Folkhälsomyndigheten
PUBLIC HEALTH AGENCY OF SWEDEN

Karolinska Institutet

Evolution of the *E. coli* ST131 sub-clone H30-Rx/C2

- high-coverage sequencing reveals greater diversity than previously reported

Sofia Ny, Erik Alm, Linus Sandegren, Christian Giske

sofia.ny@folkhalsomyndigheten.se
PhD-student

Transparency declaration: no disclosures

E. coli ST131 and its subclones

- Interesting from many perspectives
  - Subclone (H30-Rx/C2) overrepresented among clinical CTX-M producing ST131
  - Indicates variations in pathogenicity within ST131

- Well studied and characterized
  - Several phylogenetic studies
  - Suggested that spread of H30-Rx/C2 started in the late 80s and is mainly clonal
  - Suggested genomic drift 1 SNP per year

We wanted to…

- Investigate prevalence of H30-Rx/C2 in ESBL-producing ST131

- Prevalence of ST131 among ESBL-producing *E. coli*
  (Ny et al. JAC 2017)
  - Bloodstream infections 45% (63% of these were H30-Rx/C2)
  - Community carriers 17% (38% of these were H30-Rx/C2)

- Confirm H30-Rx/C2 status
  - Compared with sequencing 2 SNPs (according to Price et al. mBio 2013)
  - Compare with published genomic dataset of ST131 and H30-Rx/C2

- Investigate whole and core genome variations using
  - Whole genome based SNV analysis
  - cgMLST in EnteroBase

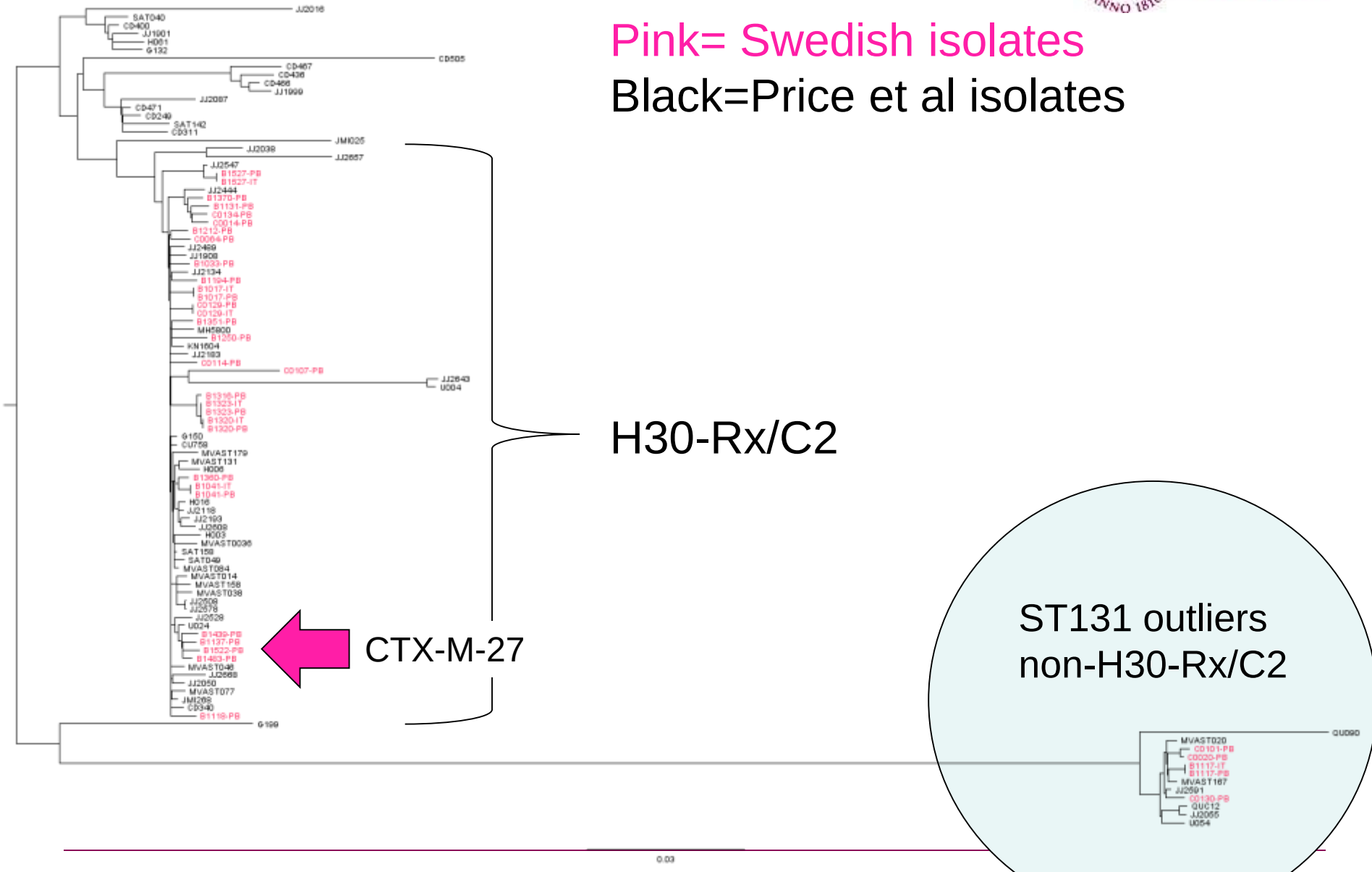
Information about sequencing

- **PacBio RS II (n=29)**
  - 4 non-H30-Rx/C2 (CTX-M-15)
  - 4 non-H30-Rx/C2 (CTX-M-27)
  - 21 H30-Rx/C2 (CTX-M-15)

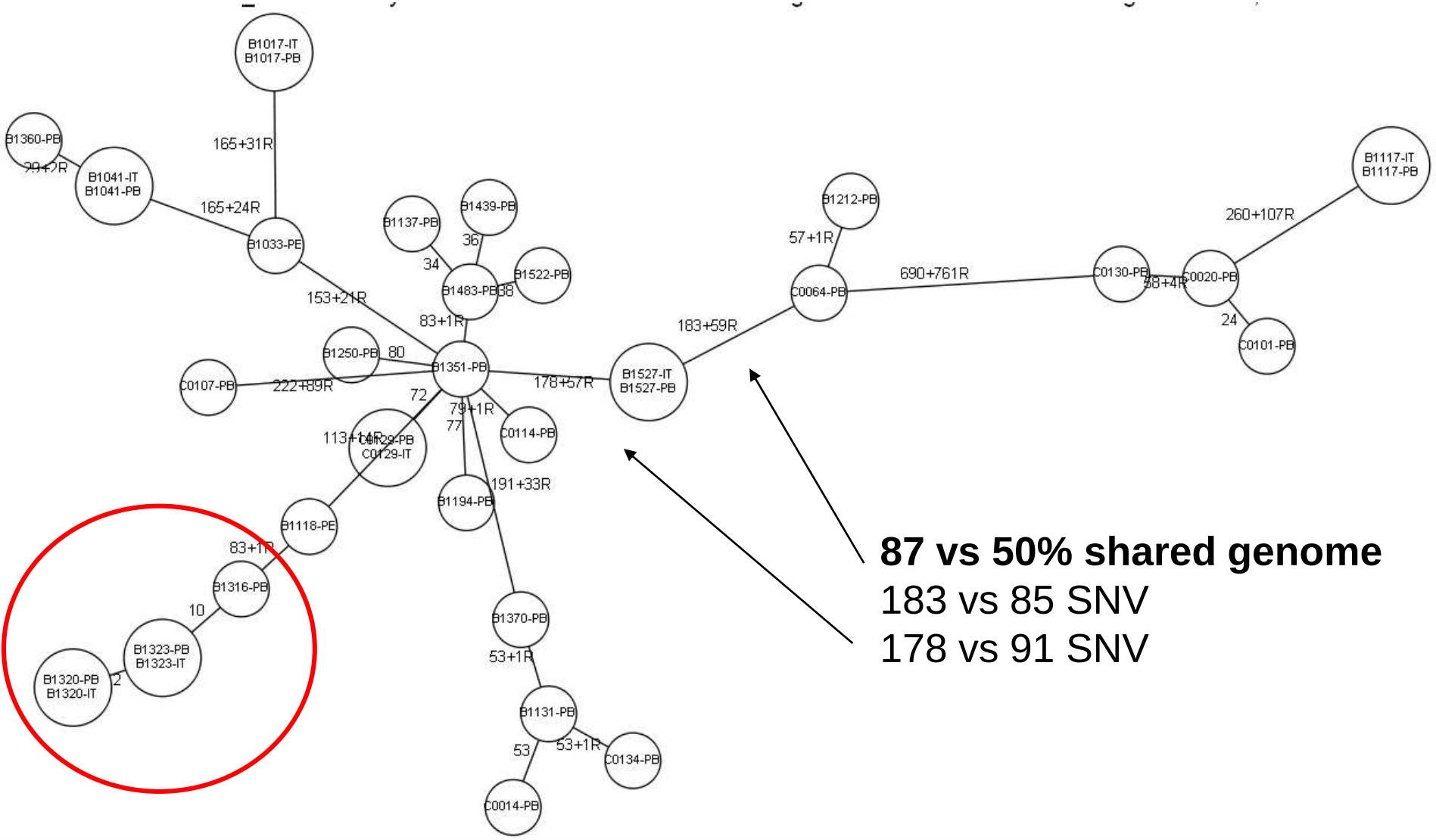
- **Ion Torrent**
  - 7 duplicates

- Reference genome EC958 (closed H30-Rx/C2 genome)
- Coverage in average >50x for PacBio
- Coverage cut-off 10x
- Compare with Price et al. mBio 2013 dataset

Maximum likelihood tree with published H30-Rx/C2 strains (50% shared genome)



Minimum spanning tree based on SNV differences - Only Swedish strains (87% shared genome)



How to relate to higher differences?

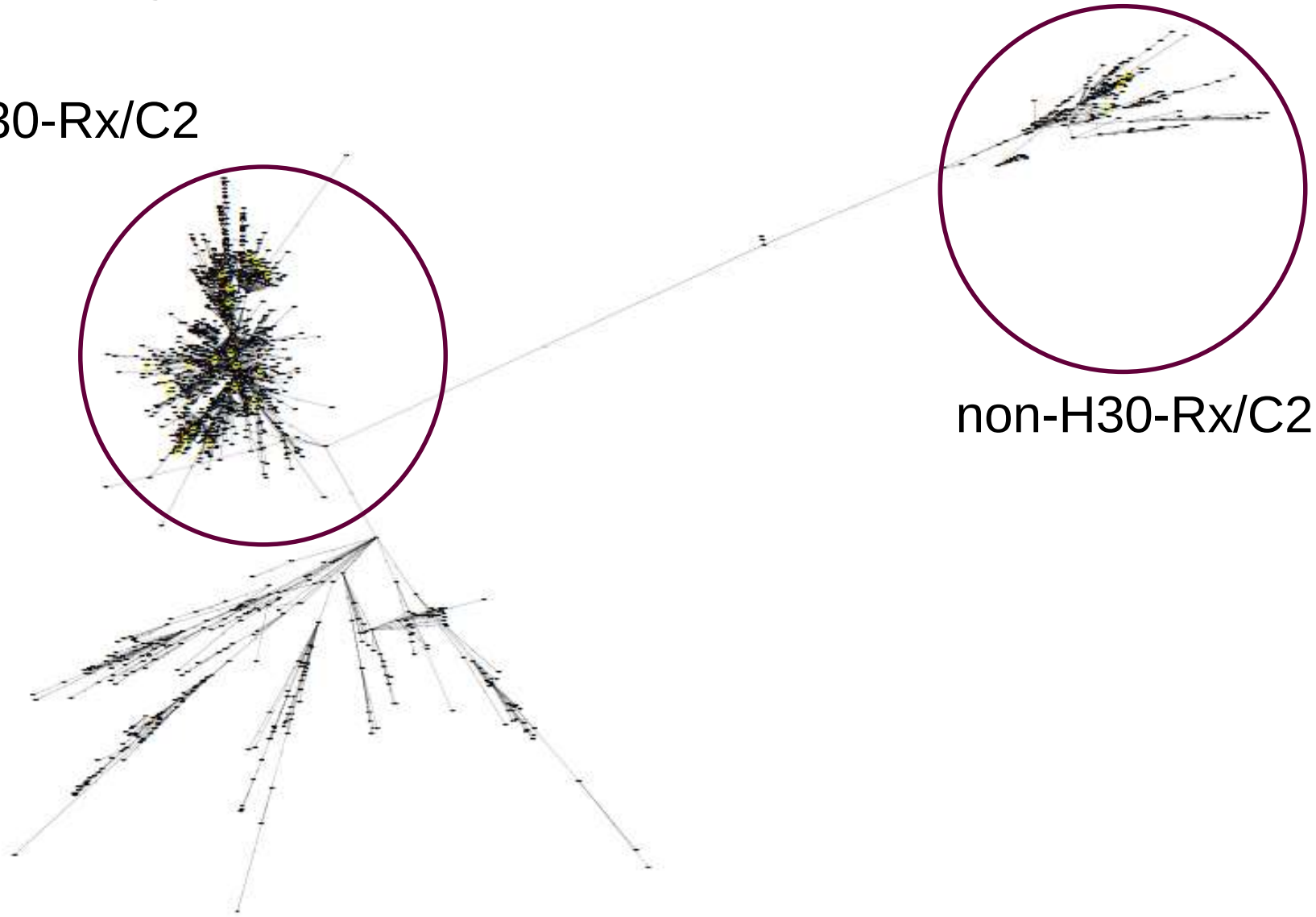
- When including Price et al strains shared genome drops from 87 to 50%
  - Logically leads to more SNV differences
  - Possibility that low coverage leads to low shared genome
- When making outbreak investigations with this method you will obtain high percentage shared genome
  - recombinant regions have to be managed

- Possible to detect closely related strains with high shared genome
  - affects SNV cut-off for defining outbreaks

- cgMLST not systematically explored for this purpose so far

ST131 cgMLST EnteroBase (2957 strains)

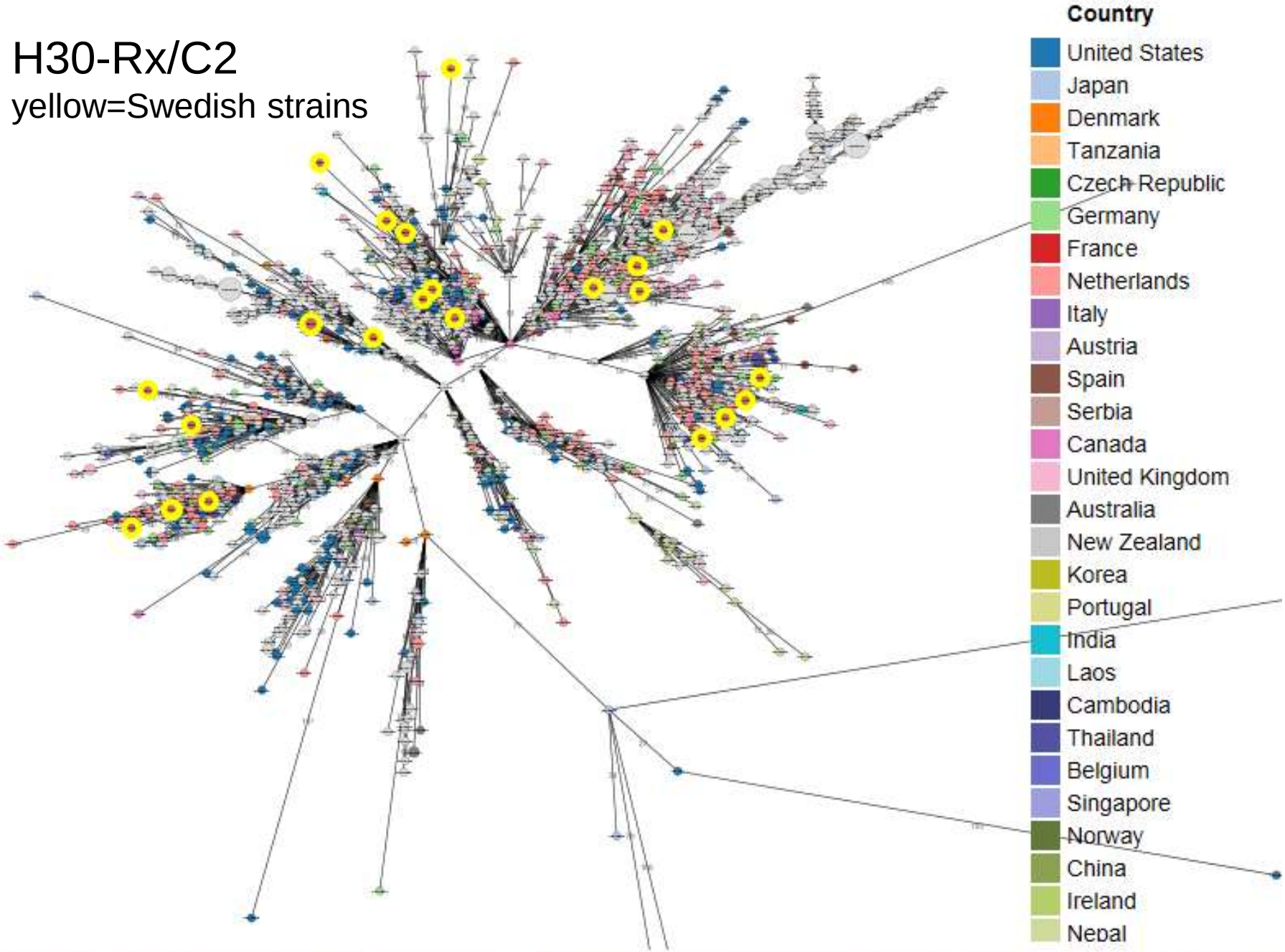
H30-Rx/C2



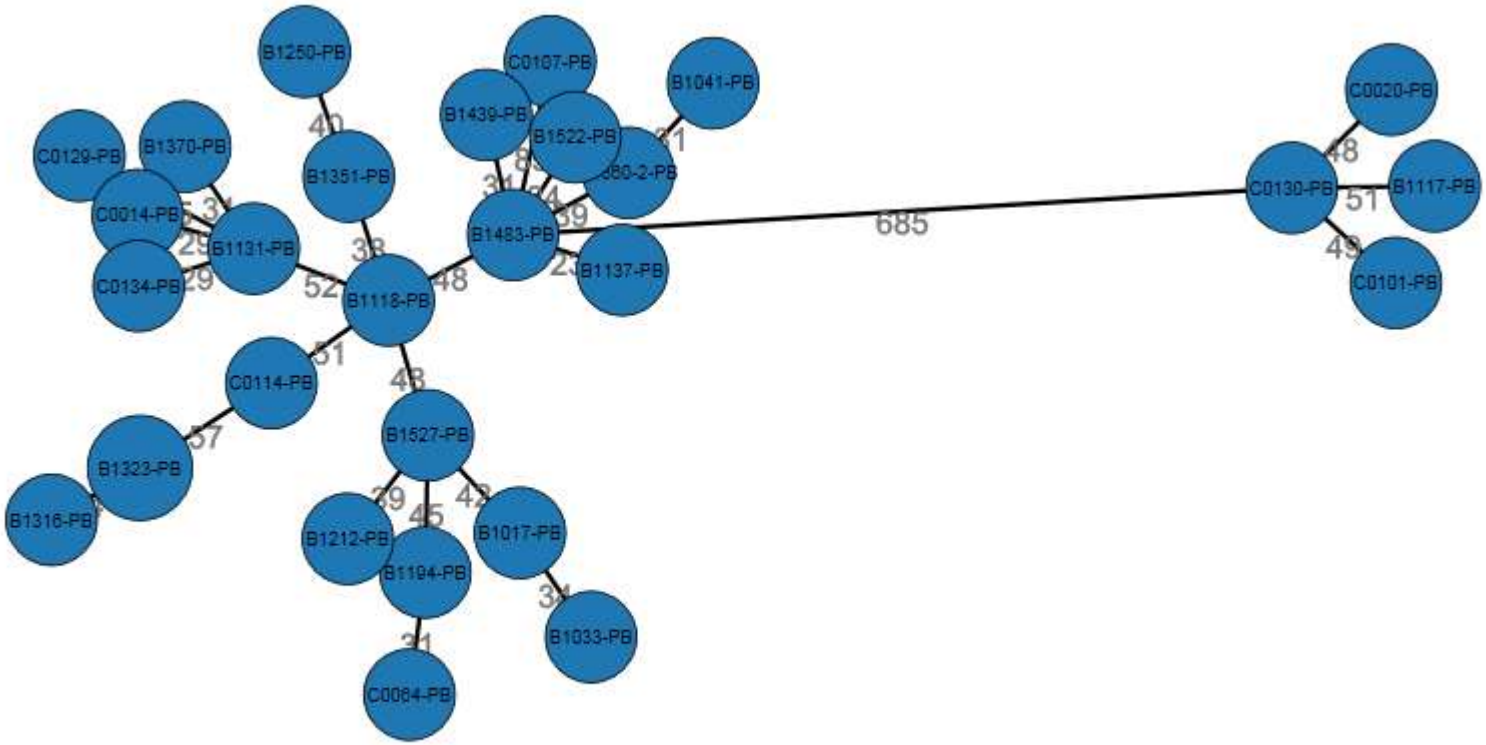
non-H30-Rx/C2

H30-Rx/C2

yellow=Swedish strains



cgMLST minimum spanning tree
EnteroBase, 2513 alleles (Swedish isolates)



Same phylogenetic pattern with cgMLST and whole genome SNV

Conclusions

- Difficult to relate to mutation rate when investigating outbreaks
  - Relevant mainly for non-recombinant parts

- How to relate to recombinant regions when calculating genomic drift?
  - Complex issue when dealing with high percentage shared genome
  - Open issue how to handle core vs accessory genome (McNally et al. PLoS Genetics 2016)

- You can identify if you have H30-Rx/C2 by cgMLST:
  - Uploading your raw Illumina reads in EnteroBase (free online software)
  - Use the workspace ”Do I have ST131 H30-Rx/C2?” (or make your own workspace from public strains)

Thank you!

Sofia Ny, Erik Alm, Linus Sandegren, Christian Giske



Special thanks to EnteroBase
Martin Sergeant
Nabil-Fareed Alikhan
https://enterobase.warwick.ac.uk/