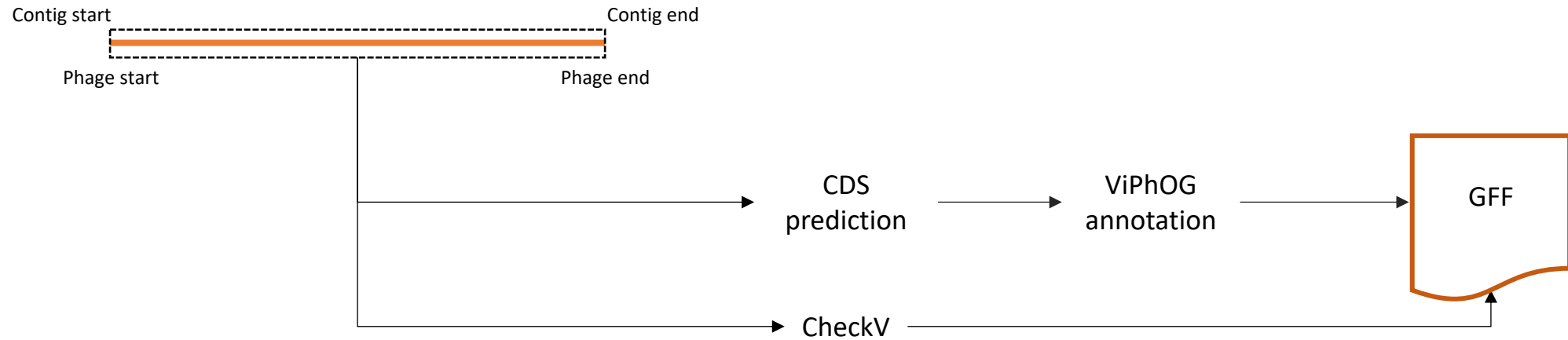
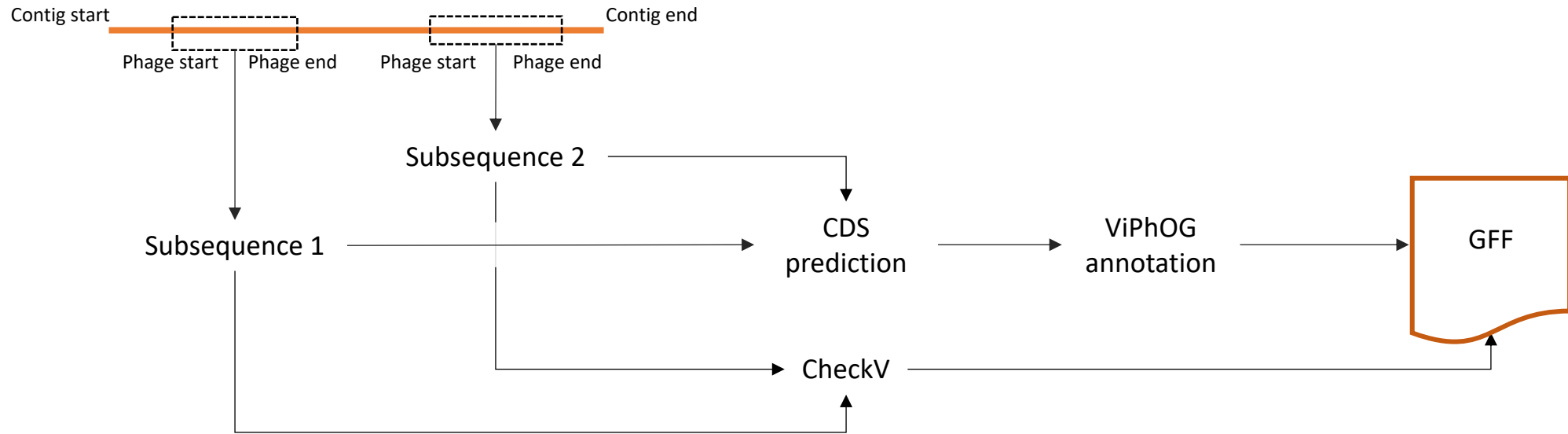


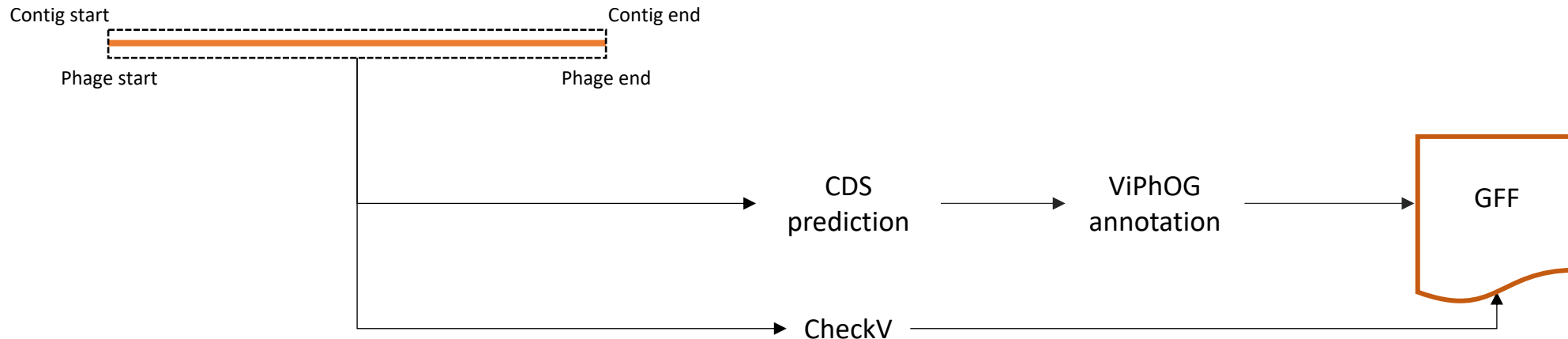
### - Viral genome in contig



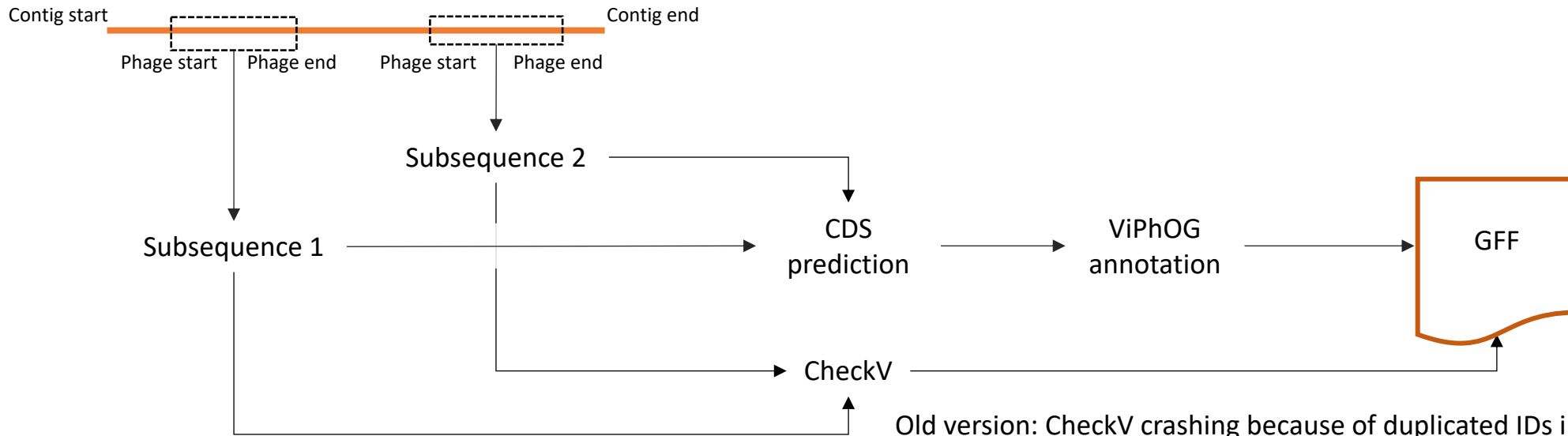
### - Prophage(s)



### - Viral genome in contig

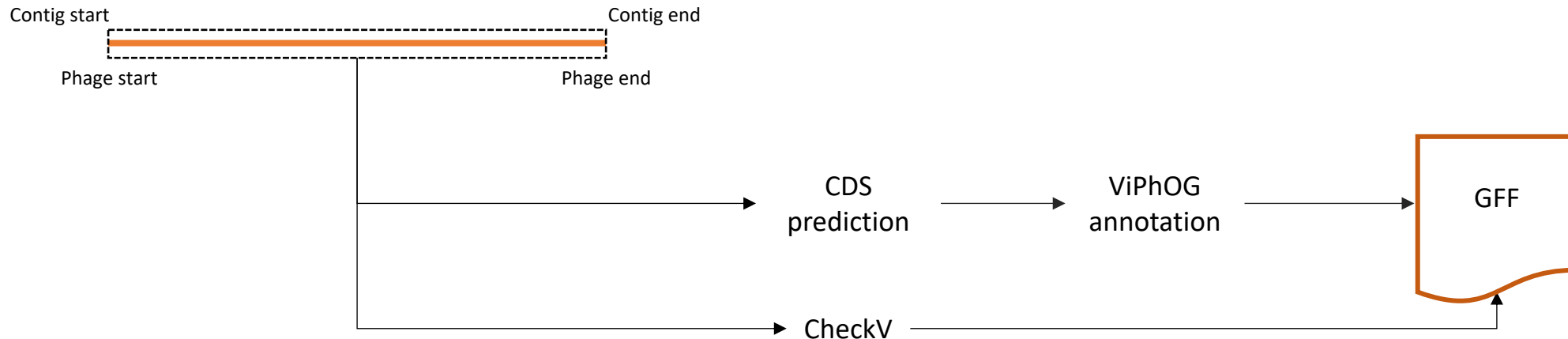


### - Prophage(s)

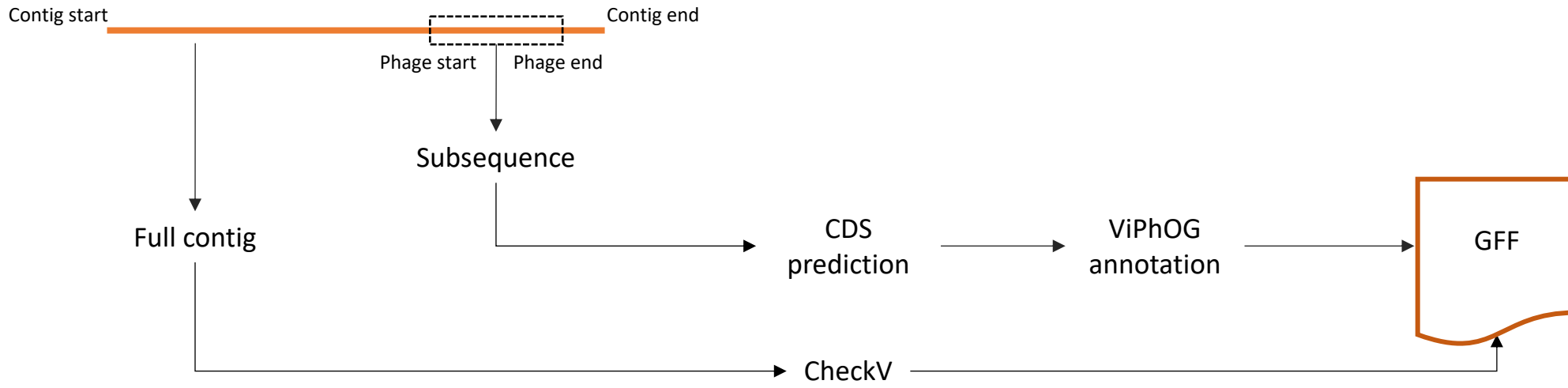


Old version: CheckV crashing because of duplicated IDs in fasta  
We should paste the coordinates to sequence ID

- Viral genome in contig



- Prophage(s)



Updated version: CheckV underestimating quality  
We should run the program per subsequence

```
(base) [ales@hl-codon-131-04 annotation]$ cat ../../prophages_original.fasta | grep '>'
>NODE_8_length_177745_cov_9_022404 prophage-157838:177257
>NODE_11_length_150587_cov_8_976344 prophage-0:22596
>NODE_13_length_149129_cov_8_941559 prophage-87235:123684
>NODE_26_length_89261_cov_9_171872 prophage-0:9742
>NODE_32_length_78548_cov_9_424853 prophage-43951:60107
>NODE_39_length_70965_cov_9_657030 prophage-34220:70894
```

Predicted prophages  
(fasta)

CDS calling and  
functional annotation on  
subseq

CDS coordinates  
relative to subseq

Transferring annotation  
to gff and metadata

Prediction coordinates  
adjusted to the first and  
last ViPhOG match

NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_1	50	820	-1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_2	905	1075	-1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_3	1278	1478	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_4	1526	1885	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_5	2344	2697	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_6	2719	3234	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_7	3231	3788	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_8	3941	4267	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_9	4264	5151	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_10	5144	5677	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_11	5679	7787	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_12	7796	8236	1	ViPhOG26107.faa	17
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_13	8279	9439	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_14	9452	9955	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_15	9970	10314	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_16	10283	10393	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_17	10484	12721	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_18	12731	13603	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_19	13578	13784	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_20	13842	14831	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_21	14864	15493	1	ViPhOG24210.faa	22
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_22	15463	15852	1	No hit	NA
NODE_32_length_78548_cov_9_424853	NODE_32_length_78548_cov_9_424853_23	15849	16106	1	No hit	NA

```
(base) [ales@hl-codon-131-04 gff]$ cat contigs_virify_contig_viewer_metadata.tsv | cut -f1
sequence_id
NODE_8_length_177745_cov_9_022404-start-277-end-11967
NODE_11_length_150587_cov_8_976344-start-2-end-583
NODE_13_length_149129_cov_8_941559-start-18043-end-32007
NODE_32_length_78548_cov_9_424853-start-7796-end-15493
NODE_39_length_70965_cov_9_657030-start-5184-end-34797
NODE_14_length_148715_cov_8_369494-start-110839-end-132744
NODE_57_length_35549_cov_8_337578-start-22881-end-23174
```

NODE_32_length_78548_cov_9_424853	ViPhOG26107	viral_sequence	7796	8236	.	+	.	ID=NODE_32_length_78548_cov_9_424853;viphog=ViPhOG26107;viphog_t
NODE_32_length_78548_cov_9_424853	ViPhOG24210	viral_sequence	14864	15493	.	+	.	ID=NODE_32_length_78548_cov_9_424853;viphog=ViPhOG24210;viphog_t

Related with the integration of predicted phages and quality

1. CheckV is running in the full contig for prophages
2. Multiple phages occurring in the same contig are missing
3. Prophages coordinates in the gff not corresponds to coordinates prediction
4. Prophages with no CDS matching ViPhOG are excluded in the gff
5. Viral genomes in full contigs are missing in the gff output

Related with the GFF file format

1. Protein IDs in gff file are missing
2. Add the `mobile_genetic_element` feature and the corresponding coordinates
3. Fix the CDS coordinates in the context of the full contig instead of the subseq