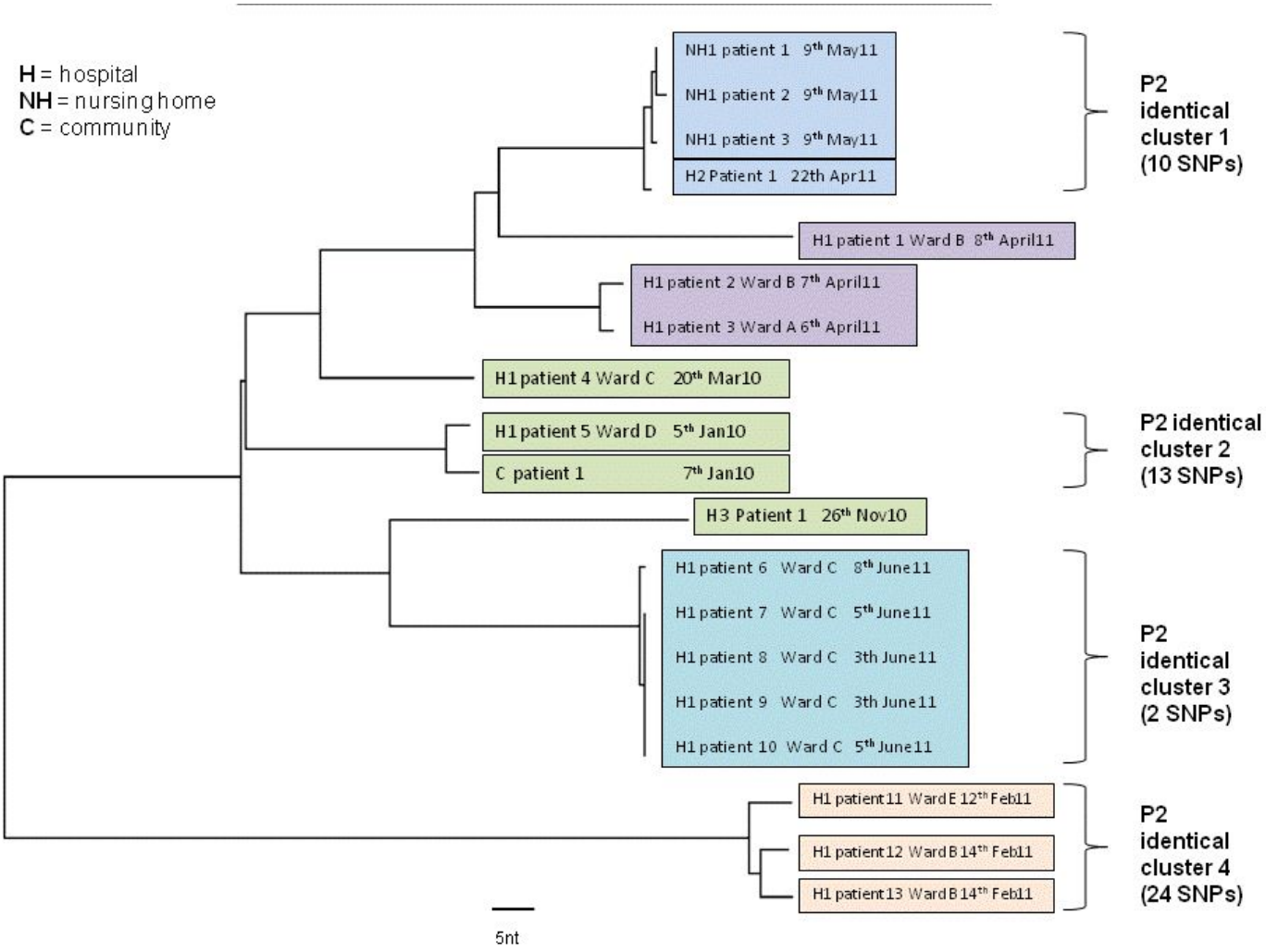
O186

Abstract (oral session)

Using whole genome sequence data to determine transmission paths and mutation rates in Norovirus outbreaks

T.H.N. Wong*, B. Dearlove, D.W Crook, T.E Peto, D.J Wilson, K.E Dingle (Oxford, UK)

Objectives: Noroviruses are estimated to cause 80-95% of all gastroenteritis cases worldwide, representing a considerable public health burden. The genotype GII.4 lineage is predominant in healthcare settings. The nucleotide sequence of the capsid P2 domain has been used to discriminate outbreaks. The aim of this study was to use whole genome sequencing to provide detail to elucidate transmission events within outbreaks. Methods: 19 complete norovirus genomes were determined by Sanger sequencing. These represented 5 separate outbreaks and 2 unlinked individuals, submitting samples between Jan 2010-June 2011. Detailed corresponding epidemiological data was collected. The BEAST (Bayesian evolutionary analysis by sampling trees) package was used to reconstruct relationships and determine the rate of evolution within the defined timeframe. Results: The 19 complete GII.4 genomes (7560 nucleotides (nt) with 569 single nucleotide polymorphisms (SNPs)) were included in our analysis. The rate of evolution was estimated at 5 x 10-3 mutations per site per year (~3 mutations/genome/month), comparable with recent reports of rates between 4.3-9.0x10-3 mutations per site per year, calculated using Bayesian analysis of the GII.4 capsid (1623nt) and polymerase (247nt) fragments of the genome. Isolates within 4 of 5 outbreak clusters were identical according to the capsid P2 domain (455nt) sequence alone. However, all but four isolates had one or more SNPs by whole genome sequencing (WGS)(see figure). WGS confirmed one cluster of 5 isolates containing 4 identical isolates and the fifth containing only 2 SNPs. In the 4 clusters each containing identical P2 domains, WGS revealed 2, 10, 13 and 24 SNPs within each cluster suggesting that these were part of larger ongoing outbreaks. Conclusion: Complete genome sequences, combined with detailed epidemiological data, provide sufficient resolution to identify transmission within hospital. The conventionally used P2 domain, however, is not adequate to discriminate between closely related isolates within outbreaks. Next generation sequencing technology will allow large numbers of whole genomic sequences to be generated rapidly informing “real time” outbreak control.


H = hospital
NH = nursing home
C = community
NH1 patient 1 9th May11
NH1 patient 2 9th May11
NH1 patient 3 9th May11
H2 Patient 1 22th Apr11
H1 patient 1 Ward B 8th April11
H1 patient 2 Ward B 7th April11
H1 patient 3 Ward A 6th April11
H1 patient 4 Ward C 20th Mar10
H1 patient 5 Ward D 5th Jan10
C patient 1 7th Jan10
H3 Patient 1 26th Nov10
H1 patient 6 Ward C 8th June11
H1 patient 7 Ward C 5th June11
H1 patient 8 Ward C 3th June11
H1 patient 9 Ward C 3th June11
H1 patient 10 Ward C 5th June11
H1 patient11 Ward E 12th Feb11
H1 patient12 Ward B 14th Feb11
H1 patient13 Ward B 14th Feb11
5nt
P2 identical cluster 1 (10 SNPs)
P2 identical cluster 2 (13 SNPs)
P2 identical cluster 3 (2 SNPs)
P2 identical cluster 4 (24 SNPs)