

# Methods to resolve this: Convert Y to (C or T)

lcl consensus	A	A	A	G	A	A	Y	G
gb OQ963804.1  [organism=Ir	A	A	A	G	A	A	C	G
gb OQ958820.1  [organism=Ir	A	A	A	G	A	A	T	G
gb OQ958812.1  [organism=Ir	A	A	A	G	A	A	T	G
gb OQ958804.1  [organism=Ir	A	A	A	G	A	A	T	G
gb OQ961564.1  [organism=Ir	A	A	A	G	A	A	T	G
gb OQ963300.1  [organism=Ir	A	A	A	G	A	A	C	G
gb OQ961548.1  [organism=Ir	A	A	A	G	A	A	C	G
gb OQ961556.1  [organism=Ir	A	A	A	G	A	A	C	G

We don't know the ground truth, this introduces bias into our dataset.

This approach does not scale.

Builds bad habits.

lcl consensus	A	A	A	G	A	A	T	G	T	C	A	C	T	G	T	T	A	C	A	C	A	T	G
gb OQ963804.1  [organism=Ir	A	A	A	G	A	A	C	G	T	C	A	C	T	G	T	T	A	C	A	C	A	T	G
gb OQ958820.1  [organism=Ir	A	A	A	G	A	A	T	G	T	C	A	C	T	G	T	T	A	C	A	C	A	T	G
gb OQ958812.1  [organism=Ir	A	A	A	G	A	A	T	G	T	C	A	C	T	G	T	T	A	C	A	C	A	T	G
gb OQ958804.1  [organism=Ir	A	A	A	G	A	A	T	G	T	C	A	C	T	G	T	T	A	C	A	C	A	T	G
gb OQ961564.1  [organism=Ir	A	A	A	G	A	A	T	G	T	C	A	C	T	G	T	T	A	C	A	C	A	T	G
gb OQ963300.1  [organism=Ir	A	A	A	G	A	A	C	G	T	C	A	C	T	G	T	T	A	C	A	C	A	T	G
gb OQ961548.1  [organism=Ir	A	A	A	G	A	A	C	G	T	C	A	C	T	G	T	T	A	C	A	C	A	T	G
gb OQ961556.1  [organism=Ir	A	A	A	G	A	A	C	G	T	C	A	C	T	G	T	T	A	C	A	C	A	T	G

# Methods to resolve this: Mask the codon

								120										130		
Id consensus	A	A	A	G	A	A	Y	G	T	C	A	C	T	G	T	T	A	C	A	C
gb OQ963804.1  [organism=Ir	A	A	A	G	A	A	C	G	T	C	A	C	T	G	T	T	A	C	A	C
gb OQ958820.1  [organism=Ir	A	A	A	G	A	A	T	G	T	C	A	C	T	G	T	T	A	C	A	C
gb OQ958812.1  [organism=Ir	A	A	A	G	A	A	T	G	T	C	A	C	T	G	T	T	A	C	A	C
gb OQ958804.1  [organism=Ir	A	A	A	G	A	A	T	G	T	C	A	C	T	G	T	T	A	C	A	C
gb OQ961564.1  [organism=Ir	A	A	A	G	A	A	T	G	T	C	A	C	T	G	T	T	A	C	A	C
gb OQ963300.1  [organism=Ir	A	A	A	G	A	A	C	G	T	C	A	C	T	G	T	T	A	C	A	C
gb OQ961548.1  [organism=Ir	A	A	A	G	A	A	C	G	T	C	A	C	T	G	T	T	A	C	A	C
gb OQ961556.1  [organism=Ir	A	A	A	G	A	A	C	G	T	C	A	C	T	G	T	T	A	C	A	C

		110										120								
		A	T	A	A	T	G	G	A	A	A	A	G	-	-	-	G	T	C	A
lcl consensus		A	T	A	A	T	G	G	A	A	A	A	G	-	-	-	G	T	C	A
gb OQ963804.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ958820.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ958812.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ958804.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ961564.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ963300.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ961548.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ961556.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ962772.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ959564.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ961068.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ958780.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ958772.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ965100.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ958764.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ958756.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ963812.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ965172.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ964212.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	C	G	T	C	A
gb OQ967076.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ964204.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ961228.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ965188.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ963252.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ960660.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ958788.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ958796.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ959068.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A
gb OQ962516.1	[organism=Influ	A	T	A	A	T	G	G	A	A	A	A	G	A	A	T	G	T	C	A