cd hit cluster 1081/SAUSA300_0652/NE1797/hypothetical protein 2 independent mutations E189G Q14R