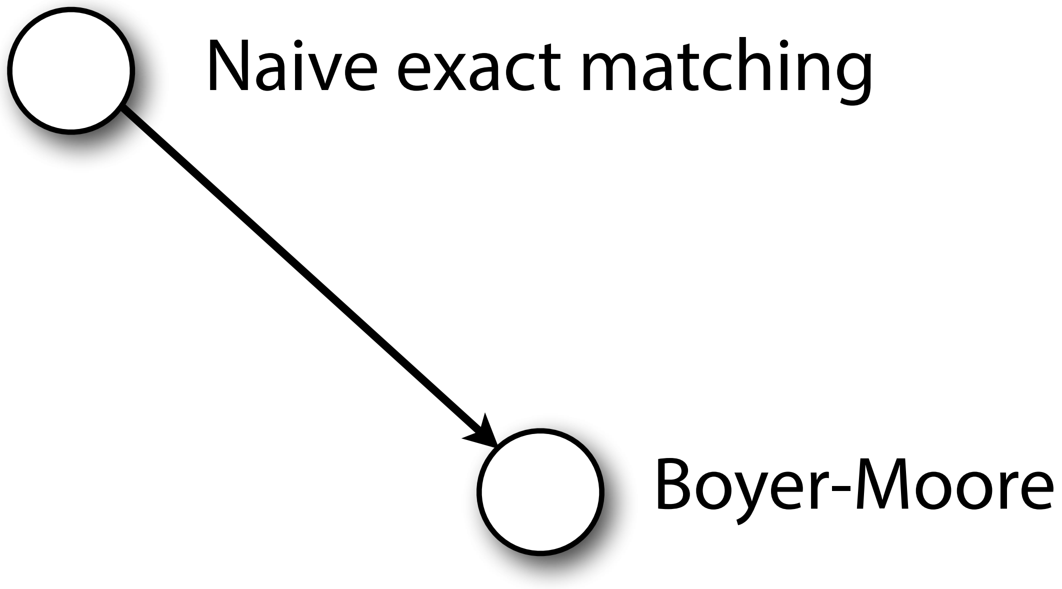
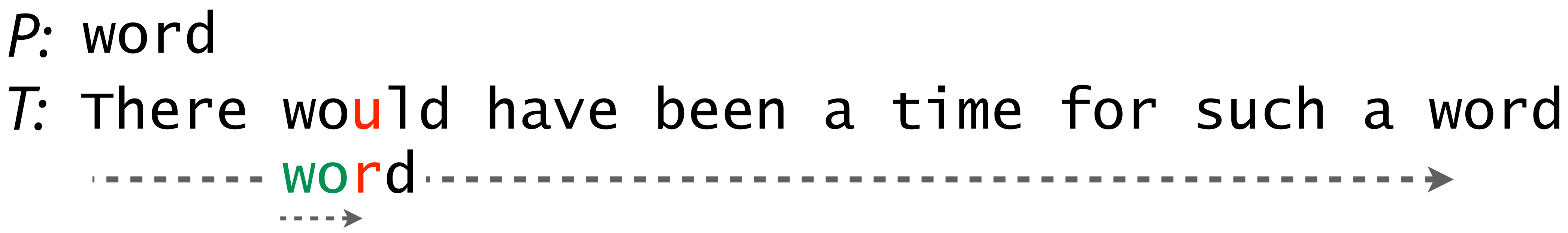
Boyer-Moore



Naive exact matching

Boyer-Moore

Exact matching: better naïve algorithm

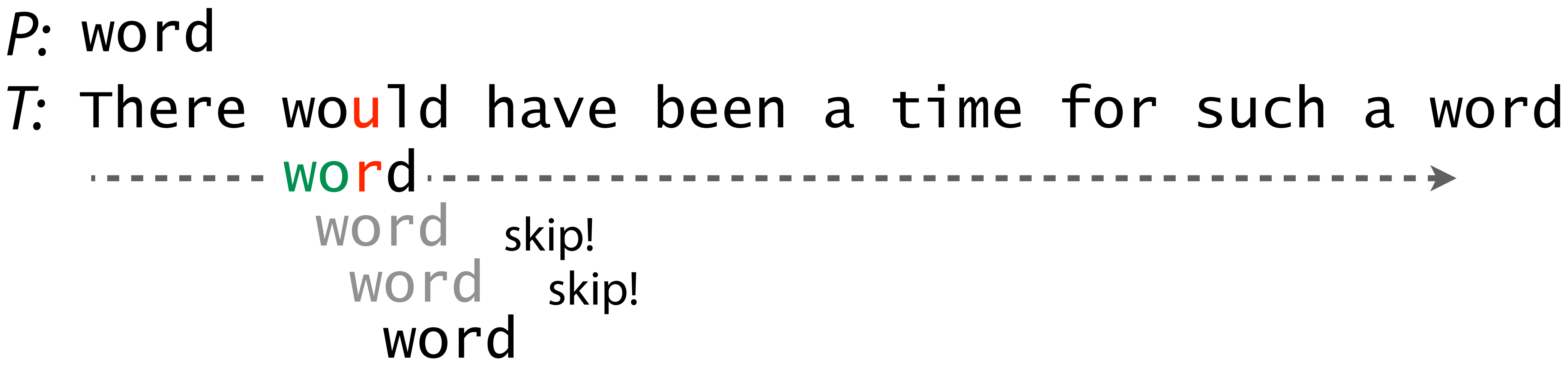


P: word

T: There would have been a time for such a word

word

u doesn't occur in *P*, so we can skip next two alignments



P: word

T: There would have been a time for such a word

word

word skip!

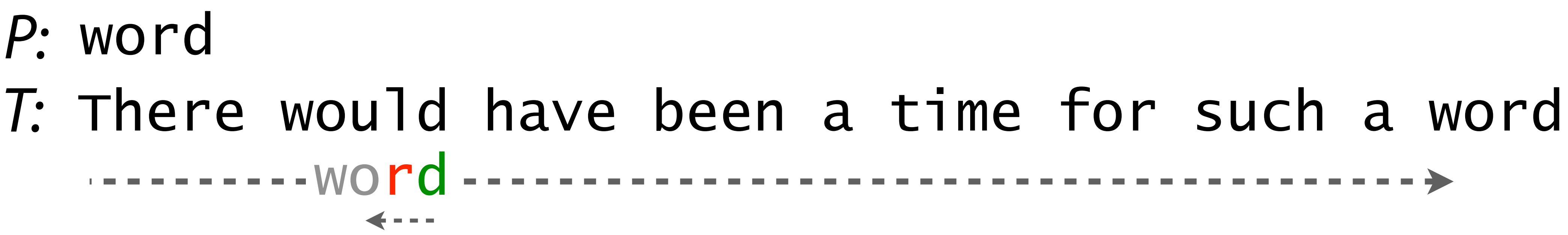
word skip!

word

Boyer-Moore

Learn from character comparisons to skip pointless alignments

Try alignments in left-to-right order, and try character comparisons in right-to-left order



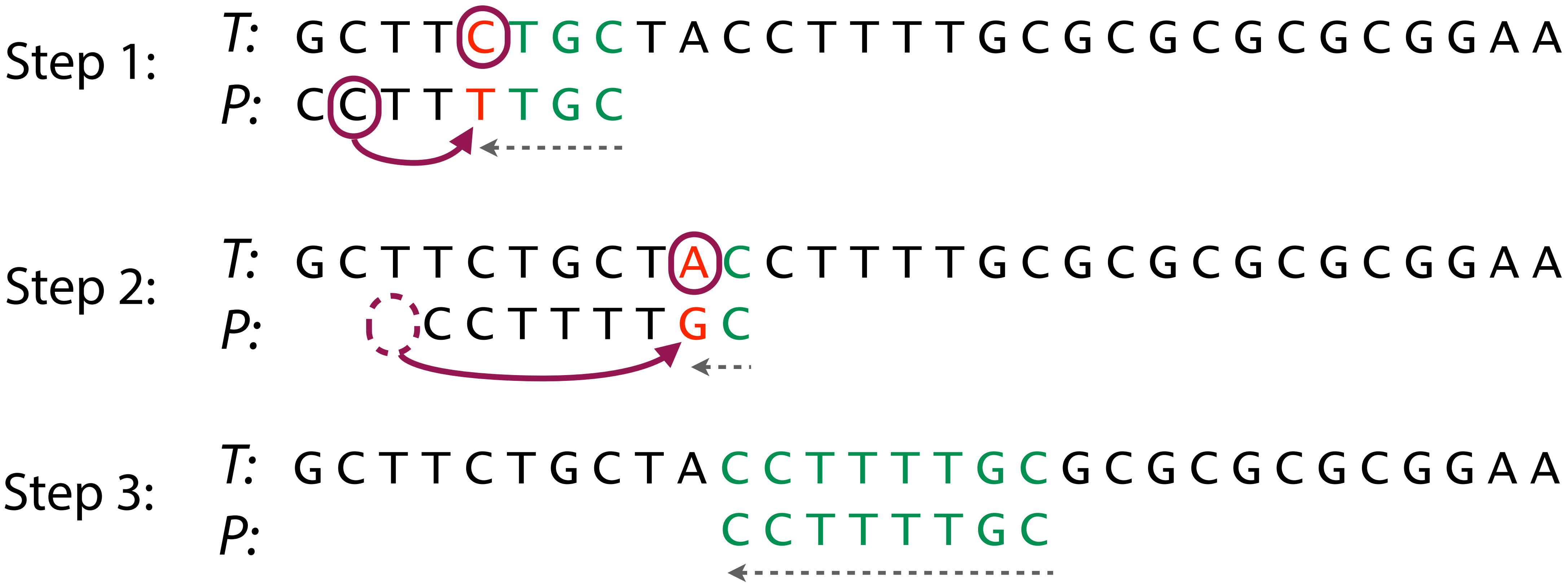
P: word

T: There would have been a time for such a word

Boyer, RS and Moore, JS. "A fast string searching algorithm." *Communications of the ACM* 20.10 (1977): 762-772.

Boyer-Moore: Bad character rule

Upon mismatch, skip alignments until (a) mismatch becomes a match, or (b) *P* moves past mismatched character



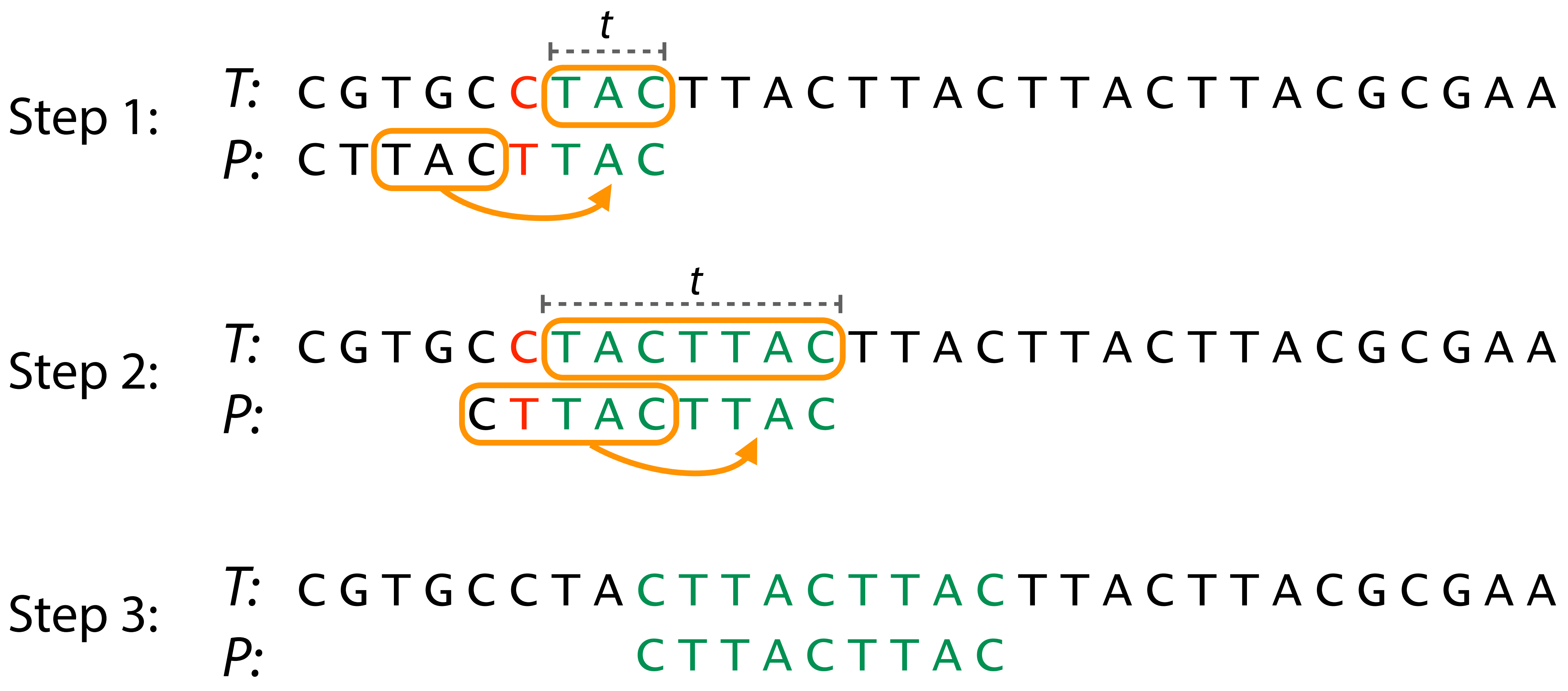
Step 1:
T: G C T T C T G C T A C C T T T T G C G C G C G C G G A A
P: C C T T T T G C

Step 2:
T: G C T T C T G C T A C C T T T T G C G C G C G C G G A A
P: C C T T T T G C

Step 3:
T: G C T T C T G C T A C C T T T T G C G C G C G C G G A A
P: C C T T T T G C

Boyer-Moore: Good suffix rule

Let *t* = substring matched by inner loop; skip until (a) there are no mismatches between *P* and *t* *or* (b) *P* moves past *t*



t

Step 1: *T*: CGTGCCTACTTACTTACTTACTTACGCGAA

P: CTTACTTAC

t

Step 2: *T*: CGTGCCTACTTACTTACTTACTTACGCGAA

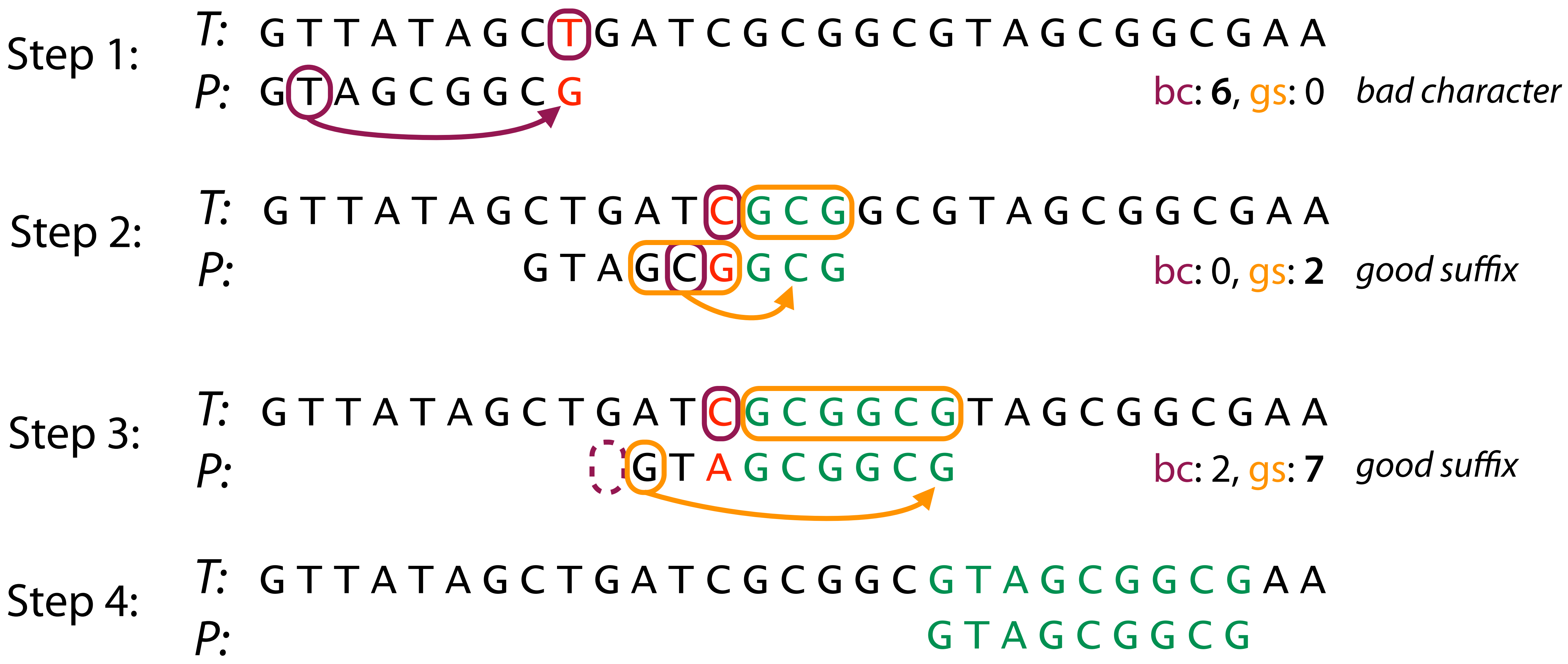
P: CTTACTTAC

Step 3: *T*: CGTGCCTACTTACTTACTTACTTACGCGAA

P: CTTACTTAC

Boyer-Moore: Putting it together

Use bad character or good suffix rule, *whichever skips more*

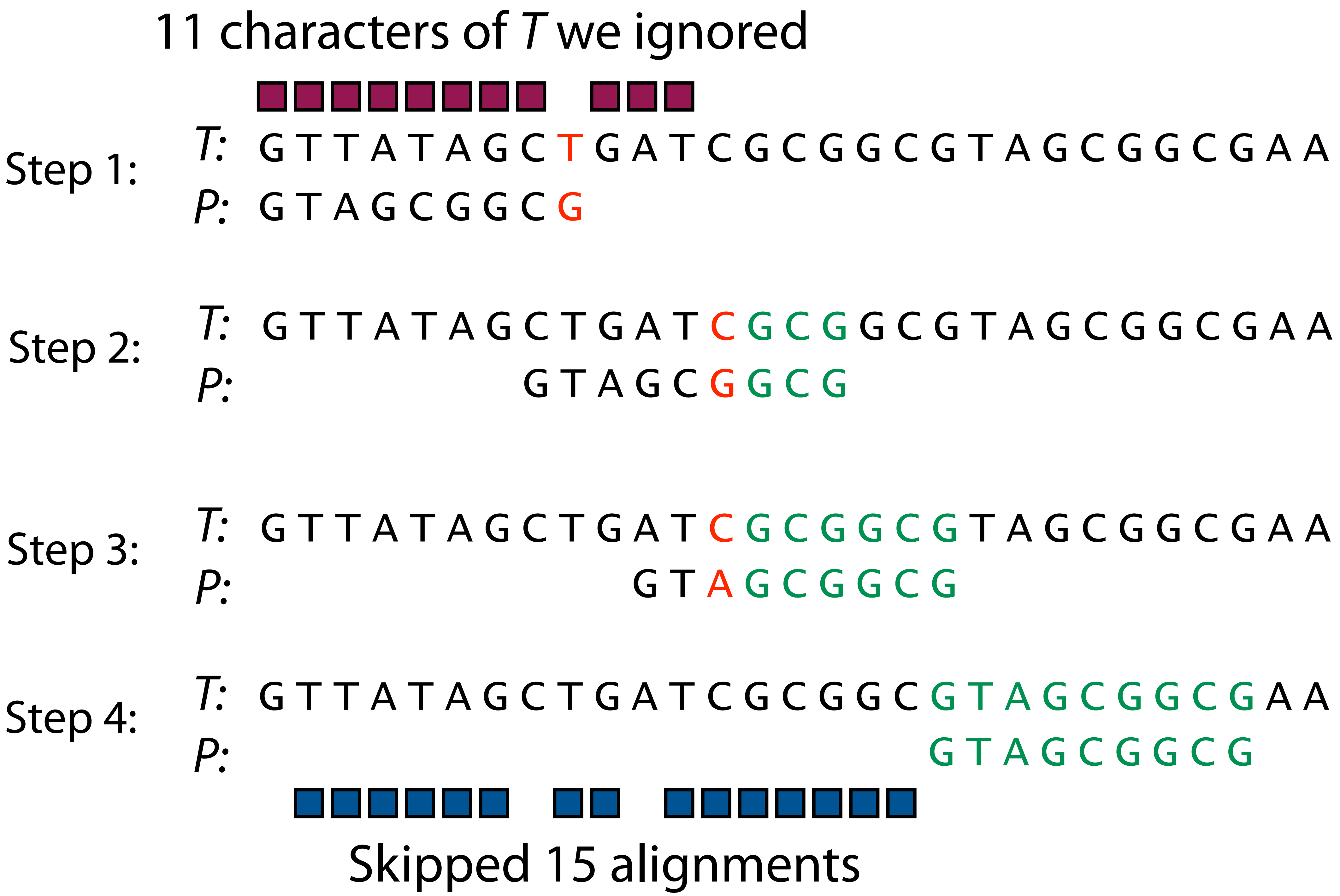


Step 1:
T: GTTATAGCTGATCGCGGCGTAGCGGCGAA
P: GTAGCGGCG
bc: **6**, gs: 0 *bad character*

Step 2:
T: GTTATAGCTGATCGCGGCGTAGCGGCGAA
P: GTAGCGGCG
bc: 0, gs: **2** *good suffix*

Step 3:
T: GTTATAGCTGATCGCGGCGTAGCGGCGAA
P: GTAGCGGCG
bc: 2, gs: **7** *good suffix*

Step 4:
T: GTTATAGCTGATCGCGGCGTAGCGGCGAA
P: GTAGCGGCG



11 characters of *T* we ignored

Step 1:
T: G T T A T A G C T G A T C G C G G C G T A G C G G C G A A
P: G T A G C G G C G

Step 2:
T: G T T A T A G C T G A T C G C G G C G T A G C G G C G A A
P: G T A G C G G C G

Step 3:
T: G T T A T A G C T G A T C G C G G C G T A G C G G C G A A
P: G T A G C G G C G

Step 4:
T: G T T A T A G C T G A T C G C G G C G T A G C G G C G A A
P: G T A G C G G C G

Skipped 15 alignments

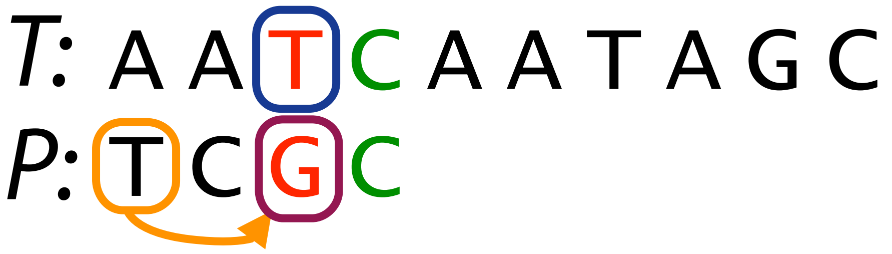
Boyer-Moore: Preprocessing

Pre-calculate skips. For bad character rule, *P* = TCGC:

P

Σ

	T	C	G	C
A	0	1	2	3
C	0	-	0	-
G	0	1	-	0
T	-	0	1	2



T: A A T C A A T A G C

P: T C G C