Evolution of virulence genes in aEPEC
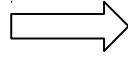
Artur Safin, Vera Emelianenko
Olga Bochkareva

Enteropathogenic *Escherichia coli* (EPEC)

EPEC

- important diarrheal pathogens of young children, don't produce toxins (ST and LT)
- have LEE and BFA
- adheres to the intestinal mucosa and causes a rearrangement of actin in the host cell
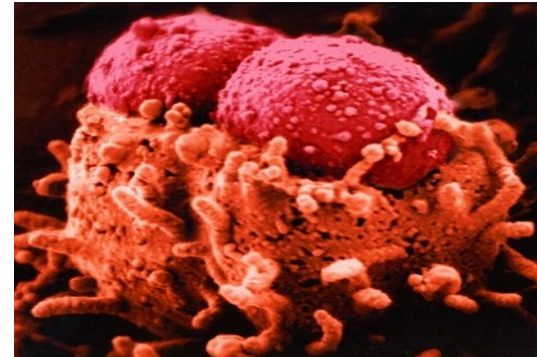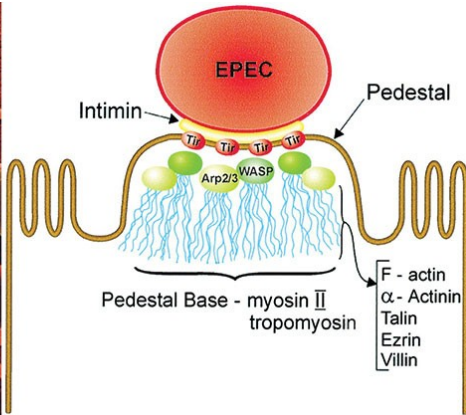
tEPEC:
- less common *E.coli* pathotype
- have LEE + BFA

aEPEC:
- don't have BFA
- more common *E.coli* pathotype
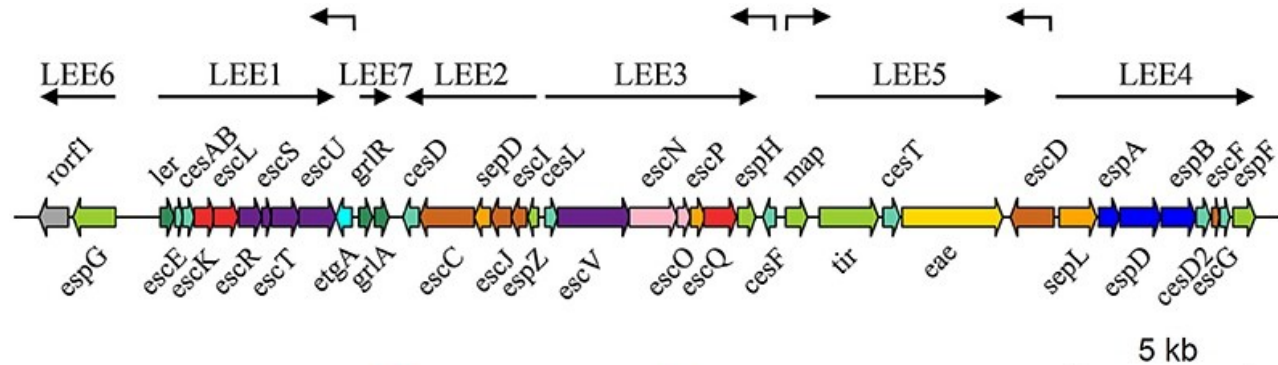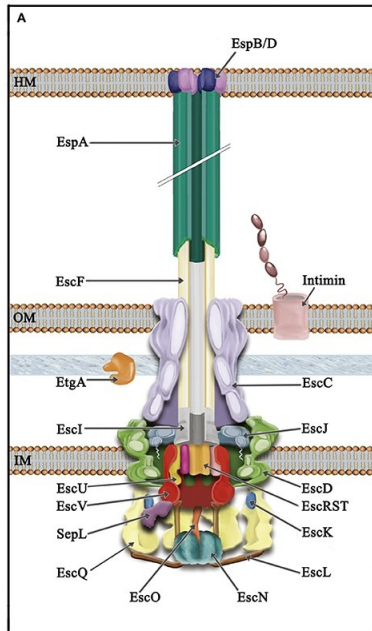
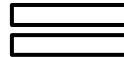# Main object of interest - genomic island LEE

components of delivery complex (**type III secretion system**)

protein **intimin** (incoded by the '*eae*' gene)
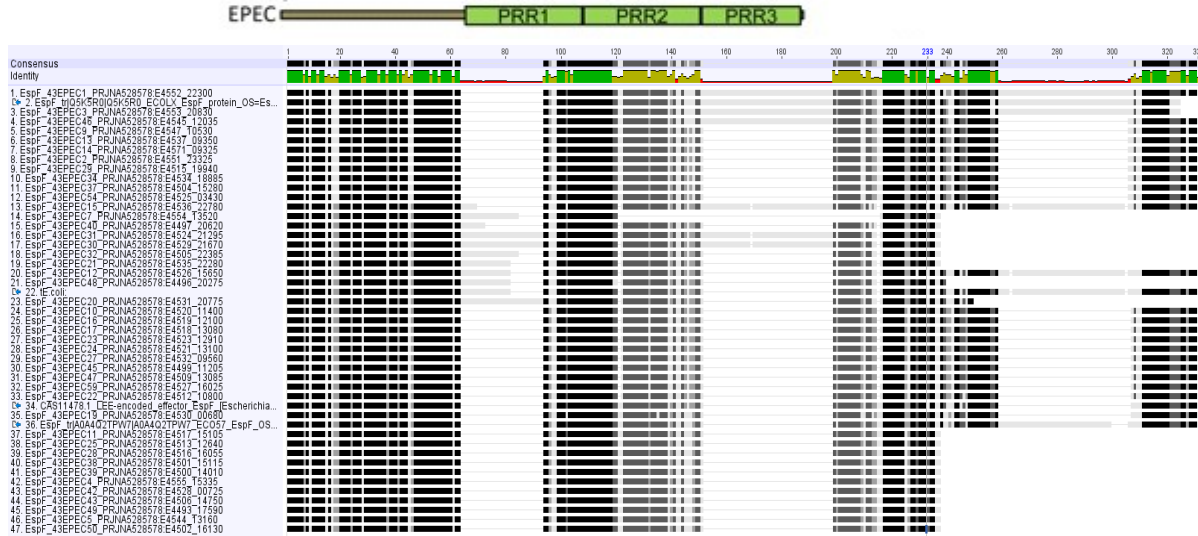
Effectors (**EspG**, **EspF**, **EspF(U)** etc.)

Data and methods
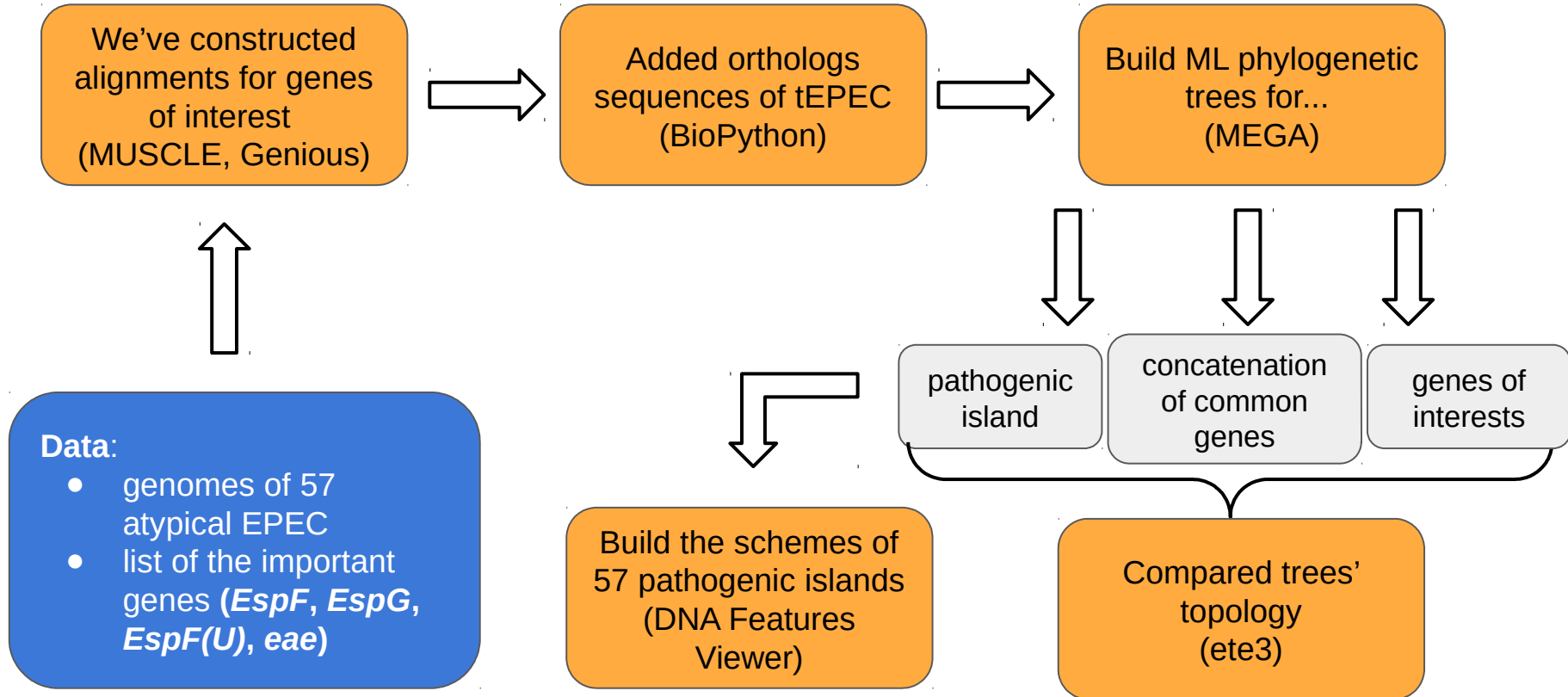
EspF alignment, domains' amplification
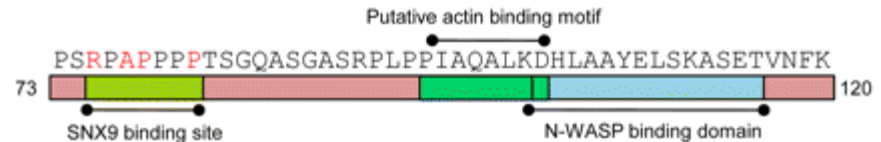

N-terminal region:
comprises the secretion signal, mitochondrial targeting signal (MTS), and nucleolar targeting domain

Proline rich repeat module:

SNX9 binding site (RxAPxxP).
binds SH3 domain of SNX9, inducing SNX9 oligomerization and increasing membrane deformation activity.

N-WASP binding site
interact with N-WASP to initiate actin fiber branching and assembly

N-terminal region

module duplication

module duplication

Insertion: MSHMSQMSR*n

Putative actin binding motif

PSRPAPPPPTSGQASGASRPLPPIAQALKDHLAAYELSKASETVNFK

73

120

SNX9 binding site

N-WASP binding domain

LEE PAI was acquired multiple times by HGT

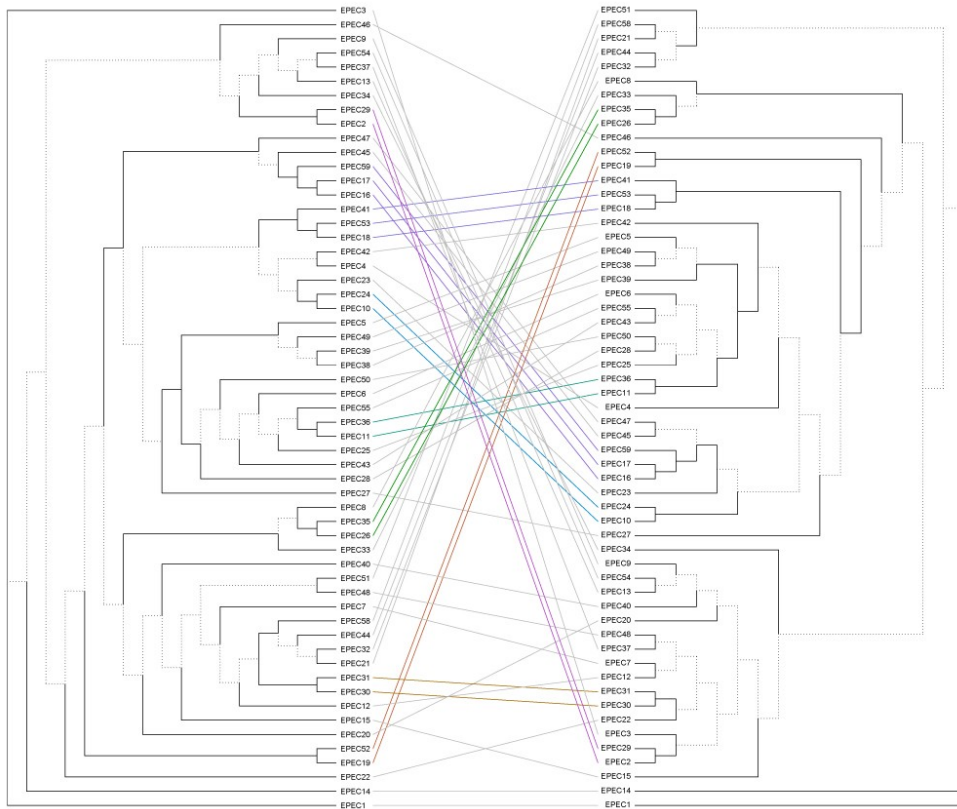**outgroup:**
E. albertii
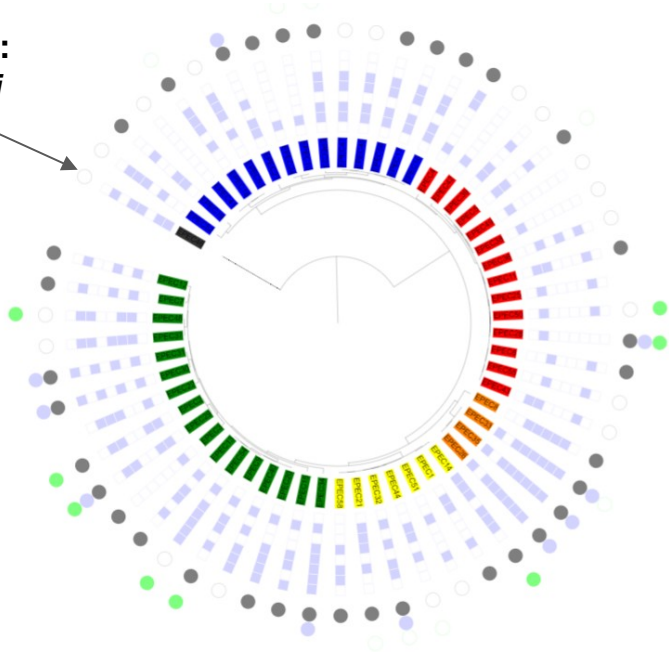
outgroup

Phylogenetic tree of aEPEC, based on alignment of 2719 common orthologous proteins. Different clades are highlighted with colors, squares indicate the presence of effector proteins, circles - phenotypes.
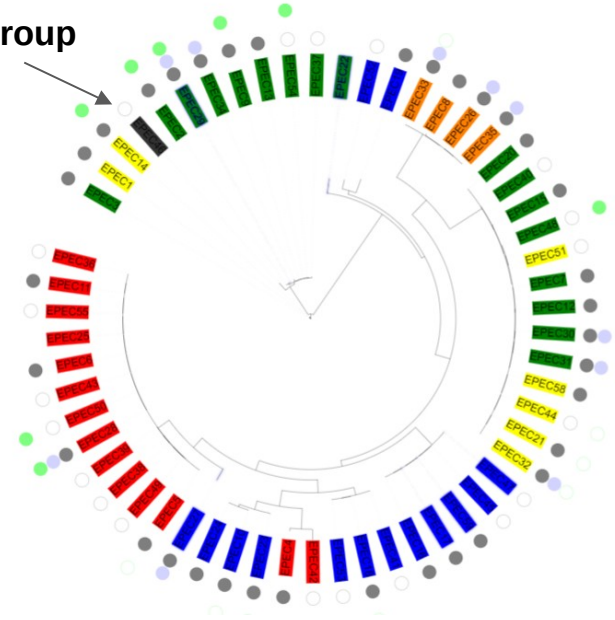
Phylogenetic tree of 28 LEE-encoded genes, based on aminoacid alignment. Different colors are mixed, which illustrated the history of horizontal transfer of this pathogenicity island.

The genes inside LEE might undergo recombination

Tanglegram, showing comparison between LEE tree (left) and the tree of all orthologs (right)
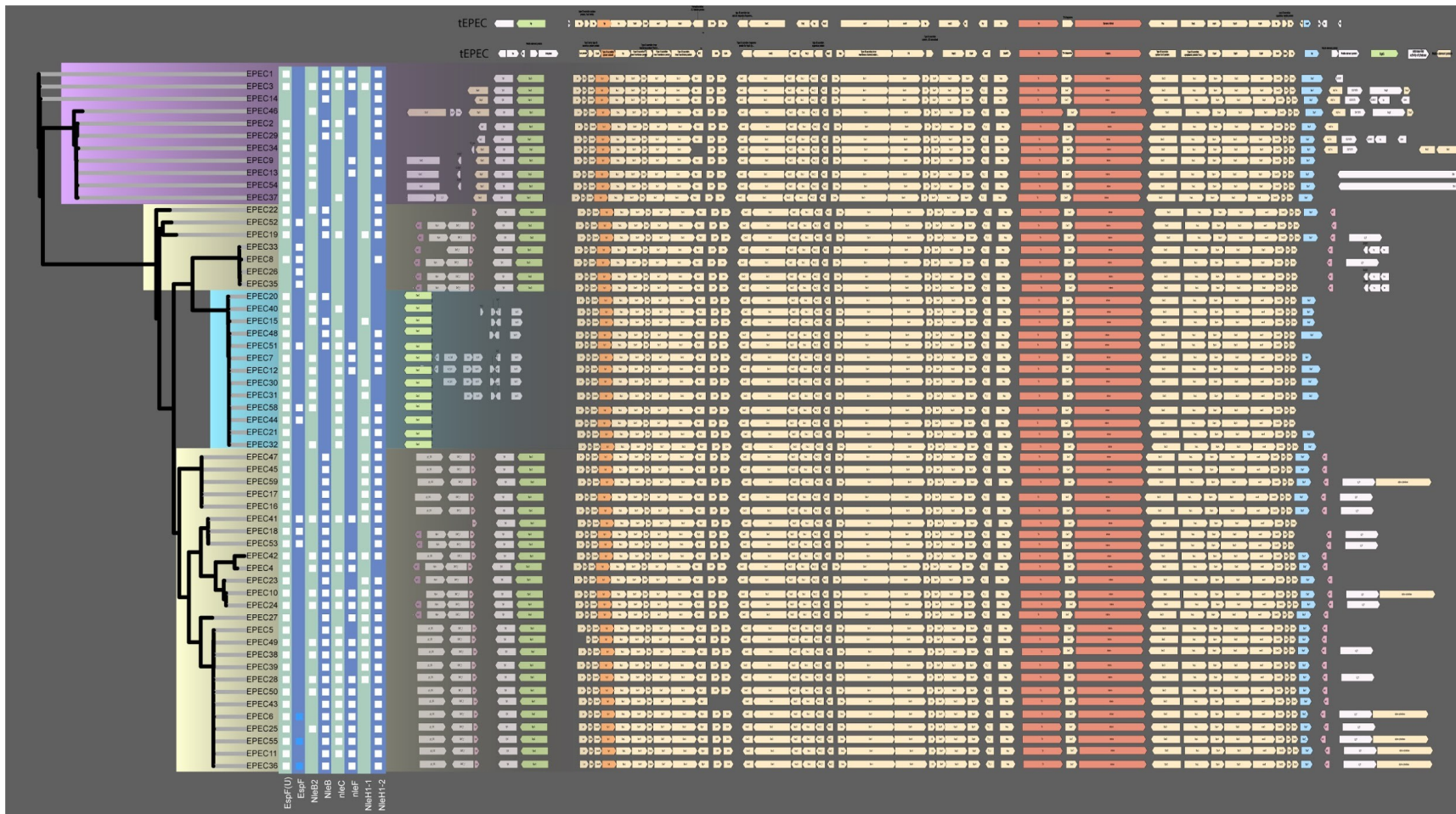Branch length and bootstrap - supports were not taken into account.

Tree comparison with python package ete3:
rf (Robinson–Foulds metric) measure of the distance between unrooted phylogenetic trees
rf = 0 - trees are identical

Tree	LEE	EspF	EspG	eae
LEE	rf 0.0 max 108.0	rf 74.0 max 80.0	rf 100.0 max 108.0	rf 74.0 max 108.0

Structure of the LEE island of a/tEPEC

The end