

To introduce SeqSaw, a mini example is illustrated. Let us take three nine nt reads and a 37-nt reference as the example. We try to align the reads to the reference allowing a gap. In this mini example, the intron is limited to 20 bp, so possible hits with a gap longer than 20 bp will be discarded. In this example, the reads are also required to cover at least two bp on each side of the gap.

See supplementary section 1.6 for allowing mismatches and processing long reads spanning more than one intron.

0 : AGCTATGGA

1 : AGCTAGCCT

2 : GCATGCGGA

a

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

0: AGCTATGGA

1: AGCTAGCCT

2: GCATGCGGA

b

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

b

0: AGCTA TGGA

1: AGCTA GCCT

2: GCATG CGGA

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

add in

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

C

0: AGCTATGGA

1: AGCTAGCCT

2: GCATGCGGA

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

0: AGCTATGGA

1: AGCTAGCCT

2: GCATGCGGA

①slide

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

C

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

C

0: AGCTATGGA

1: AGCTAGCCT

2: GCATGCGGA

① slide

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNTGGANNNNNNNNN

② retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

C

0: AGCTATGGA

1: AGCTAGCCT

2: GCATGCGGA

① slide

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNTGGANNNNNNNNN

② retrieve

③ retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

d

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

d

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNTGGANNNNNNNNN

↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

d

AGCTATGGA

id:0

AGCTAGCCT

id:1

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

d

② extend

AGCTA→TGGGA

AGCTAGCCT

id:0

id:1

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

d

② extend

AGCTATGGA

AGCTAGCCT

id:0

id:1

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

① retrieve

③ add in

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6



e

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6

e

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNTGGANNNNNNNNN

① retrieve ↓

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6

e

GCATGCGGA id:2

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

① retrieve ↓

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6

e

② extend
GCATGCGGA

id:2

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNN

① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6

e

② extend

GCATGCGGA

id:2

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNN

① retrieve

③ add in

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
QT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

f

TAGCTAGNNGCATGCGNCTNTGGANNNNNTGGANNNNNNNN

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

f

TAGCTAGNNGCATGCGNNCTNTGGANNNNNTGGANNNNNNNNN

↓ retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

f

TAGCTAGNNGCATGCGNCTNTGGANNNNNTGGANNNNNNNNN

↓ ①retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

f

AGCTAGCCT **id:1**

TAGCTAGNNGCATGCGNNCTNTGGANNNNNNNTGGANNNNNNNNNN

↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

f

[1,7)
AGCTAGCCT id:1

TAGCTAGNNGCATGCGNNCTNTGGANNNNNNNTGGANNNNNNNNNN

↓ ①retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

f

[1,7) ②extend
AGCTAGCCT id:1

TAGCTAGNNGCATGCGNNCTNTGGANNNNNNNTGGANNNNNNNNNN
↓①retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

[1,7) [18,21)
 AGCTAGCCT id:1

TAGCTAGNNGCATGCGNNCTNTGGANNNNNNTGGANNNNNNNNN

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

[1,7) [18,21) -----> ③report
 AGCTAGCCT id:1

TAGCTAGNNGCATGCGNNCTNTGGANNNNNNTGGANNNNNNNNN

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

h

TAGCTAGNNGCATGCGNNCCTNTGGA
NNNNNNTGGANNNNNNNNNN

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

h

TAGCTAGNNGCATGCGNNCCTNTGGA
NNNNNNTGGANNNNNNNNNN

retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

h

TAGCTAGNNGCATGCGNNCCTNTGGA^①NNNNNNTGGANNNNNNNNNN

↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

h

AGCTATGGA **id:0**

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

h

[1,6)
AGCTATGGA id:0

TAGCTAGNNGCATGCGNNCCTNTGGAANNNNNNNTGGANNNNNNNNNN

↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

h

← ② extend

[1,6)
AGCTATGGA id:0

TAGCTAGNNGCATGCGNNCCTNTGGAANNNNNNNTGGANNNNNNNNNN

↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

h

← ② extend
[1,6) [22,26)
AGCTATGGA id:0

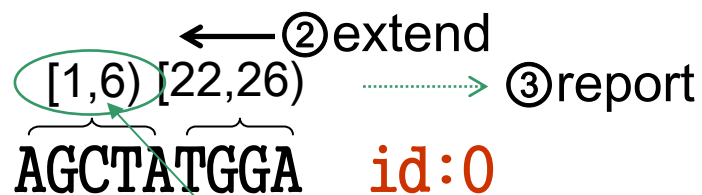
TAGCTAGNNGCATGCGNNCCTNTGGAANNNNNNNTGGANNNNNNNNNN

↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

h



TAGCTAGNNGCATGCGNNCCTNTGGAANNNNNNNTGGANNNNNNNNNN
↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

GCATGCGGA **id:2**

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

[9,16)
GCATGCGGA id:2

TAGCTAGNNGCATGCGNNCCTNTGGAANNNNNNNTGGANNNNNNNNNN

↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

i

[9,16) ← ② extend
GCATGCGGA id:2

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

[9,16) ← ② extend
GCATGCGGA id:2
[23,26)

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

[9,16) ← ② extend
GCATGCGGA id:2
[23,26)

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

j

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGA
NNNNNNNNNN

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

j

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGA

retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

j

AGCTATGGA id:0

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNNN

↓ ① retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

j

[1,6)
AGCTATGGA id:0

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNNN
↓ ①retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

j

[1,6) ← ②extend
AGCTATGGA id:0

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNNN
↓ ①retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

j

[1,6) [32,36) ← ②extend
AGCTATGGA id:0

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNNN
↓ ①retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

j

↑ ③check [1,6) [32,36) ← ②extend

AGCTATGGA id:0

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNNN

↓ ①retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

intron_len: 36(end)-1(start)-9(read_len) > 20

j

③check [1,6) [32,36) ← ②extend

AGCTATGGA id:0

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGA

①retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

intron_len: 36(end)-1(start)-9(read_len) > 20 **invalid** **j**

↑ ③check [1,6) [32,36) ← ②extend

AGCTATGGA **id:0**

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNNN

↓ ①retrieve

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

k

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNNN

①check
↓

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

k

current_pos-start-read_len > max_intron_len ?

①check

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNNN

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

k

current_pos - start - read_len > max_intron_len ?

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNNN

① check

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5
CT	id:1/start:1/matched:6
GA	id:2/start:9/matched:7

② delete

k

current_pos - start - read_len > max_intron_len ?

① check

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNNN

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0/start:1/matched:5	② delete
CT	id:1/start:1/matched:6	② delete
GA	id:2/start:9/matched:7	

k

current_pos - start - read_len > max_intron_len ?

① check

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNNN

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:0 / start:1 / matched:5	② delete
CT	id:1 / start:1 / matched:6	② delete
GA	id:2 / start:9 / matched:7	

k

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGA
NNNNNNNNNN

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:2/start:9/matched:7

k

TAGCTAGNNGCATGCGNNCCTNTGGANNNN**NTGGAN**NNNNNNNN

AGCTA	id:0
AGCTA	id:1
GCATG	id:2

GA	id:2/start:9/matched:7

0 : AGCTATGGA

I

1 : AGCTAGCCT

2 : GCATGCGGA

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

0: AGCTATGGA

1: AGCTAGCCT

2: GCATGCGGA

I

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

0: AGCTATGGA

1: AGCTAGCCT

2: GCATGCGGA



TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNN

add in

ATGGA	id:0
AGCCT	id:1
GCGGA	id:2

0: AGCTATGGA

1: AGCTAGCCT

2: GCATGCGGA

I

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNNN

add in

ATGGA	id:0
AGCCT	id:1
GCGGA	id:2

0: AGCTATGGA

1: AGCTAGCCT

2: GCATGCGGA

← slide

TAGCTAGNNGCATGCGNNCCTNTGGANNNNNNTGGANNNNNNNNNNN

add in

ATGGA	id:0
AGCCT	id:1
GCGGA	id:2