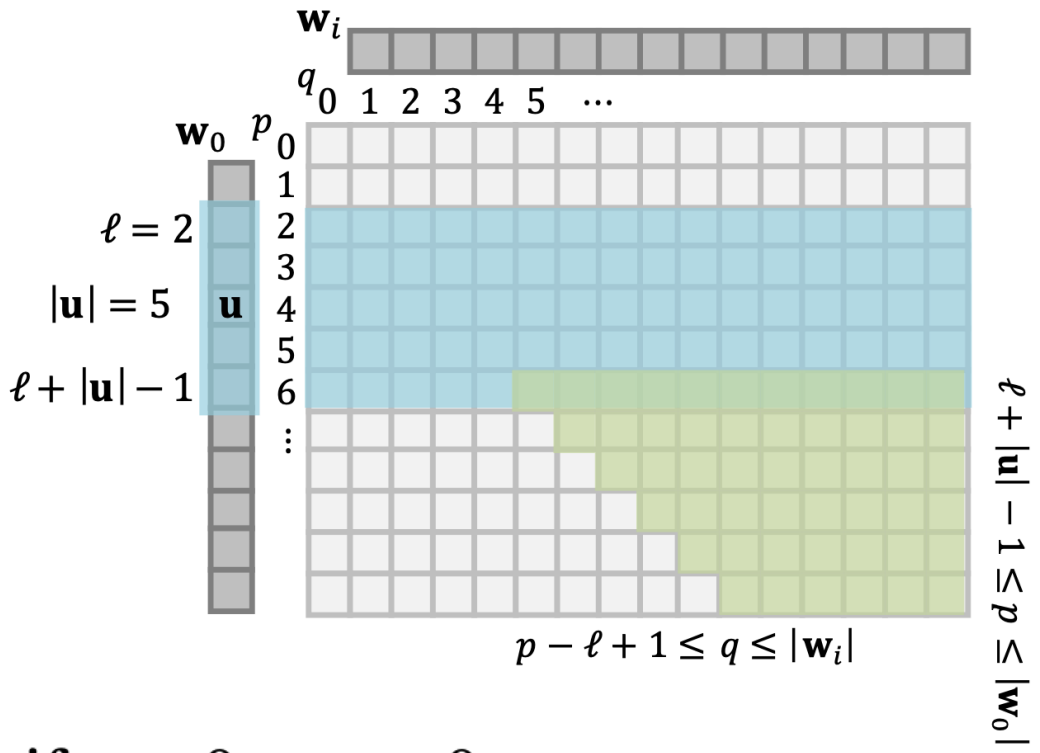
$$s[i,j] = \max \begin{cases} 0, & \text{~~if } i = 0 \text{ and } j = 0,~~} \\ s[i-1,j] + \delta(v_i, -), & \text{if } i > 0, \\ s[i,j-1] + \delta(-, w_j), & \text{if } j > 0, \\ s[i-1,j-1] + \delta(v_i, w_j), & \text{if } i > 0 \text{ and } j > 0. \end{cases}$$

$$s^* = \max_{i,j} s[i,j]$$

Method – TRS Identification (II)

Key idea 4:
Pairwise Constrained TRS-ID is a variant of local alignment with three differences, (i) alignment $A = [\mathbf{a}_0, \mathbf{a}_i]^T$ may not contain gaps, (ii) $\mathbf{a}_0$ must contain $\mathbf{u} = w_{0,\ell}, \ldots, w_{0,\ell+\omega-1}$ at some position ℓ.

$s[p, q]$ is the optimal score of aligning $w_{0,1}, \ldots, w_{0,p}$ to $w_{i,1}, \ldots, w_{i,q}$

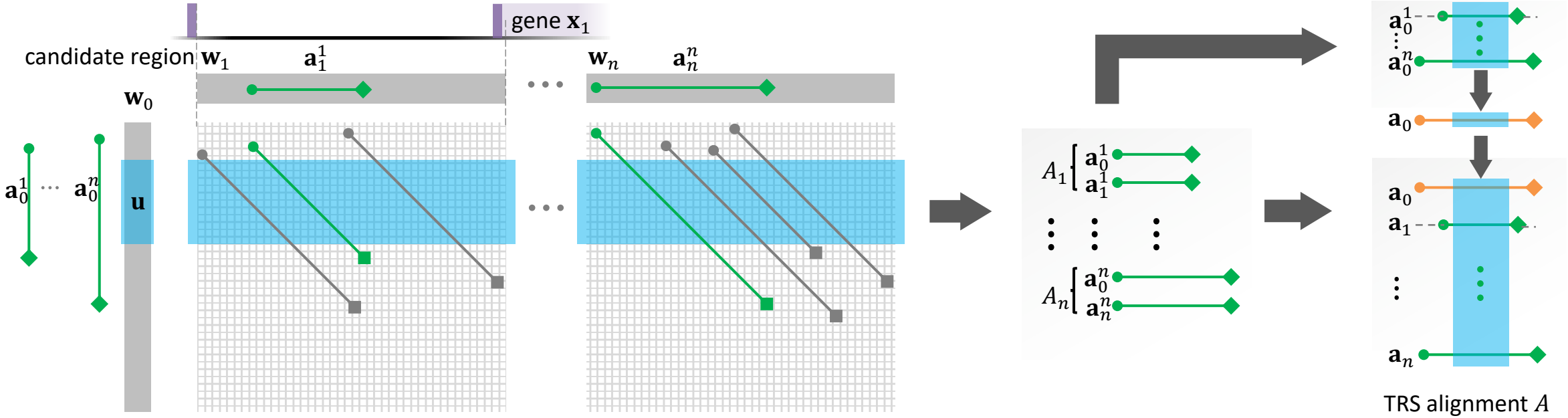


$$s[p,q] = \begin{cases} 0, & \text{if } p = 0 \text{ or } q = 0, \\ \max\{0, s[p-1, q-1] + \delta(w_{0,p}, w_{i,q})\}, & \text{if } 1 \leq p < \ell \text{ and } q \geq 1, \\ s[p-1, q-1] + \delta(w_{0,p}, w_{i,q}), & \text{if } p \geq \ell \text{ and } q \geq 1, \end{cases} \quad (2)$$

$$(p^*, q^*) = \underset{\ell+|\mathbf{u}|-1 \leq p \leq |\mathbf{w}_0|,\ p-\ell+1 \leq q \leq |\mathbf{w}_i|}{\arg\max} s[p,q]. \quad (3)$$

Methods – **CORSID-A**

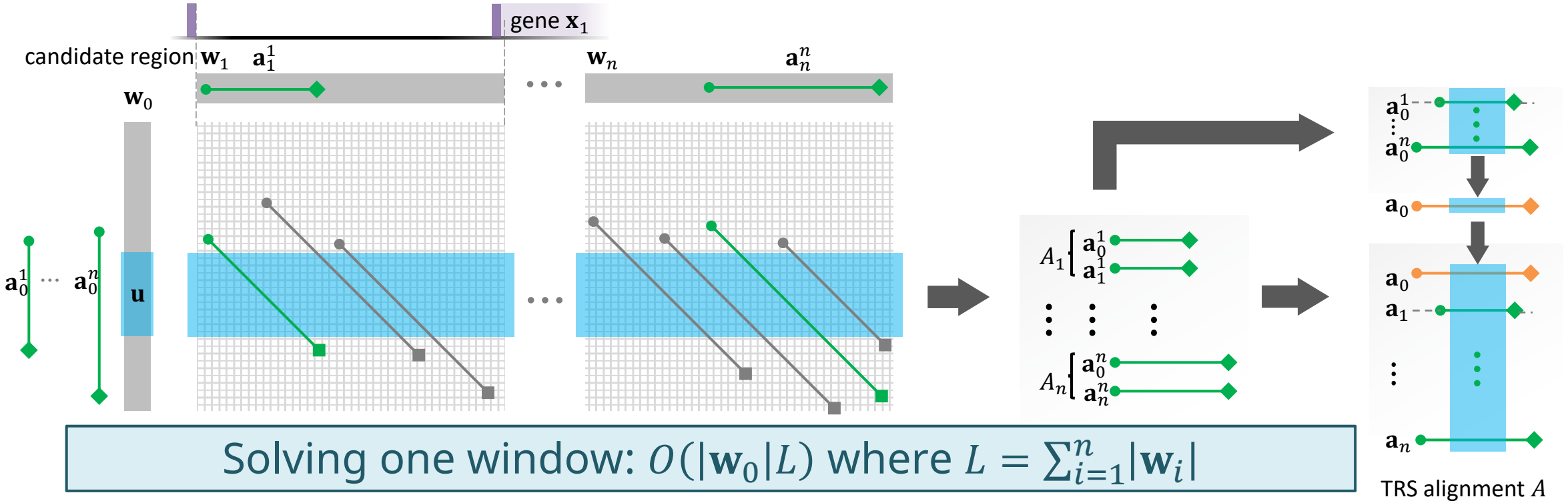
Constrained TRS Identification Problem:
Given candidate regions $\mathbf{w}_0, \mathbf{w}_1, \ldots \mathbf{w}_n$ and subsequence $\mathbf{u}$ of $\mathbf{w}_0$, find the optimal TRS alignment such that $\mathbf{u}$ is contained in the core sequence.



Solving one window: $O(|\mathbf{w}_0|L)$ where $L = \sum_{i=1}^{n} |\mathbf{w}_i|$

Methods – **CORSID-A**

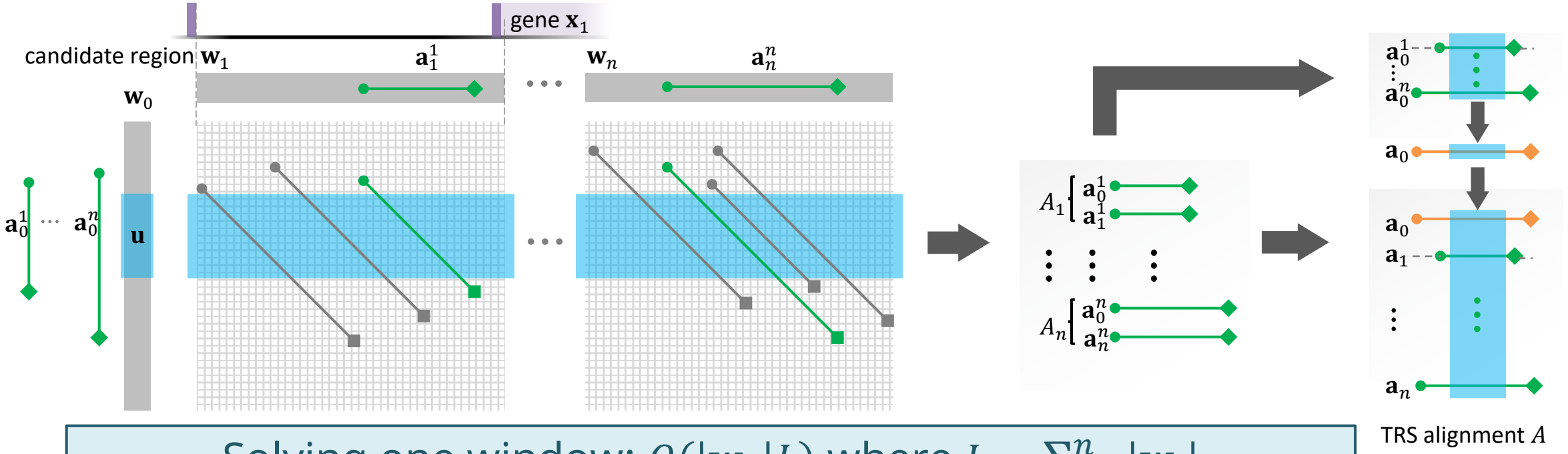
Constrained TRS Identification Problem:
Given candidate regions $\mathbf{w}_0, \mathbf{w}_1, \ldots \mathbf{w}_n$ and subsequence $\mathbf{u}$ of $\mathbf{w}_0$, find the optimal TRS alignment such that $\mathbf{u}$ is contained in the core sequence.



Solving one window: $O(|\mathbf{w}_0|L)$ where $L = \sum_{i=1}^{n}|\mathbf{w}_i|$

Methods – **CORSID-A**

Constrained TRS Identification Problem:
Given candidate regions $\mathbf{w}_0, \mathbf{w}_1, \ldots \mathbf{w}_n$ and subsequence $\mathbf{u}$ of $\mathbf{w}_0$, find the optimal TRS alignment such that $\mathbf{u}$ is contained in the core sequence.

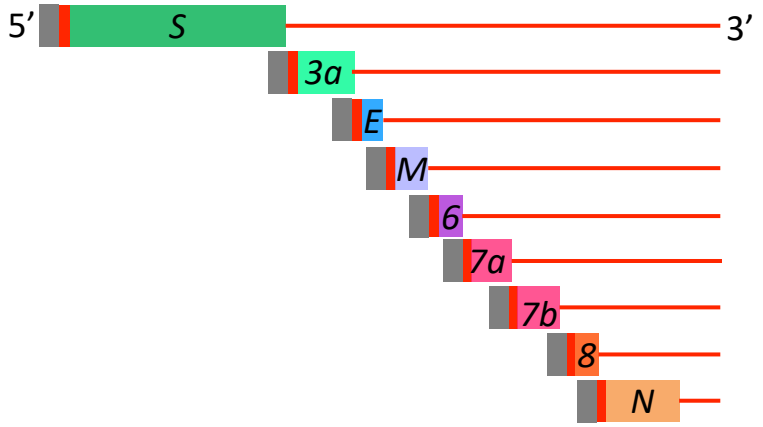


Solving one window: $O(|\mathbf{w}_0|L)$ where $L = \sum_{i=1}^{n}|\mathbf{w}_i|$

Solving all $O(|\mathbf{w}_0|)$ sliding windows: $O(|\mathbf{w}_0|^2 L)$ where $L = \sum_{i=1}^{n}|\mathbf{w}_i|$

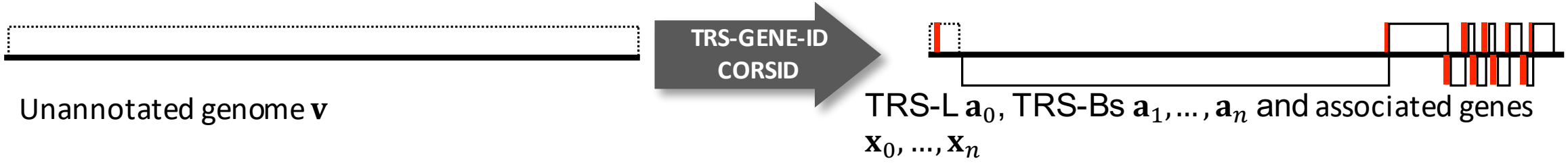
Outline

- TRS Identification
  - Problem Statement
  - Method: CORSID-A
- TRS + Gene Identification
  - Problem Statement
  - Method: CORSID
- Results
- Conclusion

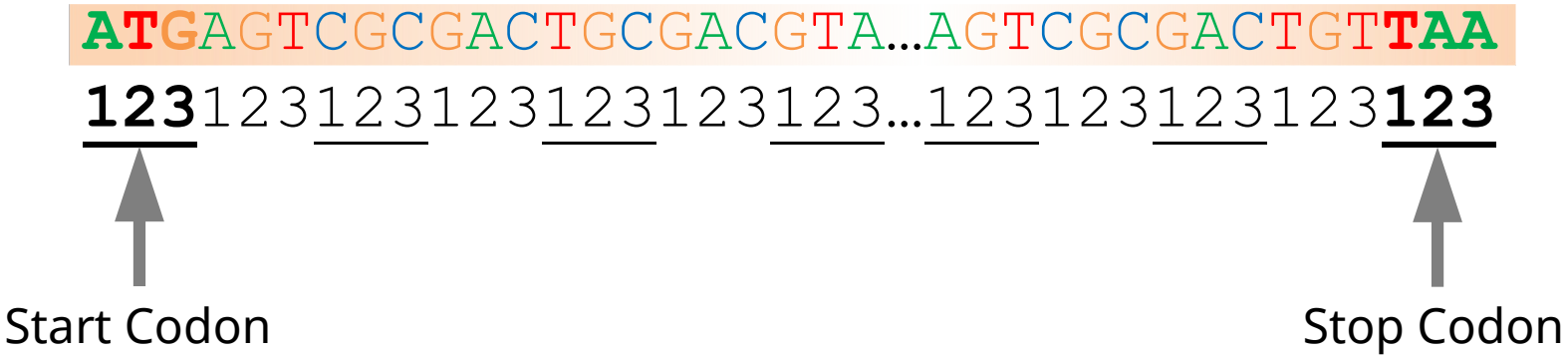



```
L    TCTAAACGAACTTT
S    -CTAAACGAAC---
3a   --TAAACGAACTT-
E    -----TACGAACTT-
M    TCTAAACGAACT--
6    -----CACGAAC---
7a   --TAAACGAAC---
7b   --TAAAAGAACCTT
8    -CTAAACGAAC---
N    TCTAAACGAAC---
```

Problem Statement – **TRS-Gene Identification**



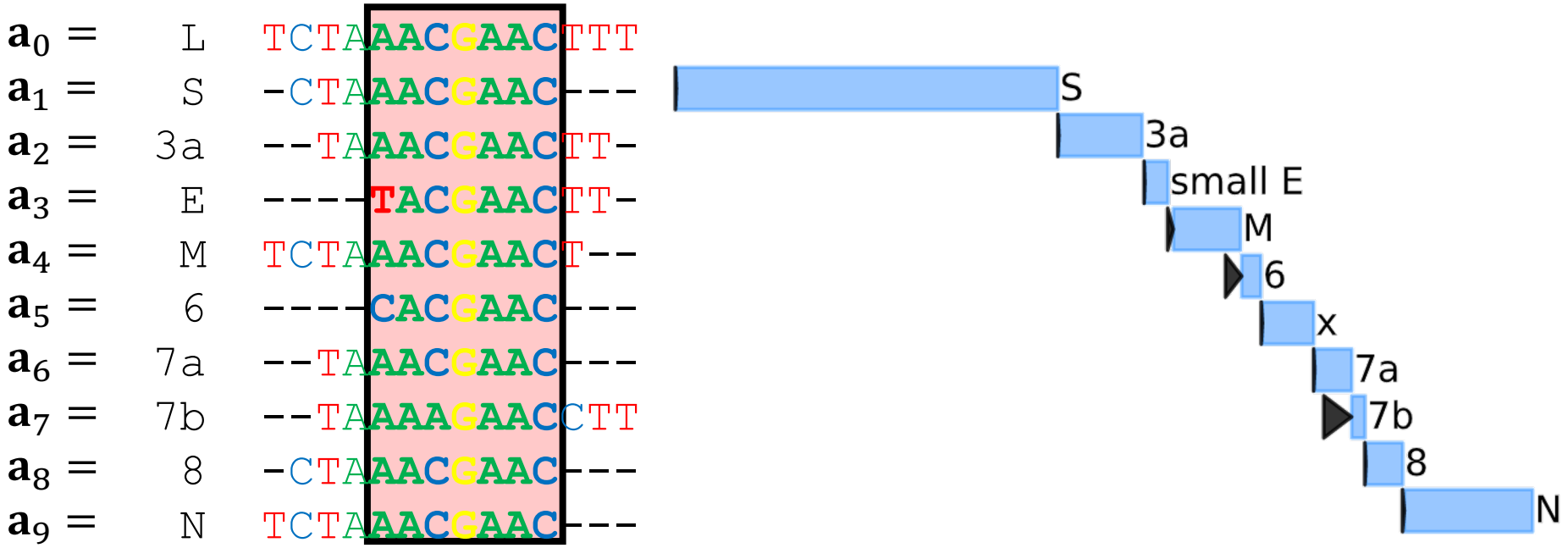
- Enumerate ORFs for genes



- Genes should not overlap – Independent set problem
- Genome coverage $g(A)$ should be large
- Genes should have similar TRS sequences

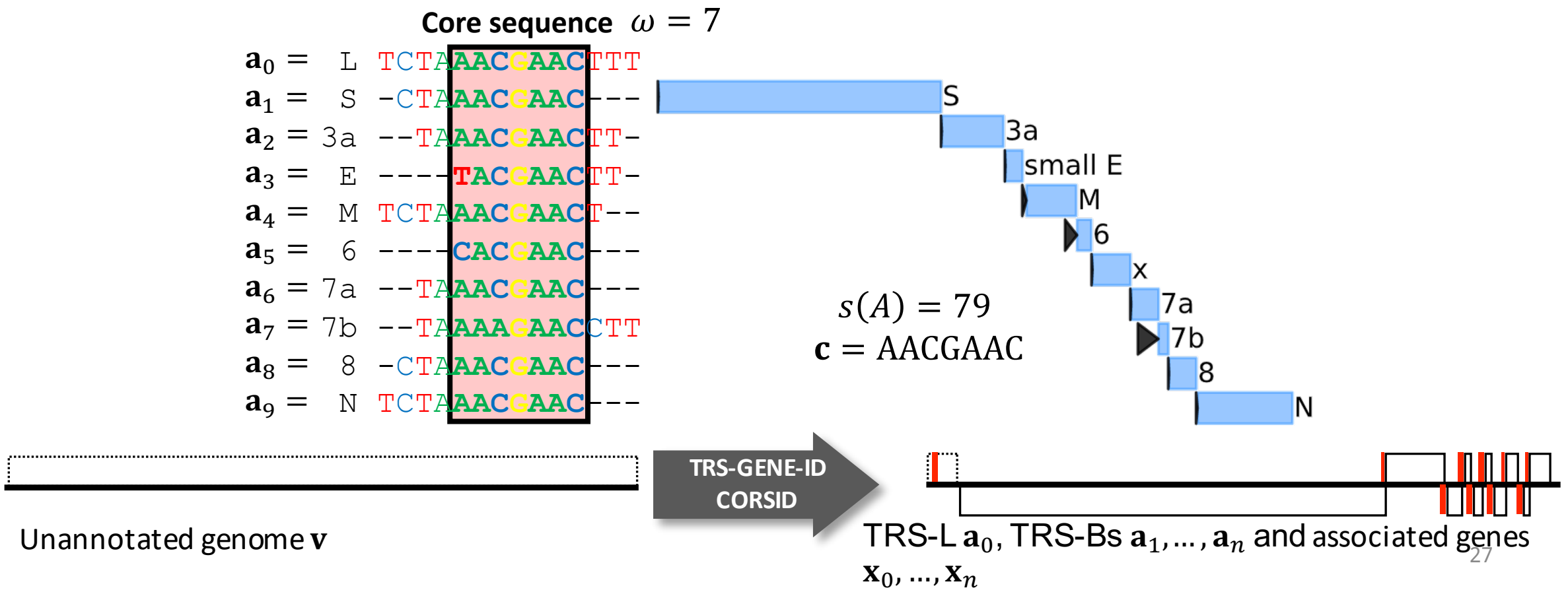
Problem Statement – **TRS-Gene Identification**

- TRS alignment A induces a gene set $\Gamma(A)$



Problem Statement – **TRS-Gene Identification**

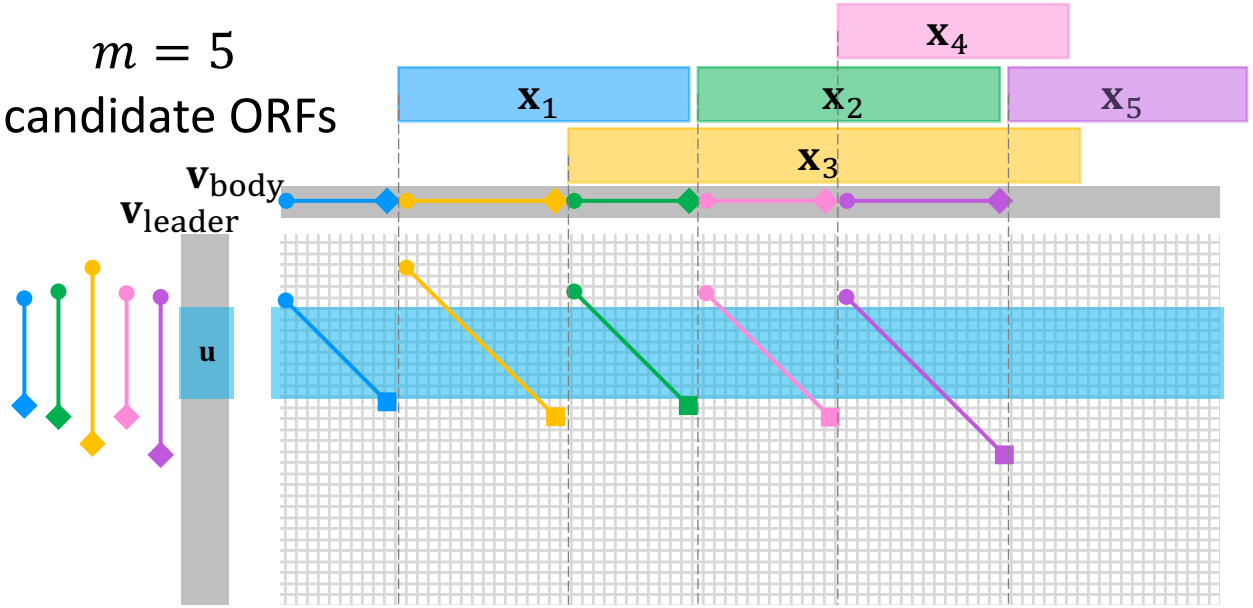
TRS-Gene Identification Problem: Given leader region $\mathbf{v}_{\text{leader}}$, body region $\mathbf{v}_{\text{body}}$, core-sequence length $\omega > 0$, find a TRS alignment $A = [\mathbf{a}_i]$ such that (i) $\mathbf{a}_0$ corresponds to a subsequence in $\mathbf{v}_{\text{leader}}$, (ii) $\mathbf{a}_i$ corresponds to a subsequence in $\mathbf{v}_{\text{body}}$ for all $i \geq 1$, (iii) the core sequence $\mathbf{c}(A)$ has length at least ω, (iv) A induces a set $\Gamma(A)$ of genes with maximum genome coverage $g(A)$ and subsequently maximum score $s(A)$.

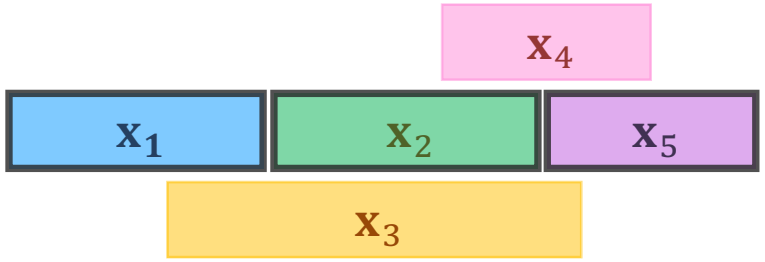


Methods – **CORSID**

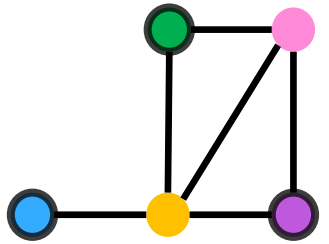
$$O\left(|\mathbf{v}_{\text{leader}}|^2|\mathbf{v}_{\text{body}}| + |\mathbf{v}_{\text{leader}}| \cdot m\right)$$

$m = 5$
candidate ORFs





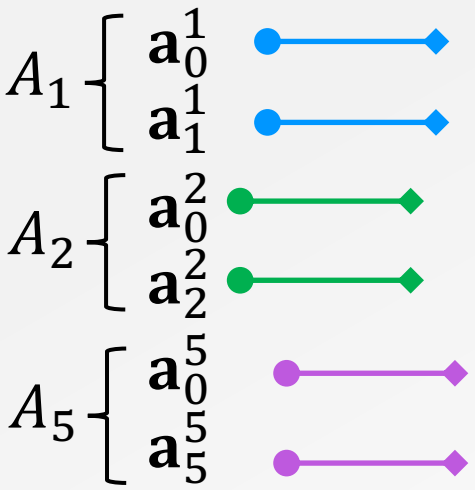
$\Leftrightarrow$

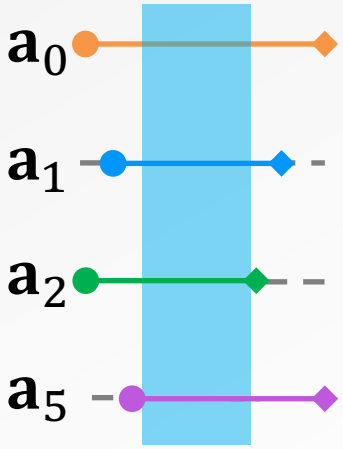


$\text{weight}_i = |\mathbf{x}_i|$

Maximum Weight Independent Set (MWIS)

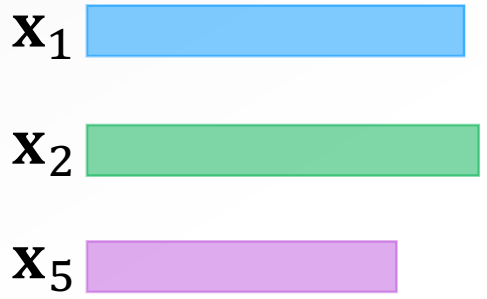
$X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5\}$





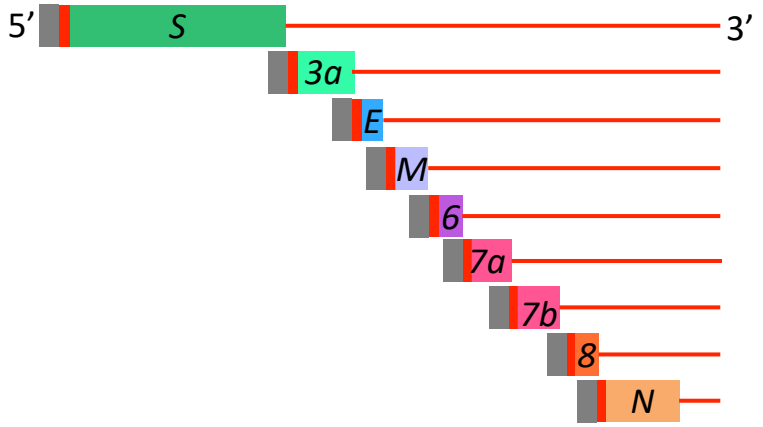
TRS alignment A

Induced genes $\Gamma(A)$



Outline

- TRS Identification
  - Problem Statement
  - Method: CORSID-A
- TRS + Gene Identification
  - Problem Statement
  - Method: CORSID
- Results
- Conclusion

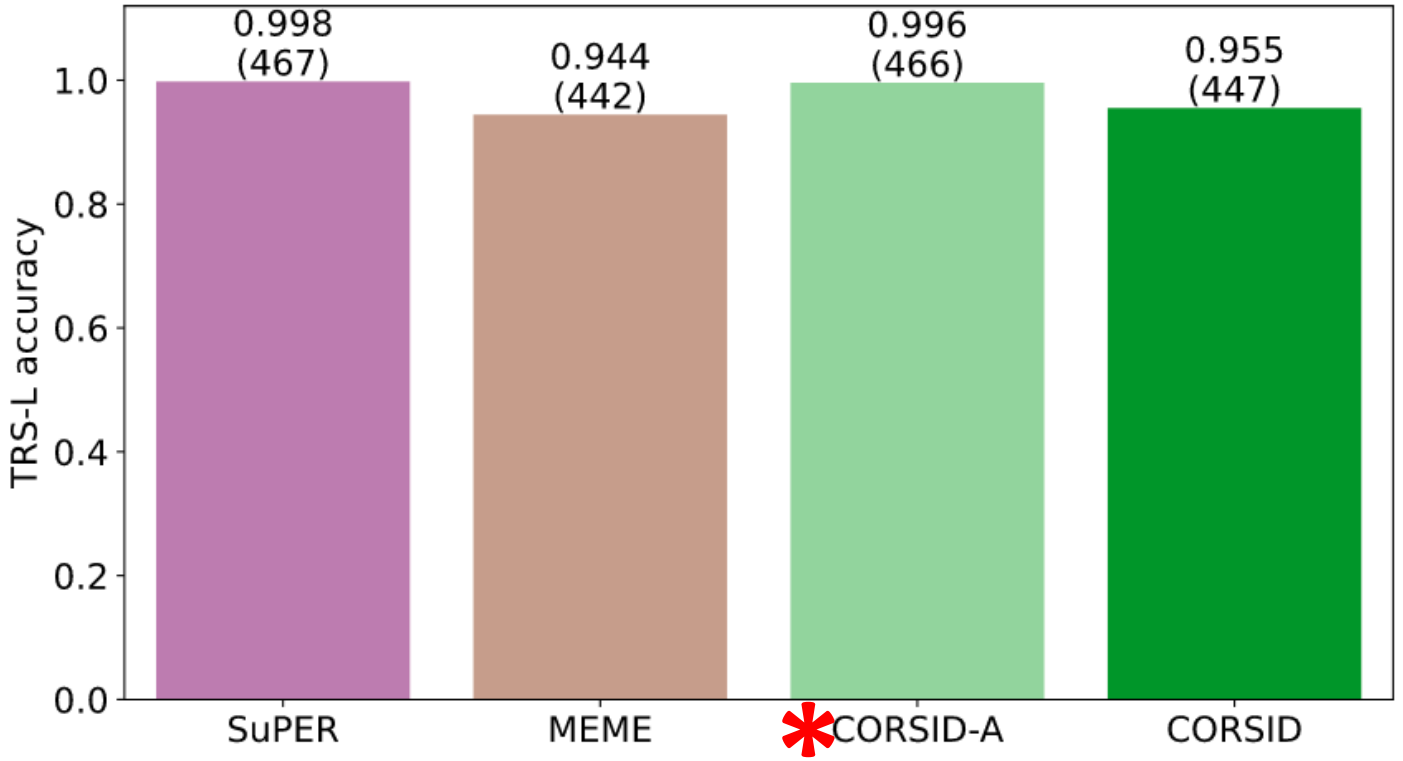



```
L   TCTAAACGAACTTT
S   -CTAAACGAAC---
3a  --TAAACGAACTT-
E   ----TACGAACTT-
M   TCTAAACGAACT--
6   ----CACGAAC---
7a  --TAAACGAAC---
7b  --TAAAAGAACCTT
8   -CTAAACGAAC---
N   TCTAAACGAAC---
```

Results – TRS-L Accuracy

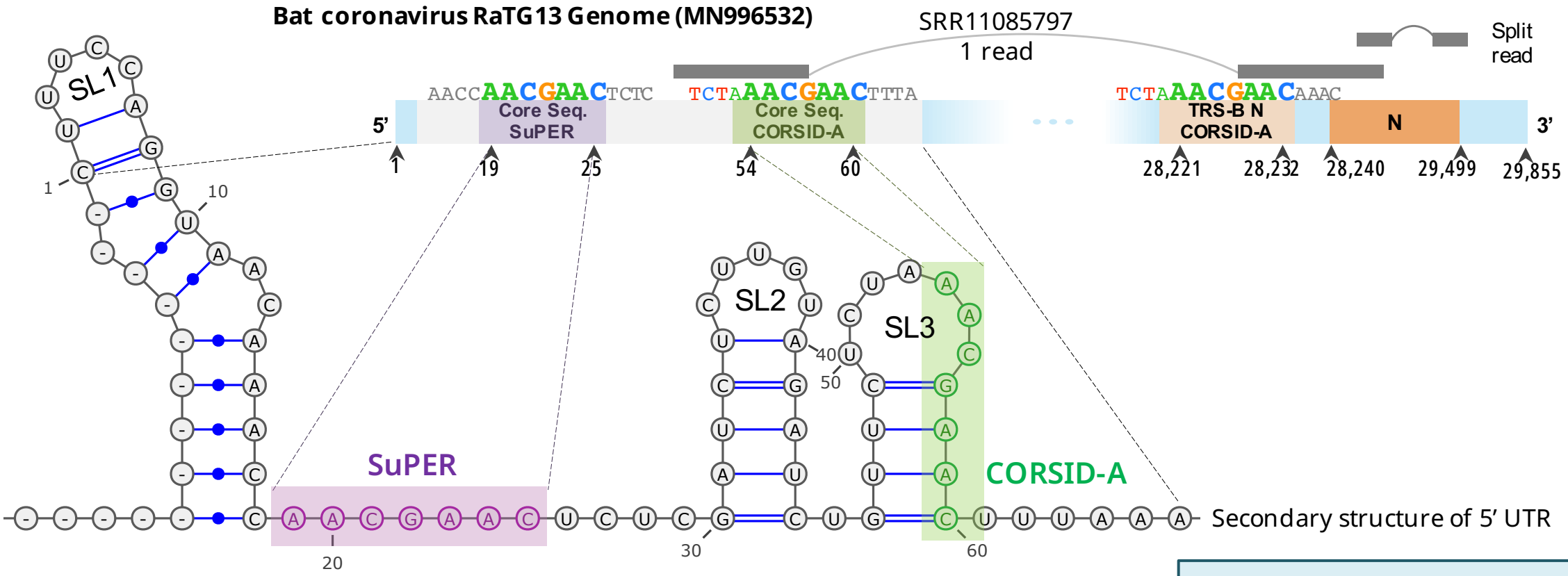
- GenBank: 468 genomes of coronaviruses
- Ground truth based on secondary structure + MSA of leader region

Methods	TRS	Gene	Additional Signal
CORSID	✓	✓	-
CORSID-A	✓	✗	-
SuPER	✓	✗	Secondary structure Predefined motifs
MEME	✓	✗	-
Glimmer3	✗	✓	
Prodigal	✗	✓	
VADR	✗	✓	



CORSID-A correctly identifies TRS-L

Results – CORSID-A finds the correct TRS-L site

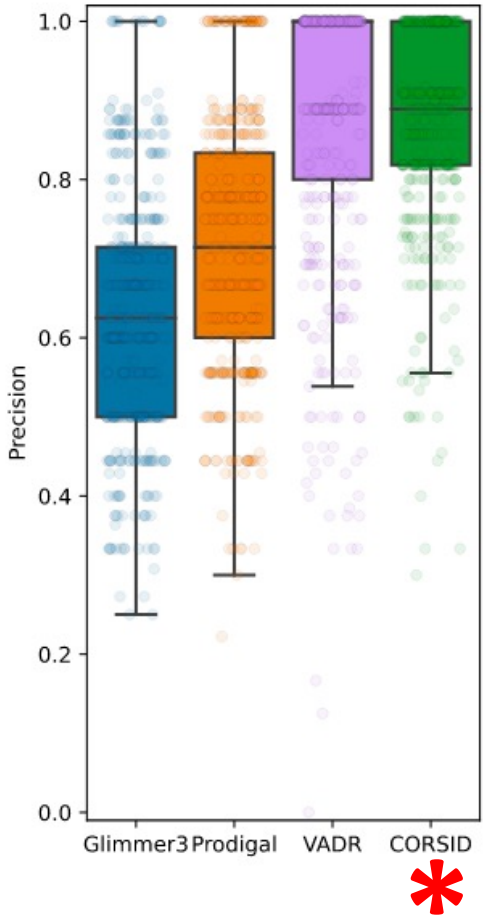


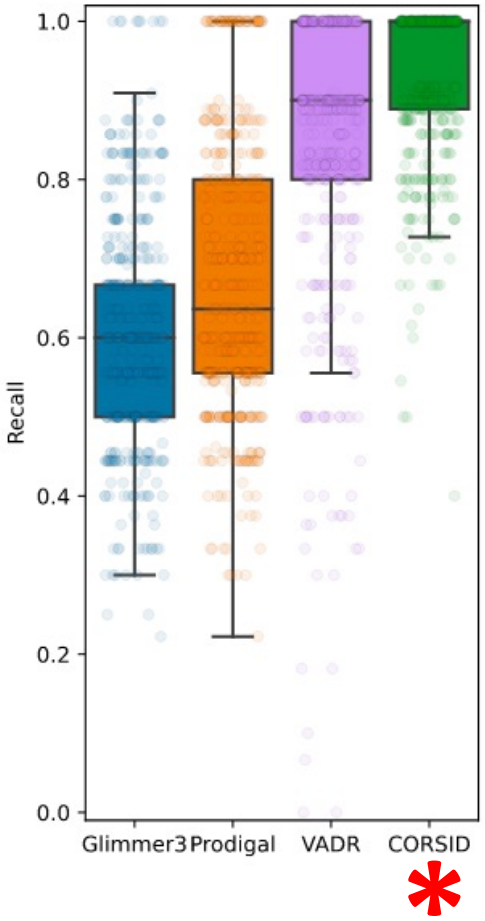
L	AACCAACGAACTCTC	TCTAAACGAACTTTAAAA
S	ACTAAACGAACCATG	-CTAAACGAAC------
3	CATAAACGAACTTAT	--TAAACGAACTT-----
E	TGAGTACGAACTTAT	----TACGAACTT-----
M	TCTAAACGAACTAAA	TCTAAACGAACT------
6	ACATCACGAACGCTT	---AAACTAACATT----
7a	GATTAACGAACATGA	----AACGAACATGAAAA
7b	CTCAAAAGAAAAACA	TCTAAAAGAACCTT----
8	CCTAAACGAACATGA	-CTAAACGAACATGAAA-
N	TCTAAACGAACAAAC	TCTAAACGAAC------

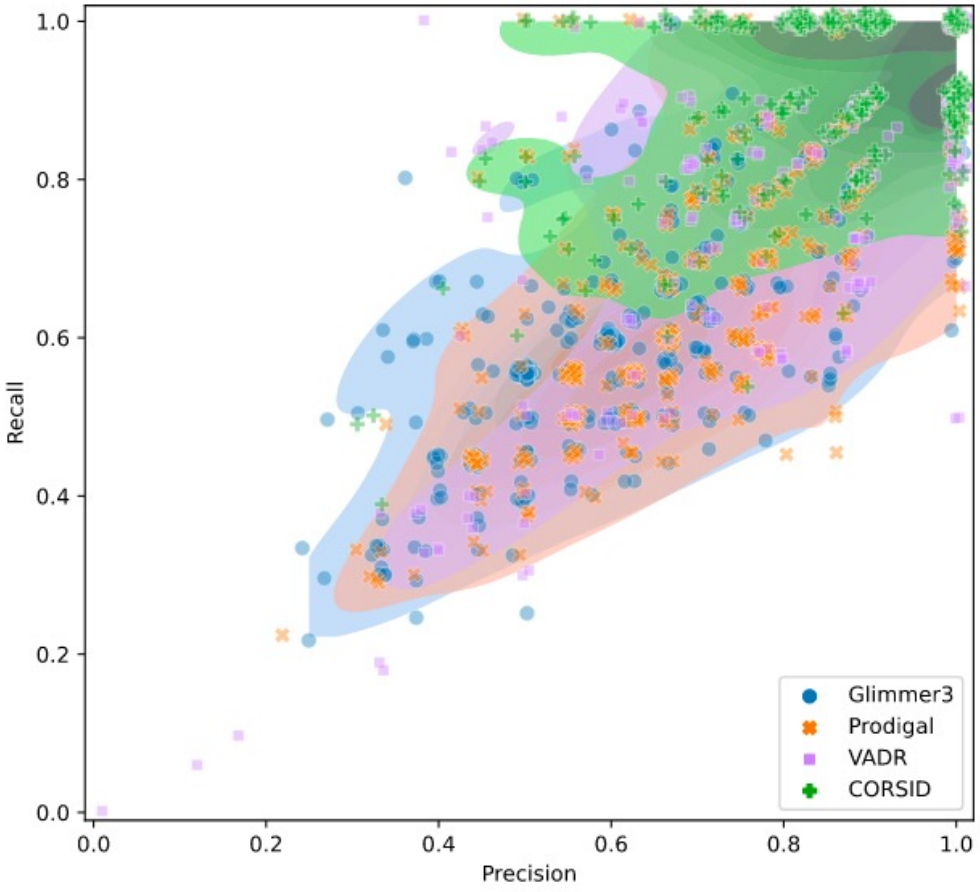
Flanking nucleotides around core sequence help identify TRS-L

Results – CORSID – Gene Identification

Methods	TRS	Gene
CORSID	✓	✓
CORSID-A	✓	✗
SuPER	✓	✗
MEME	✓	✗
Glimmer3	✗	✓
Prodigal	✗	✓
VADR	✗	✓







WebApp

Michael Xiang

Yichi Zhang

Ayesha Kazi

Home Help

Welcome to the CORSID Visualization Tool

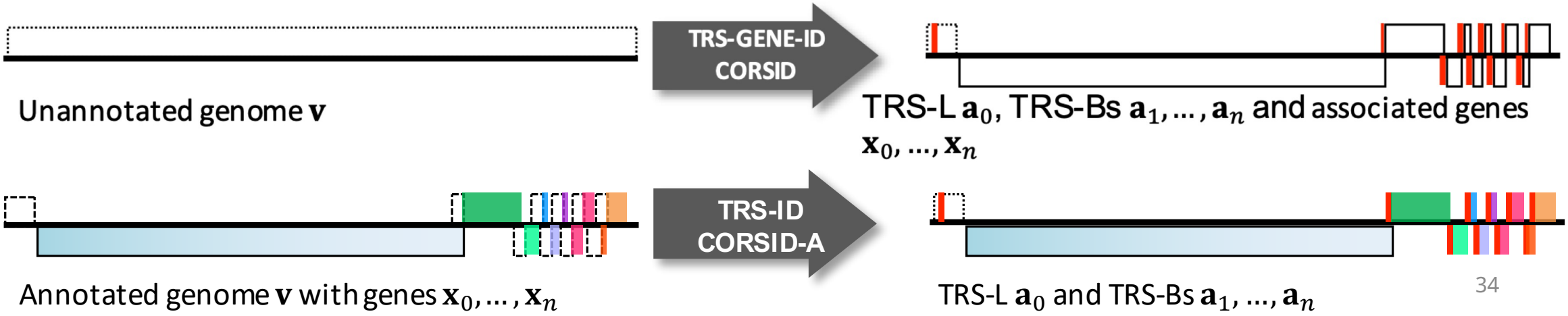
In this demo we show 468 samples used in the analysis. You can start exploring results of CORSID/CORSID-A by clicking any link in the table.

Search: Genus: Subgenus:

Sample	Genus	Subgenus	CORSID viz	CORSID-A viz
AC_000192	Betacoronavirus	Embecovirus		
AF220295	Betacoronavirus	Embecovirus		
AY319651	Gammacoronavirus	Igacovirus		
AY514485	Gammacoronavirus	Igacovirus		
DQ011855	Betacoronavirus	Embecovirus		
DQ022305	Betacoronavirus	Sarbecovirus		
DQ071615	Betacoronavirus	Sarbecovirus		
DQ415899	Betacoronavirus	Embecovirus		
DQ648794	Betacoronavirus	Merbecovirus		
DQ648856	Betacoronavirus	Sarbecovirus		
DQ648857	Betacoronavirus	Sarbecovirus		
DQ811787	Alphacoronavirus	Tegacovirus		
DQ848678	Alphacoronavirus	Tegacovirus		
EF065510	Betacoronavirus	Merbecovirus		
EF065514	Betacoronavirus	Nobecovirus		
EF065515	Betacoronavirus	Nobecovirus		
EF065516	Betacoronavirus	Nobecovirus		
EF446615	Betacoronavirus	Embecovirus		
EU022525	Gammacoronavirus	Igacovirus		
EU022526	Gammacoronavirus	Igacovirus		
EU074218	Alphacoronavirus	Tegacovirus		
EU186072	Alphacoronavirus	Tegacovirus		
EU420137	Alphacoronavirus	Minunacovirus		
EU637854	Gammacoronavirus	Igacovirus		
EU714029	Gammacoronavirus	Igacovirus		

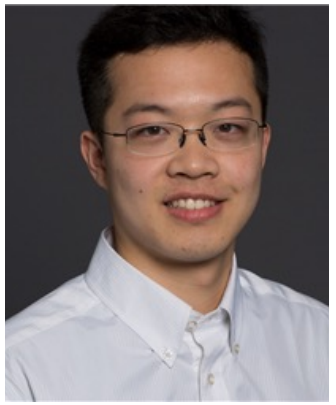
Conclusion

- CORSID:
  - The first method to simultaneously identify TRS sites and genes
  - *De novo* identification of TRS sites and genes
  - Outperforms state-of-the-art methods
- Future direction:
  - Alternative start codon, Kozak sequence
  - RNA-seq, split reads
  - Incomplete genome without TRS-L



Acknowledgements

- Joint first authors:
  Chuanyi Zhang, Palash Sashittal
- Michael Xiang, Yichi Zhang, and Ayesha Kazi
- Grants
  - NSF
    - CRII: CCF-1850502
    - RAPID: CCF-2027669
    - CAREER: CCF-2046488
  - AWS
    Greg Gulick Honorary Research Award



Chuanyi Zhang

Palash Sashittal

- Availability
  - Code

    install with bioconda | pypi package 0.1.3

    https://github.com/elkebir-group/CORSID
  - WebApp

    https://elkebir-group.github.io/CORSID-viz