oriC environ speciates staphylococci

What can be found by comparative genomics of related bacterial species

2005.12.16 @Juntendo
Fumihiko Takeuchi
J. Bacteriology (2005) vol.187, p.7292

Background: We compare the genomes of

- *Staphylococcus aureus*
  - Major pathogen for hospital- and community-acquired infection
  - Notorious for its virulence and antibiotic resistance as in Methicillin-resistant *S. aureus* (MRSA)
- *Staphylococcus epidermidis*
  - Colonizes the skin of almost all humans
  - Benign but causes foreign-body infection
- *Staphylococcus haemolyticus*
  - Also among the normal skin flora
  - Remarkable for its historically early acquisition of antibiotic resistance

A fundamental concept: *homologs*

- open reading frames (*orfs*)
  - Genes encoding proteins

S. aureus							*S. epidermidis*		*S. haemolyticus*
COL	MRSA252	MSSA476	Mu50	N315	NCTC8325	MW2	ATCC12228	RP62a	JCSC1435
2615	2656	2579	2714	2594	2810 (draft)	2632	2419	2494	2678

- *homologs*
  - *orfs* grouped by their nucleotide sequence similarity
  - Definition: two *orfs* with quality ratio ≥70% by alignment belong to the same *homolog*
  - e.g. the *homolog* of intercellular adhesion A (*icaA*) genes comprises
    - COL, #2565, icaA, intercellular adhesion protein A
    - MRSA252, #2608, icaA, glucosaminyltransferase
    - MSSA476, #2533, -, glucosaminyltransferase
    - Mu50, #2666, icaA, intercellular adhesion protein A
    - N315, #2458, icaA, intercellular adhesion protein A
    - NCTC8325, #2761
    - MW2, #2586, icaA, intercellular adhesion protein A
    - RP62A, #2243, icaA, intercellular adhesion protein A
  - Two *orfs* that are homologous/orthologous/paralogous should belong to the same *homolog*

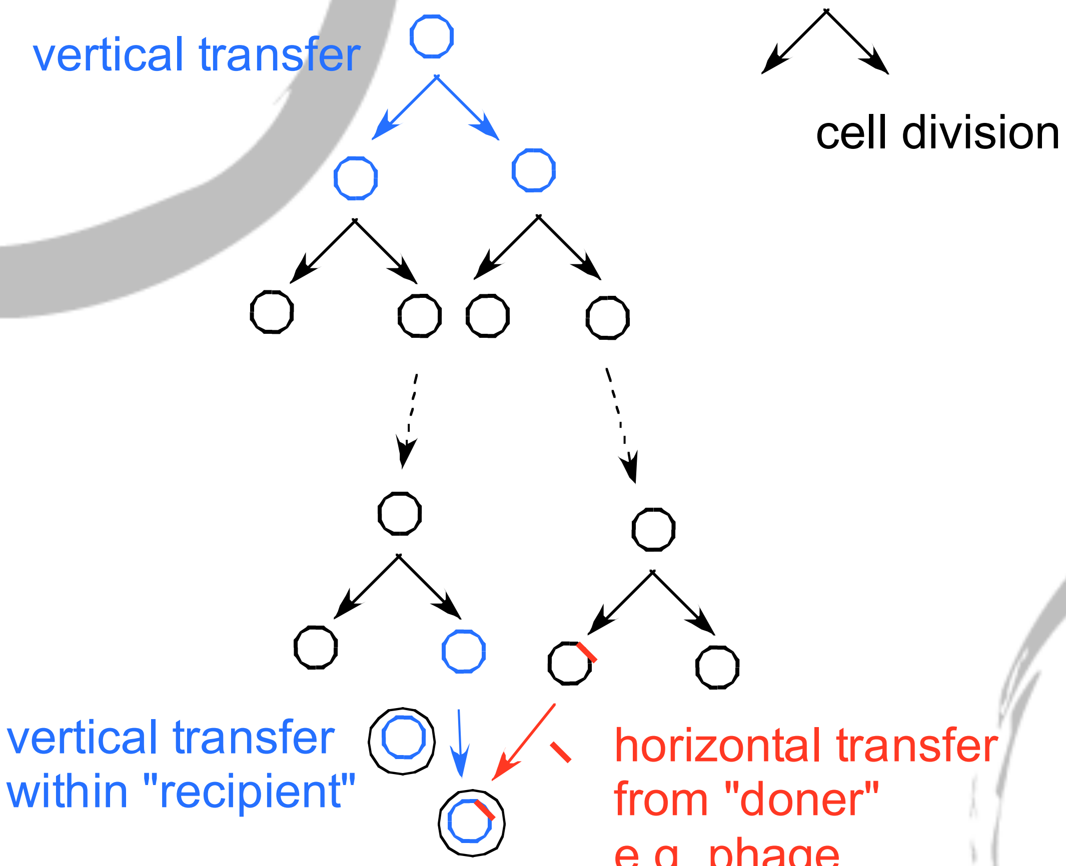
Two basic profiles of *homologs*

- Phylogenetic profile
  - Which species/strains share a *homolog*
- Location profile
  - Where on chromosome are the *orfs* (of a *homolog*) located
- e.g. *icaA homolog*

	S. aureus							*S. epidermidis*		*S. haemolyticus*
	COL	MRSA252	MSSA476	Mu50	N315	NCTC8325	MW2	ATCC12228	RP62a	JCSC1435
Phylogenetic profile	+	+	+	+	+	+	+	-	+	-
	Common to *S. aureus*							Partially in *S. epidermidis*		Not in *S. haemolyticus*
Location profile	#2565 2763320..2764558	#2608	#2533	#2666	#2458	#2761	#2586		#2243	

- Viewpoint from comparative genomics:
  association of these profiles might tell us something…

Vertical and horizontal transfer of DNA



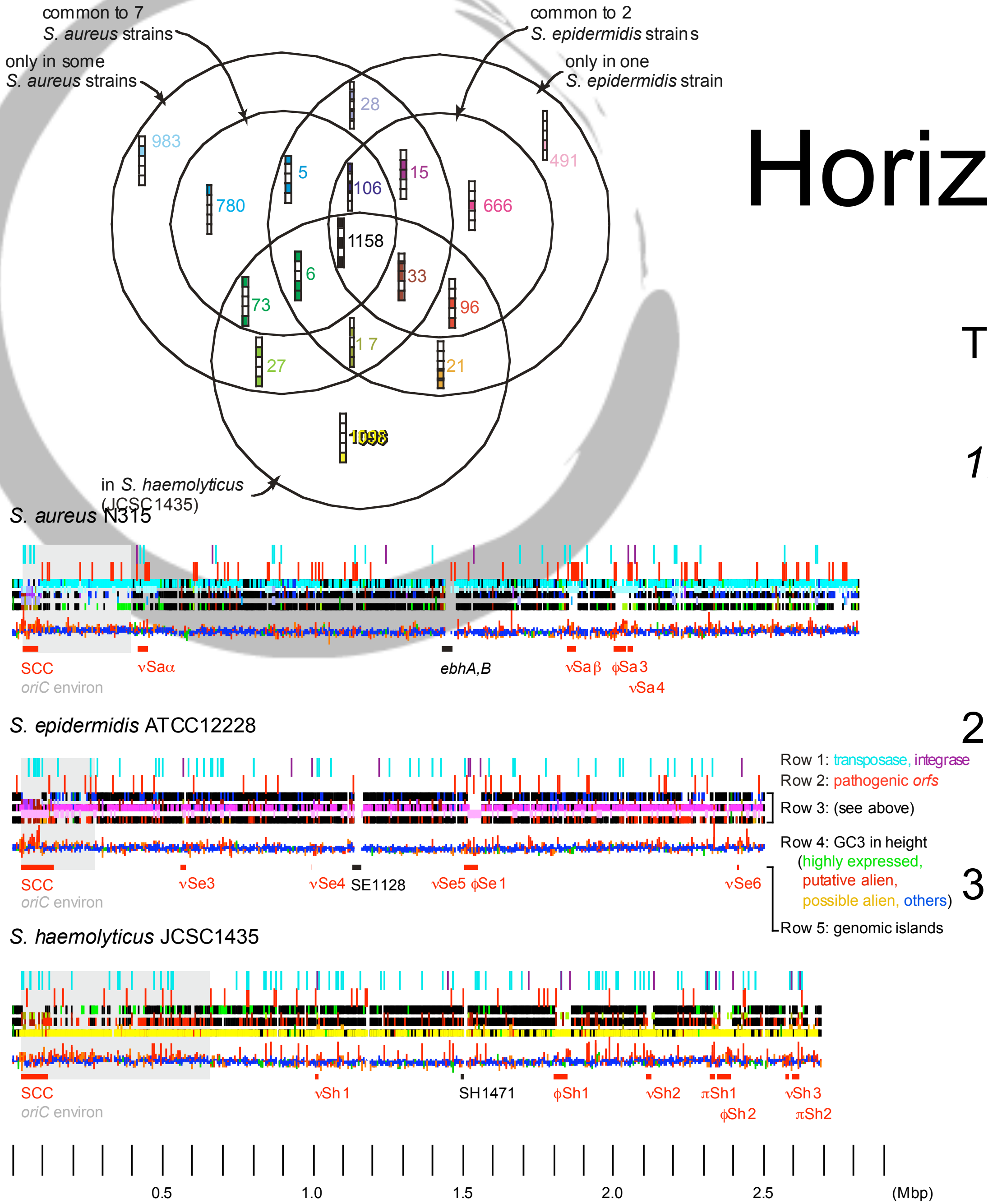
vertical transfer

cell division

vertical transfer
within "recipient"

horizontal transfer
from "doner"
e.g. phage

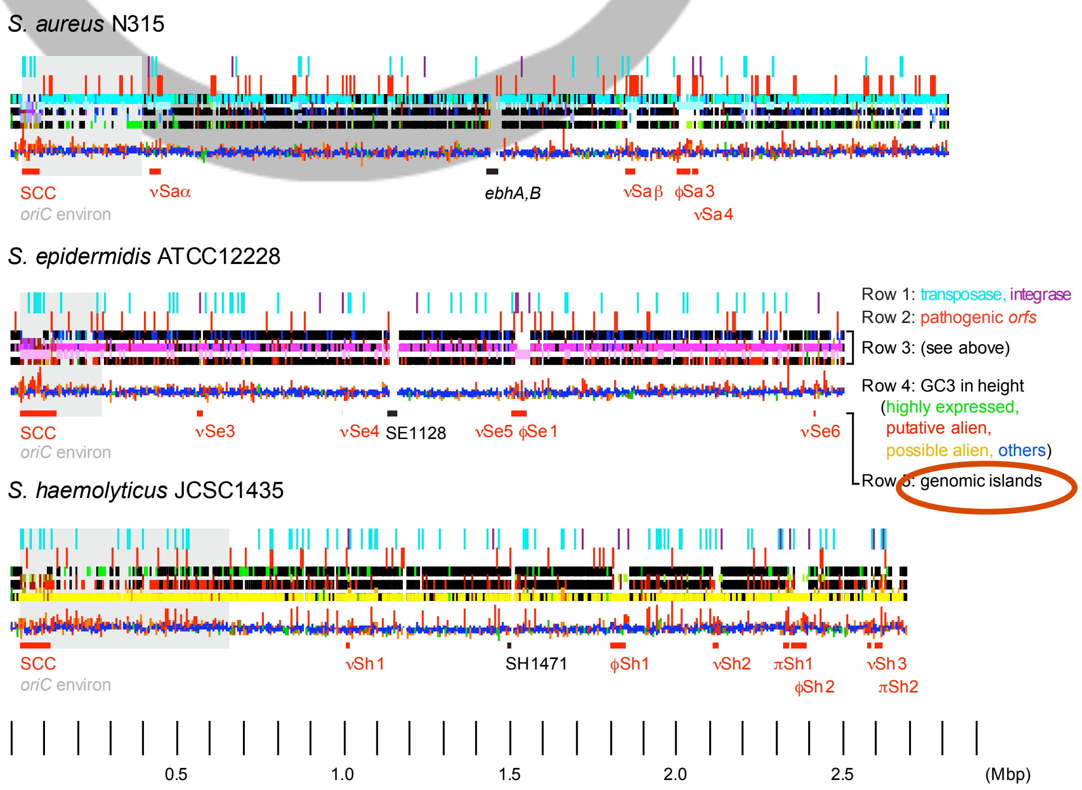
Horizontal transfer



The phylogenetic profile & location profile of *homologs* identify horizontal transfer.

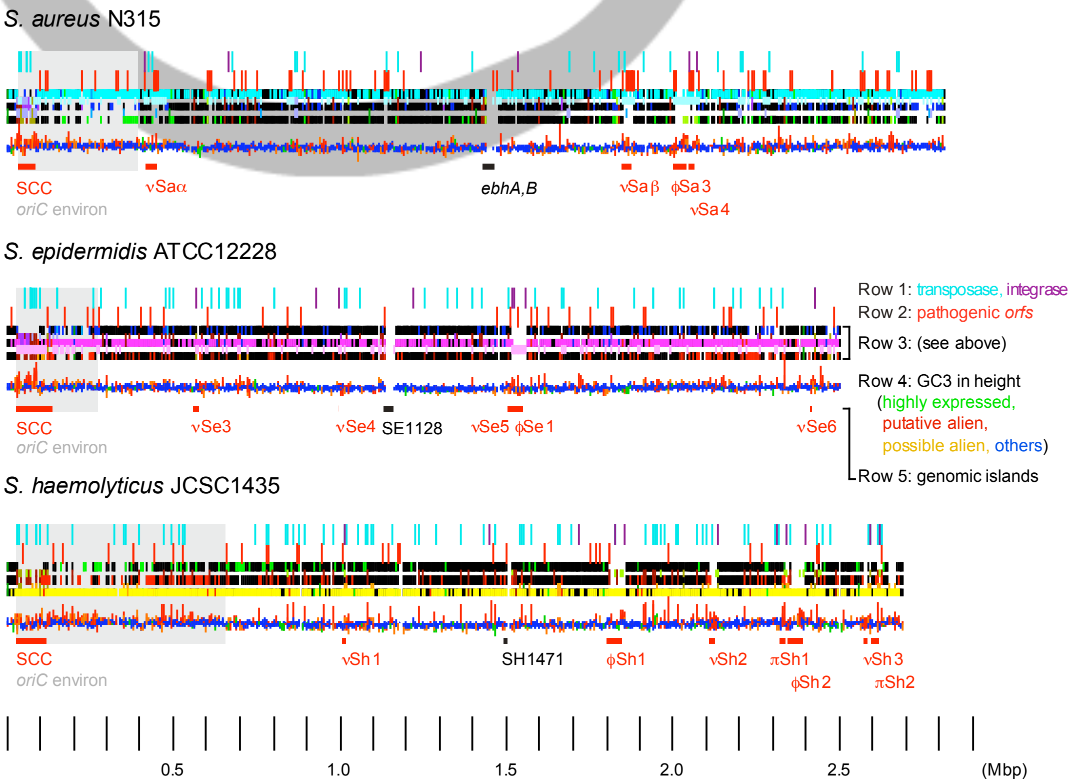
1. *homologs* are classified and color-coded according to their phylogenetic profile; *homologs* common to three species are represented by black segmented bars
2. Each *orf* of a genome is colored according to the phylogenetic profile (of its belonging *homolog*)
3. Chromosomal regions lacking the common *homologs* are suggested to be introduced by horizontal transfer
    - Genomic islands (prophages, integrated plasmids, SCC, etc.)
    - *oriC* environ

Genomic islands



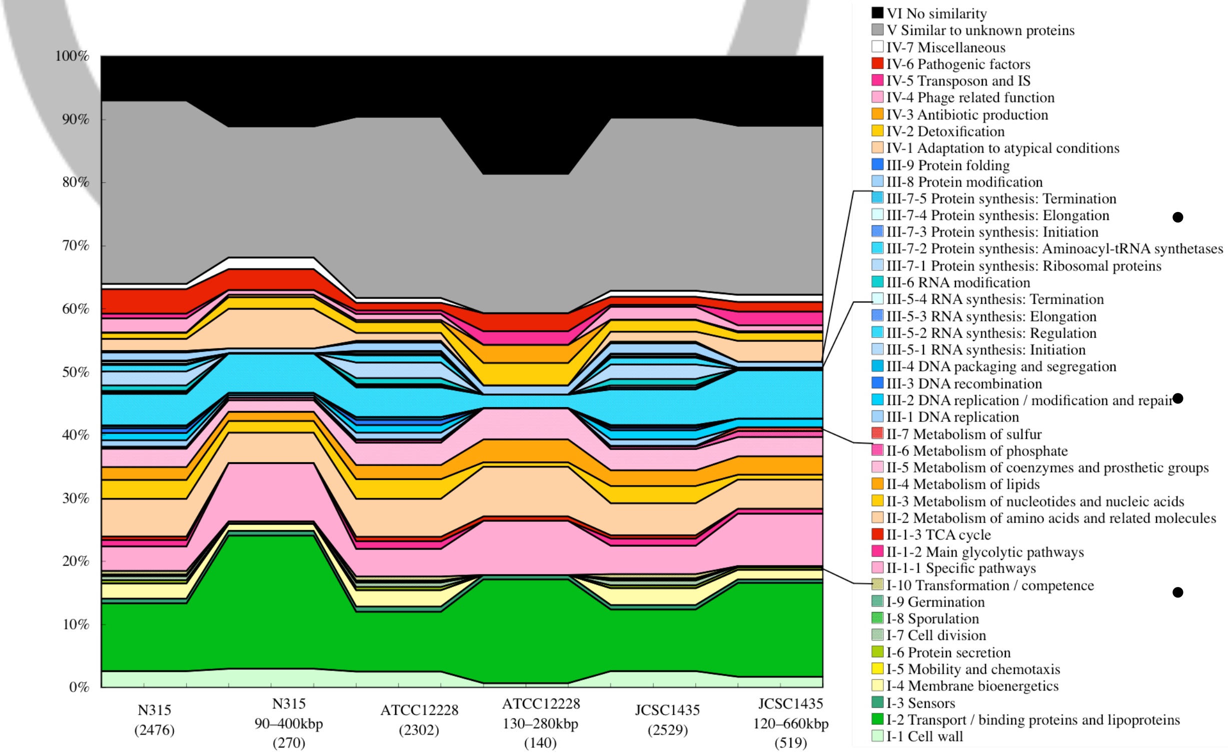
- Chromosomal regions (>10 kbp) putatively introduced by horizontal transfer
- Anomalous codon usage, biased GC3, and existence of integrase and transposase support their alien origin
- Differ between strains (even within a species)

oriC environ



- Located in the downstream of the origin of DNA replication (*oriC*), and defined as stretches of 20-kbp windows with <45% common *homologs*
  - 34,153–400,000 bp in *S. aureus* N315
  - 32,492–280,000 bp in *S. epidermidis* ATCC12228
  - 29,763–660,000 bp in *S. haemolyticus* JCSC1435
- Staphylococcus Cassette Chromosome (SCC) occupies the left part
- The remaining right part is specific to species (and conserved among strains within a species)
- Codon usage and GC3 is similar to the total chromosome
- The rest of the talk focuses on *oriC* environ

oriC environ (other than SCC): functional profile



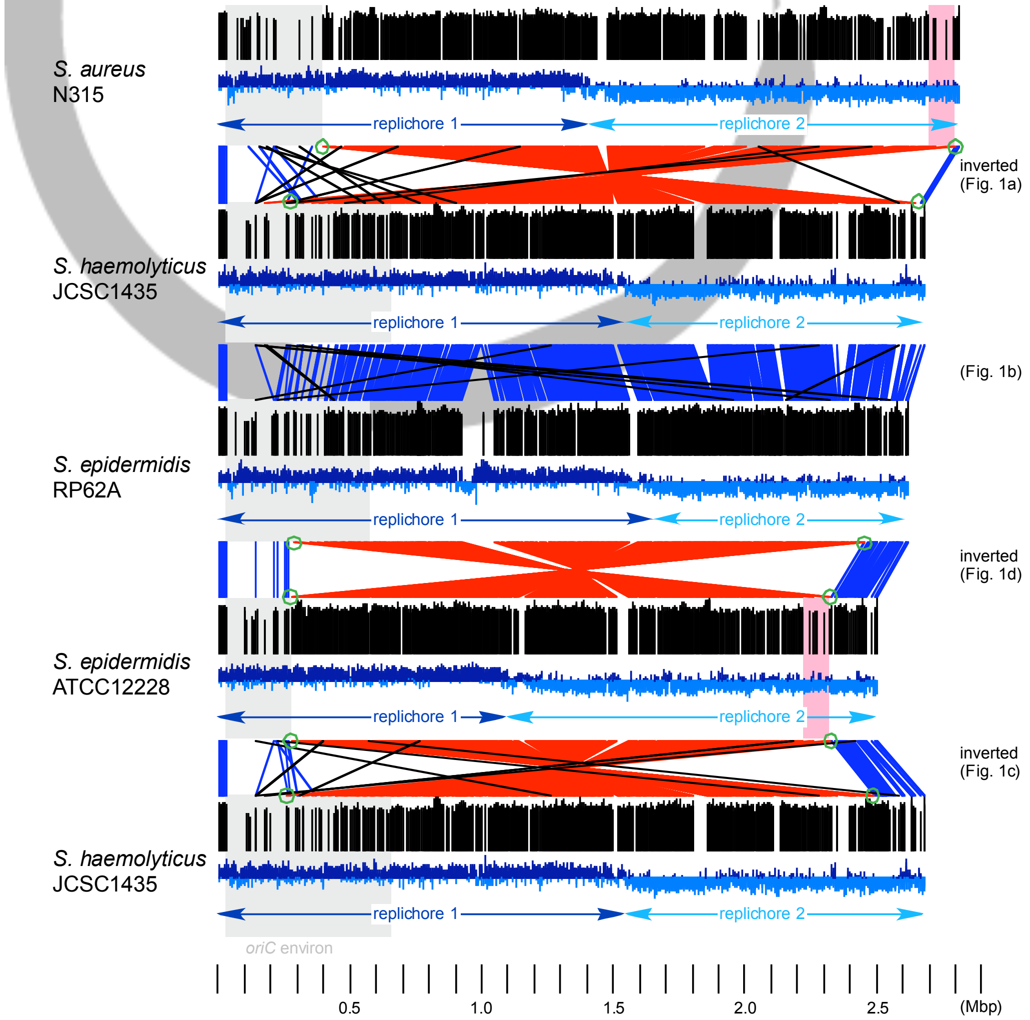
- In *oriC* environ (other than SCC) of any species, higher proportion was observed for
  - Transport/binding proteins and lipoproteins (I-2)
  - Specific pathways (II-1-1)
- In *S. aureus* N315 90–400 kbp
  - protein A (*spa*), coagulase (*coa*), capsular polyscaccharide biosynthesis operon (*cap5[8]A–P*)
- In *S. epidermidis* ATCC12228 130–280 kbp
  - Detoxification (IV-2)
  - Antibiotic production (IV-3) such as iturin A and surfactin
- In *S. haemolyticus* JSCS1435 120–660 kbp
  - Capsule operon (*capA–M*) belonging to adaptation to atypical condition (IV-1)
  - Mannitol-specific phosphotransferase system (PTS)

oriC environ (other than SCC): BLASTP top-hit species profile

- BLASTP top-hit species of and *orf* suggests the previous (or next) hosting species of the *orf*
- *oriC* environ (other than SCC) had fewer top-hit in the staphylococci's most related *Bacillus* species, implying their acquisition by horizontal transfer.

%	*S. aureus* N315	*S. aureus* N315 87–400 kbp	*S. epidermidis* ATCC12228	*S. epidermidis* ATCC12228 134–280 kbp	*S. haemolyticus* JCSC1435	*S. haemolyticus* JCSC1435 120–520 kbp
Bacillus	46	26	47	37	44	25
Listeria	13	6	11	10	11	7
Oceanobacillus	11	14	11	10	11	15
Streptococcus	5	14	5	9	6	12

Chromosomal rearrangements



The phylogenetic profile and location profile of *homologs* also indicate chromosomal rearrangements during vertical transfer

- For the common *homolog* that has only one *orf* in each genome, we connected the *orfs* by lines. This indicates the history of chromosomal rearrangements (inversion in red)

- Probable ancestor type was *S. haemolyticus* and *S. epidermidis* RP62a
- Their *ica* operons were in *oriC* environ

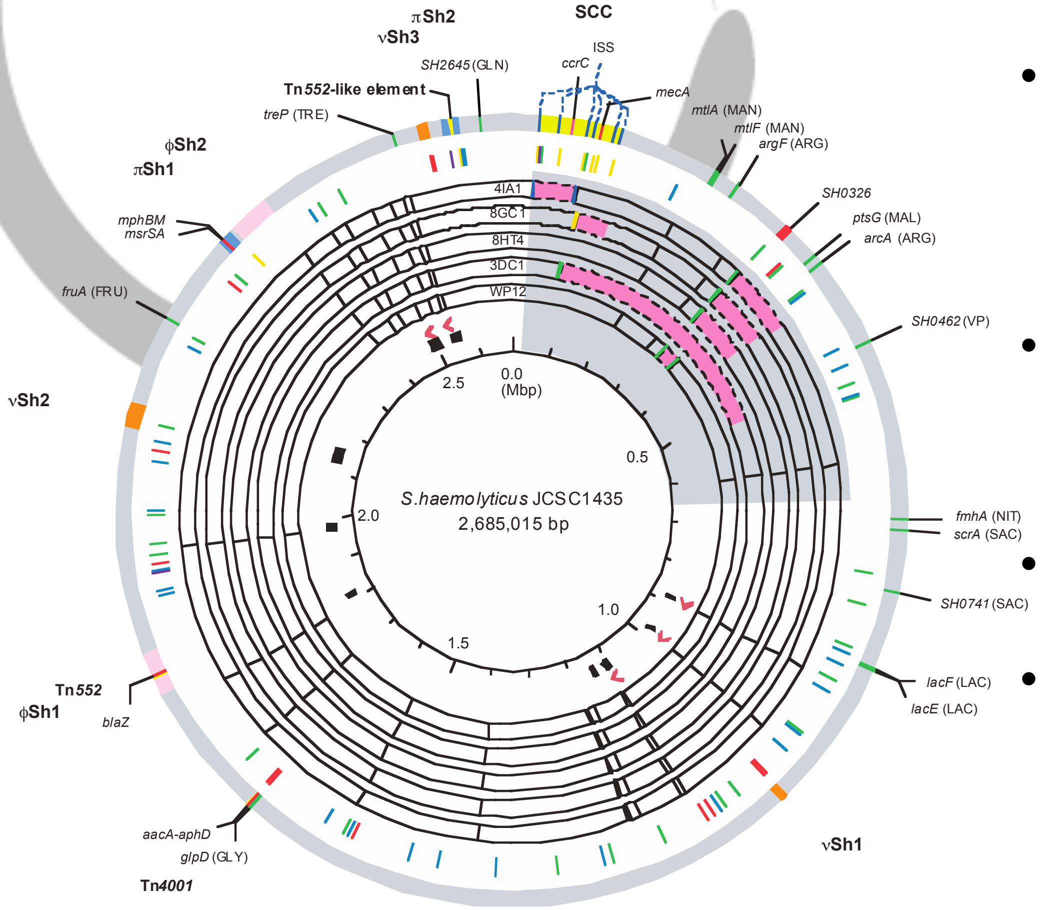
- *S. aureus* and *S. epidermidis* ATCC12228 were inverted around *oriC*, and remnats of *oriC* environ were found transolocated (pink highlight).
- *ica* operon of *S. aureus* was in the translocated region (2,550 kbp in N315).
- Loci flanking *ica* in *S. epidermidis* RP62a was found in 280 and 2,330 kbp of *S. epidermidis* ATCC12228 (= breakpoints of rearrangements) and the *ica* operon was missing in ATCC12228.

Summary

Phylogenetic profile and location profile of *homologs* could extract

- Chromosomal regions transferred horizontally—genomic islands and *oriC* environ
  - Genomic islands were introduced in specific strains
  - *oriC* environ included SCC (varies by strains) in its left part, and a species-specific region in the right part
- Chromosomal rearrangements during vertical transfer
  - The inverted genomes had a remnant of *oriC* environ translocated in the upstream of *oriC*

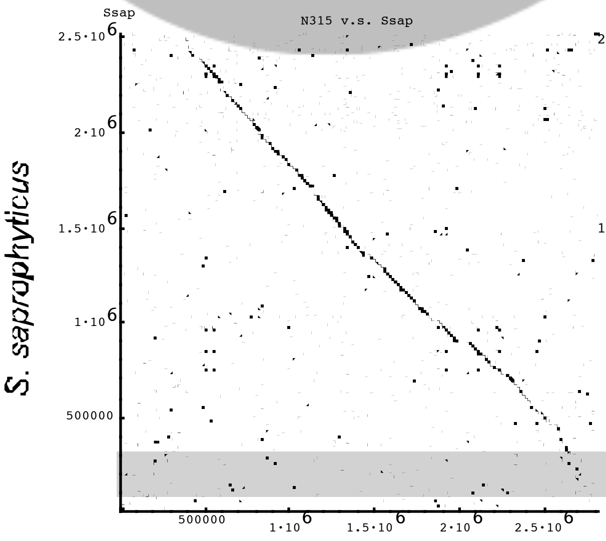
Epilogue 1: *in vitro* chromosomal rearrangements in *oriC* environ [Watanabe]



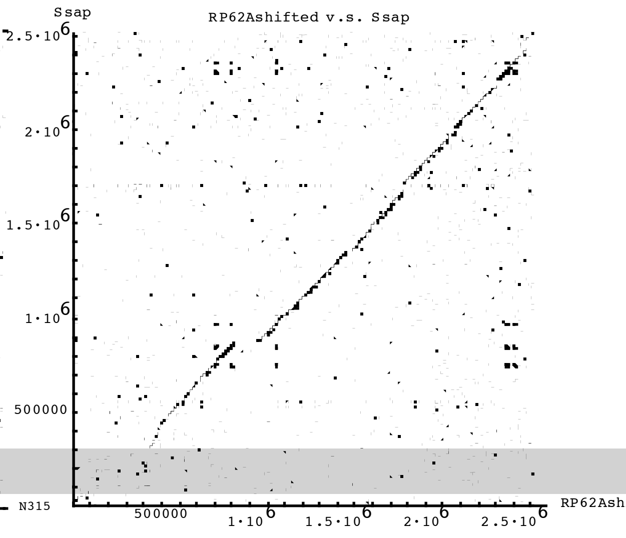
- Method: prolonged (11 or 18 days) culture of *S. haemolyiticus* JCSC1435, followed by isolation of susceptible mutants
- Deletions (pink boxes) exclusively in *oriC* environ mediated by the numerous IS's
- Changes in biochemical properties
- --> *oriC* environ is not essential for cell viability but biochemically characterizes the species

Epilogue 2: *oriC* environ in *S. saprophyticus*

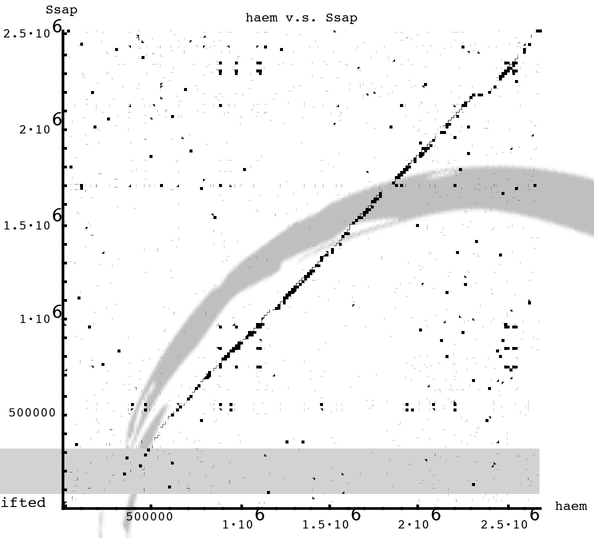
- Does *oriC* environ exist in the recently sequenced *Staphylococcus saprophyticus* genome?
- --> Yes



S. aureus N315



S. epidermidis RP62A



S. haemolyticus JCSC1435

Epilogue 3: a region similar to *oriC* environ in *Bacillus*

- Does *oriC* environ exist in other species?
- --> Species-specific region around *oriC* exists in *Bacillus* species. However, *orfX* (where the SCC is integrated in staphylococci) is not located in this region (4,134,123..4,133,647 in *B. subtilis*)