

A

Distinct CDS

Protein ID

Genome
integer IDs

ATGGTACAA...ATCAGATAG

[49, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]



CDS	Genome IDs
ATGGTACAA...ATCAGATAG	[49, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
ATGAGCAAG...CCAAAATAA	[1762, 1, 2, 3, 4, 5, 7, 8, 9, 11]
ATGAGGAAA...AAGAAATGA	[413, 18, 30, 80, 89, 118]

↓ Hashing
(SHA-256)

↓ Modified polyline
encoding

Hashed CDS	Encoded IDs
6ab9534de...ab41432c6	?aB~AAAAAAAAAAAA
070c540e0...f8c973ea8	?cmB`mBAAAACAAC
ab26aea08...469ed9f5b	?yXtWWcBQy@

B

Distinct
translated CDSPairs of protein and
genome integer IDs

MKILVTTFD...LKSVGGATH

[1219, 123, 1761, 92]



Translated CDS	Genome IDs
MKILVTTFD...LKSVGGATH	[1219, 123, 1761, 92]
MKILVWGFD...LKRVS GATH	[1092, 155, 1239, 220]
MKILVMGFD...LKMVG GATH	[639, 1, 1338, 57, 713, 250]

↓ Hashing
(SHA-256)

↓ Modified polyline
encoding

Hashed translated CDS	Encoded IDs
8f47a94cb...4227c684f	?ekAncAkeBhgB
300b1e083...7edd81574	?gcApy@wbAt~@
00c7ff34a...cd638d97f	?}f@zf@qrA`oA_h@