cd hit cluster 564/SAUSA300_2101/NE372/hypothetical protein 3 independent mutations 197N LS88RI Y93F Ter99Lext*? V207G