cd hit cluster 1138/SAUSA300_2158/hypothetical protein 2 independent mutations G101E Y88F