cd hit cluster 514/agrC/SAUSA300\_1991/NE873/hypothetical protein 4 independent mutations D284G 1139fs 145fs L225\* I138K V392fs