

Journal Club: *Salmonella*

Course 27104

The Scientific Communication of Comparative Genomics

Thursday 15 September 2011

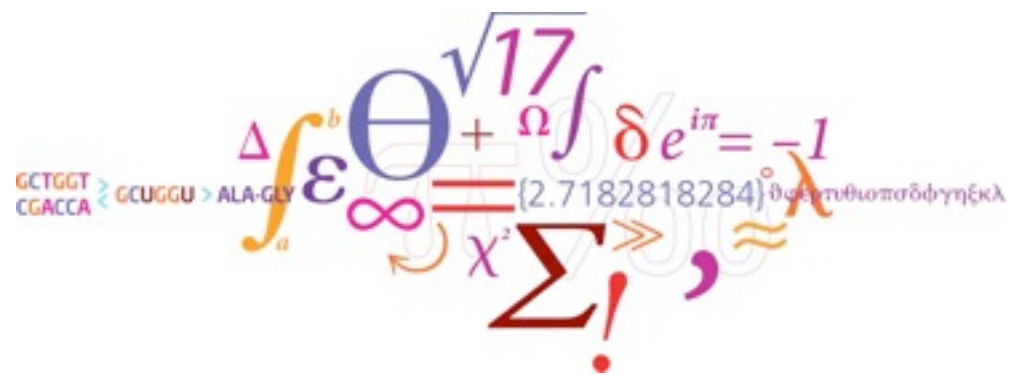
Julie Rank, S071283

Sebastian Wingard Thrane, S082424

Sandra Wingard Thrane, S082403

Tanni Leth Christensen, S072807

Martin Frederik Laursen, S072490





The *Salmonella enterica* Pan-genome

Annika Jacobsen · Rene S. Hendriksen ·
Frank M. Aarestrup · David W. Ussery · Carsten Friis

Received: 1 January 2011 / Accepted: 8 May 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract *Salmonella enterica* is divided into four subspecies containing a large number of different serovars, several of which are important zoonotic pathogens and some show a high degree of host specificity or host preference. We compare 45 sequenced *S. enterica* genomes that are publicly available (22 complete and 23 draft genome sequences). Of these, 35 were found to be of sufficiently good quality to allow a detailed analysis, along with two *Escherichia coli* strains (K-12 substr. DH10B and the avian pathogenic *E. coli* (APEC O1) strain). All genomes were subjected to standardized gene finding, and the core and pan-genome of *Salmonella* were estimated to be around 2,800 and 10,000 gene families, respectively. The constructed pan-genomic dendrograms suggest that gene

content is often, but not uniformly correlated to serotype. Any given *Salmonella* strain has a large stable core, whilst there is an abundance of accessory genes, including the *Salmonella* pathogenicity islands (SPIs), transposable elements, phages, and plasmid DNA. We visualize conservation in the genomes in relation to chromosomal location and DNA structural features and find that variation in gene content is localized in a selection of variable genomic regions or islands. These include the SPIs but also encompass phage insertion sites and transposable elements. The islands were typically well conserved in several, but not all, isolates—a difference which may have implications in, e.g., host specificity.

Salmonella enterica

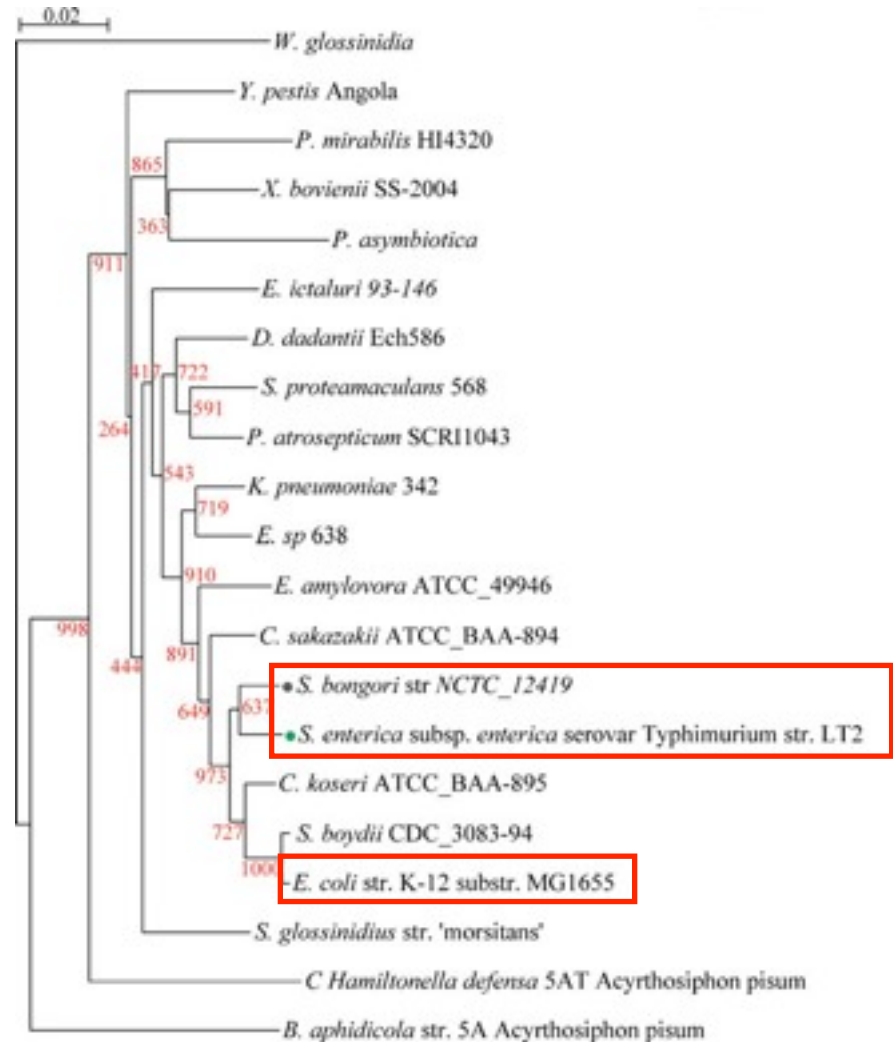
- 2 species: *S. enterica* and *S. bongori*
S. enterica: subspecies - serovars
- Intracellular pathogen
 - Typhoid fever
 - Foodborne illness
- Virulence and host specificity:
Salmonella pathogenicity islands
(SPI's)



Table 1 General properties of the *S. enterica* and *E. coli* genomes used in this study

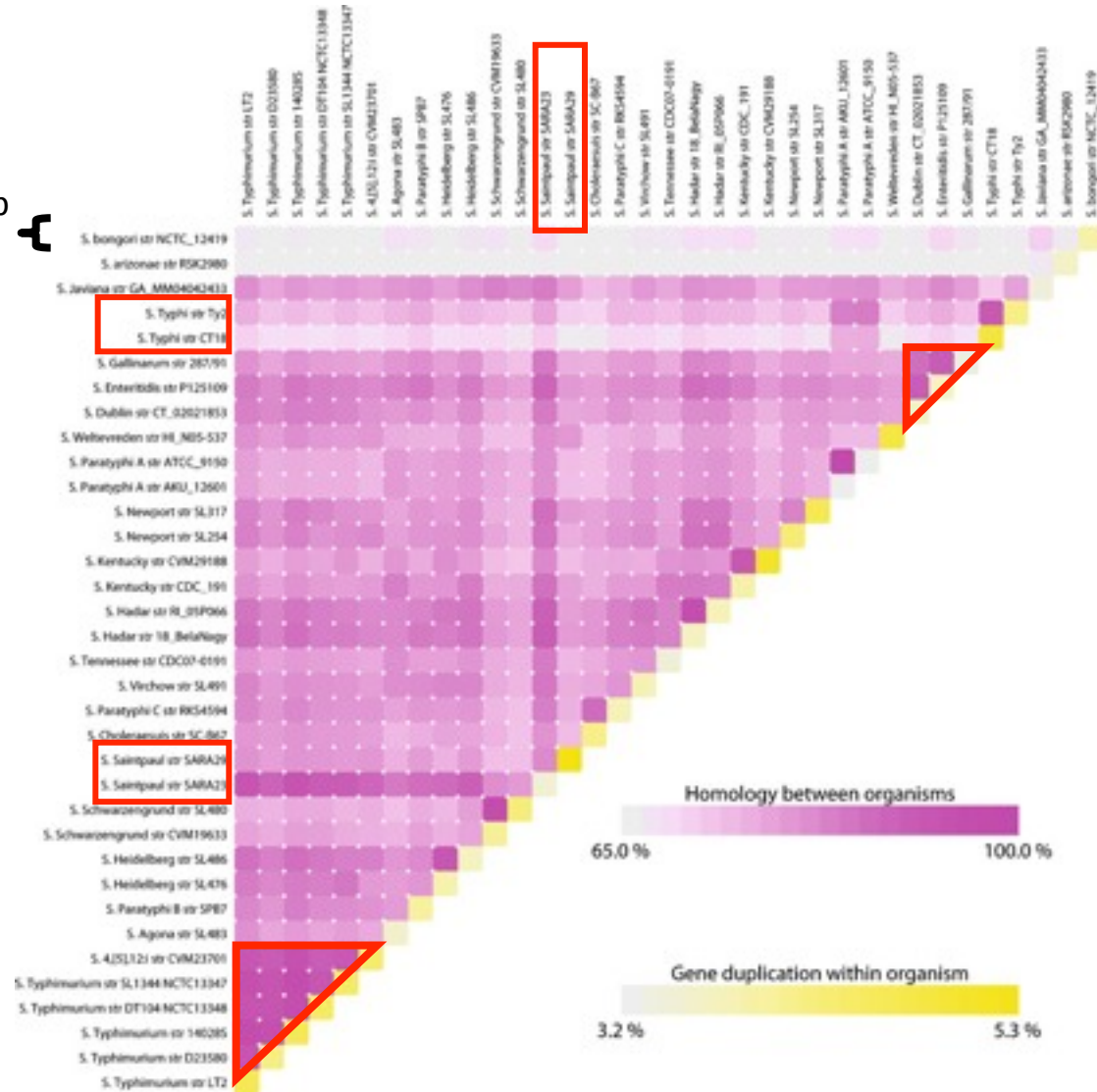
Organism* (publication reference)	Genome Size	Contigs	Quality Score	Accession	PID	Genes	Specificity	Sero group (O antigen)
<i>S. Panteyphi</i> A str. ARJ_12601 [6]	4.58 MB	1	1	FM200053	30943	4,351	Human-restricted	O:2
<i>S. Panteyphi</i> A str. ATCC 9150 [18]	4.59 MB	1	1	CP000026	13086	4,348	Human-restricted	O:2
<i>S. 4,[5],12:i:-</i> str. CVM23701 [45]	4.90 MB ^b	113	3	ABA000000000	19465	4,694	Ubiquitous	O:4
<i>S. Agona</i> str. SL483	4.84 MB	2	1	CP001138	20063	4,508	Ubiquitous	O:4
<i>S. Heidelberg</i> str. SL476	4.89 MB	3	1	CP001120	20045	4,680	Ubiquitous	O:4
<i>S. Heidelberg</i> str. SL486	4.75 MB ^b	48	3	ABE100000000	20065	4,432	Ubiquitous	O:4
<i>S. Panteyphi</i> B str. SPB7	4.86 MB	1	1	CP000886	27803	4,555	Ubiquitous	O:4
<i>S. Saintpaul</i> str. SARA23	4.72 MB ^b	2	1	ABAM000000000	19461	4,350	Ubiquitous	O:4
<i>S. Saintpaul</i> str. SARA29	4.93 MB ^b	182	5	ABAN000000000	19463	4,757	Ubiquitous	O:4
<i>S. Schwarzengrund</i> str. CVM19633	4.71 MB	3	1	CP001127	19459	4,551	Ubiquitous	O:4
<i>S. Schwarzengrund</i> str. SL480	4.76 MB ^b	67	3	ABE100000000	20071	4,547	Ubiquitous	O:4
<i>S. Typhimurium</i> str. 140285 [72]	4.76 MB	2	1	CP001363	33067	4,653	Ubiquitous	O:4
<i>S. Typhimurium</i> str. D23580 [13]	4.88 MB	1	1	FN424405	40625	4,804	Ubiquitous	O:4
<i>S. Typhimurium</i> str. DT104	4.93 MB	2	1	–	–	4,752	Ubiquitous	O:4
<i>S. Typhimurium</i> str. LT2 [4]	4.86 MB	2	1	AB006468	241	4,635	Ubiquitous	O:4
<i>S. Typhimurium</i> str. SL1344	4.88 MB	4	1	–	–	4,774	Ubiquitous	O:4
<i>S. Choleraesuis</i> str. SC-867 [69]	4.76 MB	3	1	AB017220	9618	4,792	Porcine-adapted	O:7
<i>S. Panteyphi</i> C str. RKS4994 [52]	4.83 MB	2	1	CP000857	20993	4,690	Ubiquitous	O:7
<i>S. Tennessee</i> str. CDC07-0191	4.79 MB ^b	94	3	ACHF000000000	30831	4,546	Ubiquitous	O:7
<i>S. Virchow</i> str. SL491	4.88 MB ^b	5	2	ABFH000000000	20995	4,596	Ubiquitous	O:7
<i>S. Hadar</i> str. R1_05P066	4.79 MB ^b	50	3	ABFG000000000	20993	4,487	Ubiquitous	O:8
<i>S. Hadar</i> str. 18 Bela Nagy	4.79 MB	3	1	–	–	4,448	Ubiquitous	O:8
<i>S. Kentucky</i> str. CDC 191	4.70 MB ^b	53	3	ABE100000000	20069	4,383	Ubiquitous	O:8
<i>S. Kentucky</i> str. CVM29188 [70]	4.79 MB ^b	4	2	ABAK000000000	19457	4,745	Ubiquitous	O:8
<i>S. Newport</i> str. SL317	4.95 MB ^b	63	3	ABE100000000	20047	4,720	Ubiquitous	O:8
<i>S. Newport</i> str. SL254	4.83 MB	3	1	CP001113	18747	4,710	Ubiquitous	O:8
<i>S. Dublin</i> str. CT_02021853	4.84 MB	2	1	CP001144	19467	4,682	Bovine-adapted	O:9
<i>S. Enteritidis</i> str. PI 25109 [5]	4.69 MB	1	1	AM933172	30687	4,363	Ubiquitous	O:9
<i>S. Gallinarum</i> str. 28781	4.66 MB	1	1	AM933173	30689	4,466	Avian-restricted	O:9
<i>S. Javiana</i> str. GA_MS014041433	4.35 MB ^b	19	4	ABE100000000	20049	4,221	Ubiquitous	O:9
<i>S. Typhi</i> str. CT18 [19]	4.81 MB	3	1	AL513382	236	5,065	Human-restricted	O:9
<i>S. Typhi</i> str. Ty2 [58]	4.79 MB	1	1	AB014613	371	4,632	Human-restricted	O:9
<i>S. Typhi</i> str. J185 ^c	4.74 MB ^b	1065	6	CAAW000000000	28303	5,626 ^d	Human-restricted	O:9
<i>S. Typhi</i> str. M223 ^c	5.02 MB ^b	3024	6	CAAX000000000	28305	7,442 ^d	Human-restricted	O:9
<i>S. Typhi</i> str. B98-066 ^c	4.71 MB ^b	3939	6	CAAU000000000	28299	7,259 ^d	Human-restricted	O:9
<i>S. Typhi</i> str. B98-206 ^c	4.76 MB ^b	3682	6	CAAV000000000	28301	7,637 ^d	Human-restricted	O:9
<i>S. Typhi</i> str. B98-313 ^c	4.60 MB ^b	415	5	CAA2000000000	28309	5,051	Human-restricted	O:9
<i>S. Typhi</i> str. 4049 ^c	4.68 MB ^b	6441	6	CAAQ000000000	28289	10,055 ^d	Human-restricted	O:9
<i>S. Typhi</i> str. AG3 ^c	4.75 MB ^b	7336	6	CAAY000000000	28307	10,675 ^d	Human-restricted	O:9
<i>S. Typhi</i> str. B00-786 ^c	4.76 MB ^b	1445	6	CAAR000000000	28291	5,749 ^d	Human-restricted	O:9
<i>S. Typhi</i> str. B01-675 ^c	4.58 MB ^b	4564	6	CAAS000000000	28293	8,315 ^d	Human-restricted	O:9
<i>S. Typhi</i> str. B02-118 ^c	4.71 MB ^b	422	5	CAAT000000000	28295	5,097	Human-restricted	O:9
<i>S. Weltevreden</i> str. H_ N05-537	5.05 MB ^b	81	3	ABFF000000000	20591	4,784	Ubiquitous	O:9
<i>S. orishana</i> 62-04x23- str. R5X2880	4.60 MB	1	1	CP000880	13030	4,278	Ubiquitous	O:3,10
<i>S. bongori</i> str. NCTC_12419	4.79 MB	1	1	–	–	4,049	Ubiquitous	
<i>E. coli</i> str. K-12 substr. DH10B [71]	4.69 MB	1	1	CP000948	20079	4,398	Ubiquitous	
<i>E. coli</i> APBC O1 [73]	5.08 MB	3	1	CP000468	16718	5,259	Ubiquitous	

16s rRNA relating enteric species



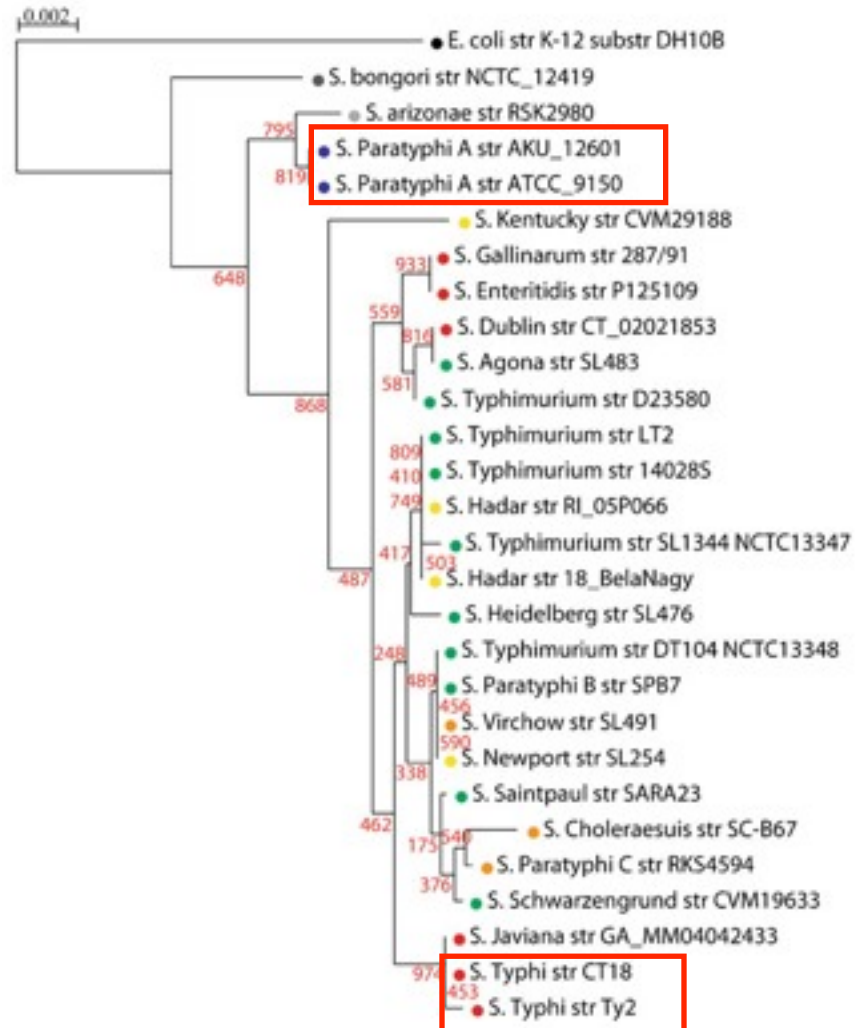
Blast Matrix

Other sub specie

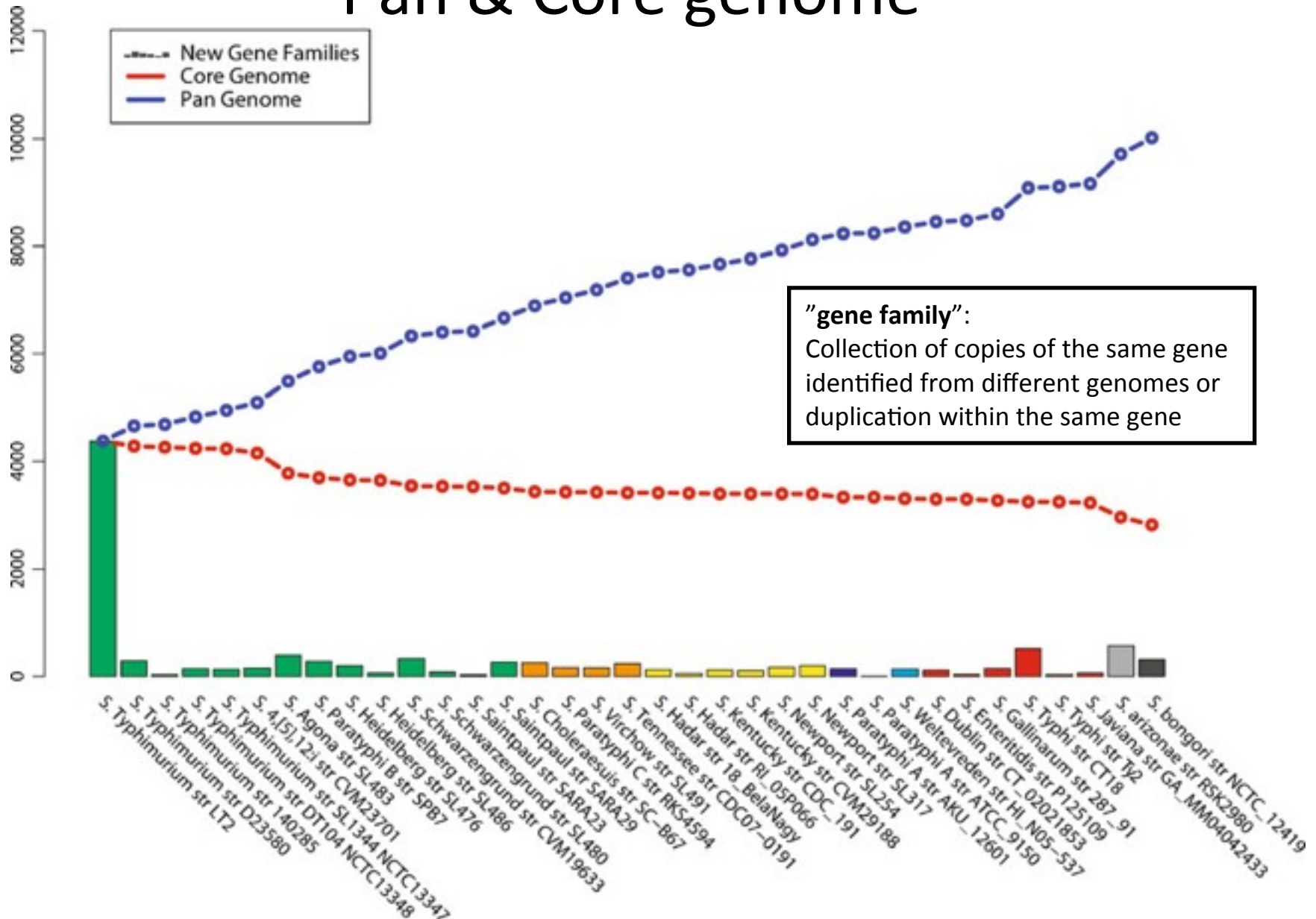


16s rRNA relating Salmonella serotypes

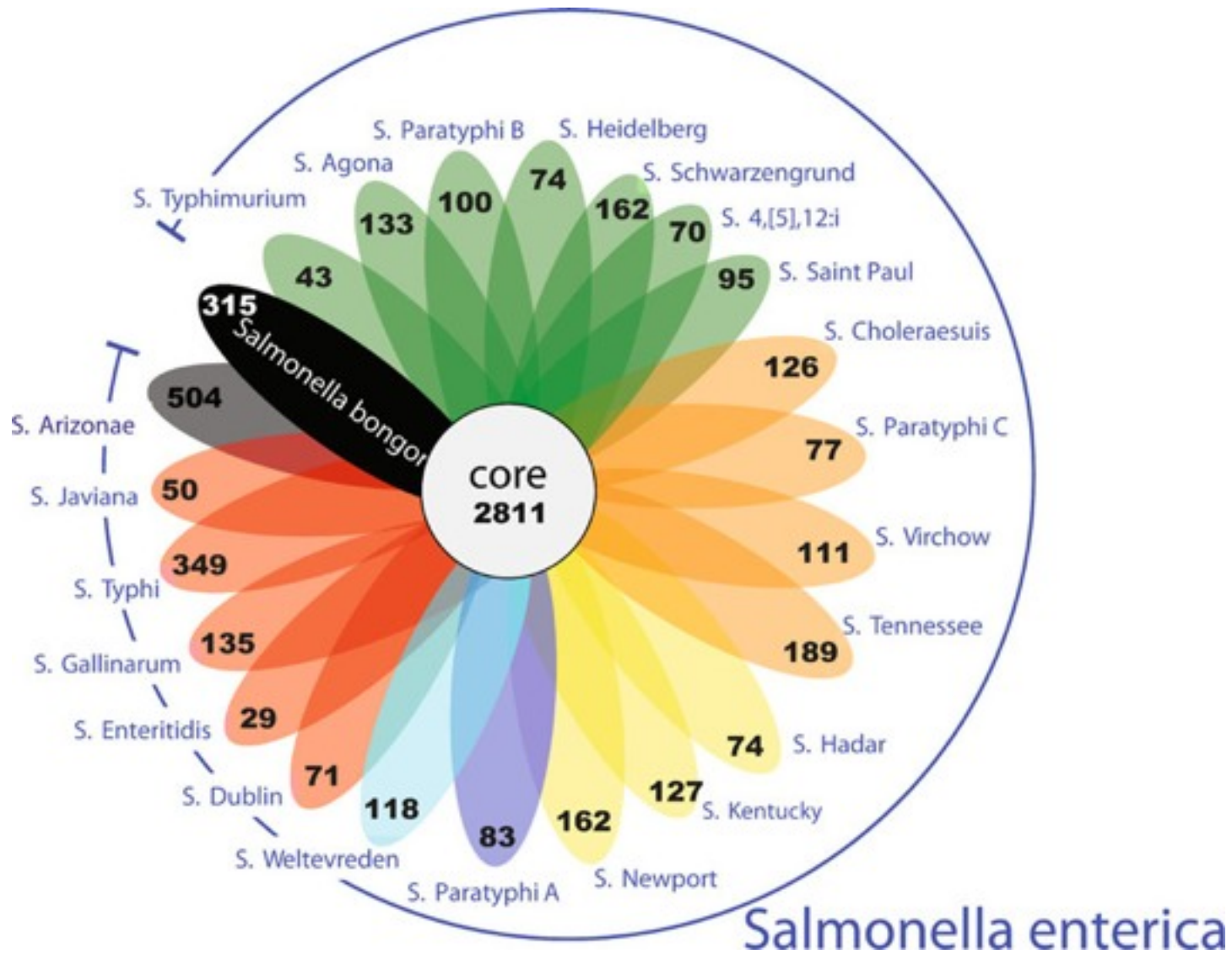
- *S. enterica*
- Other subsp.
- *E. coli*



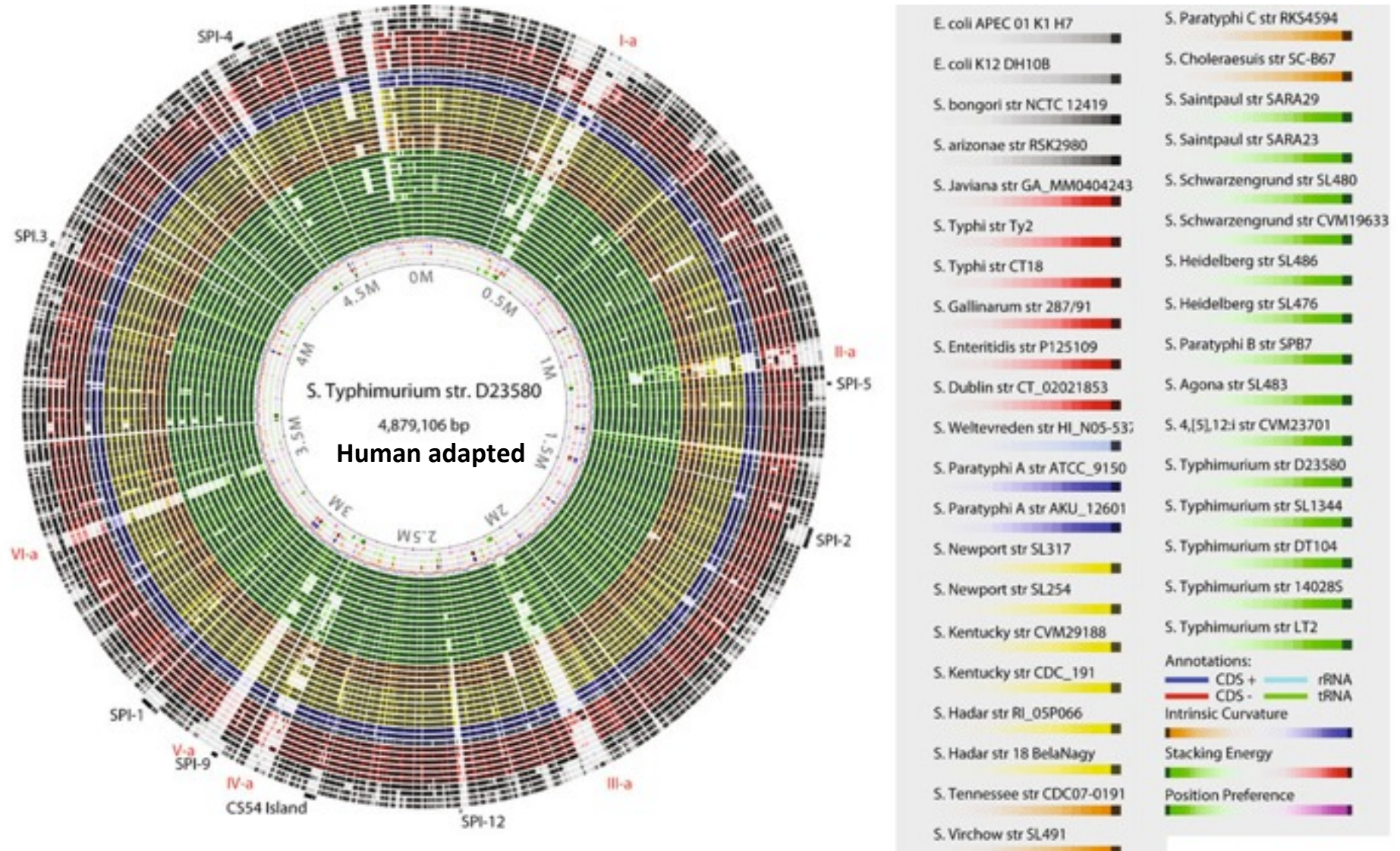
Pan & Core genome



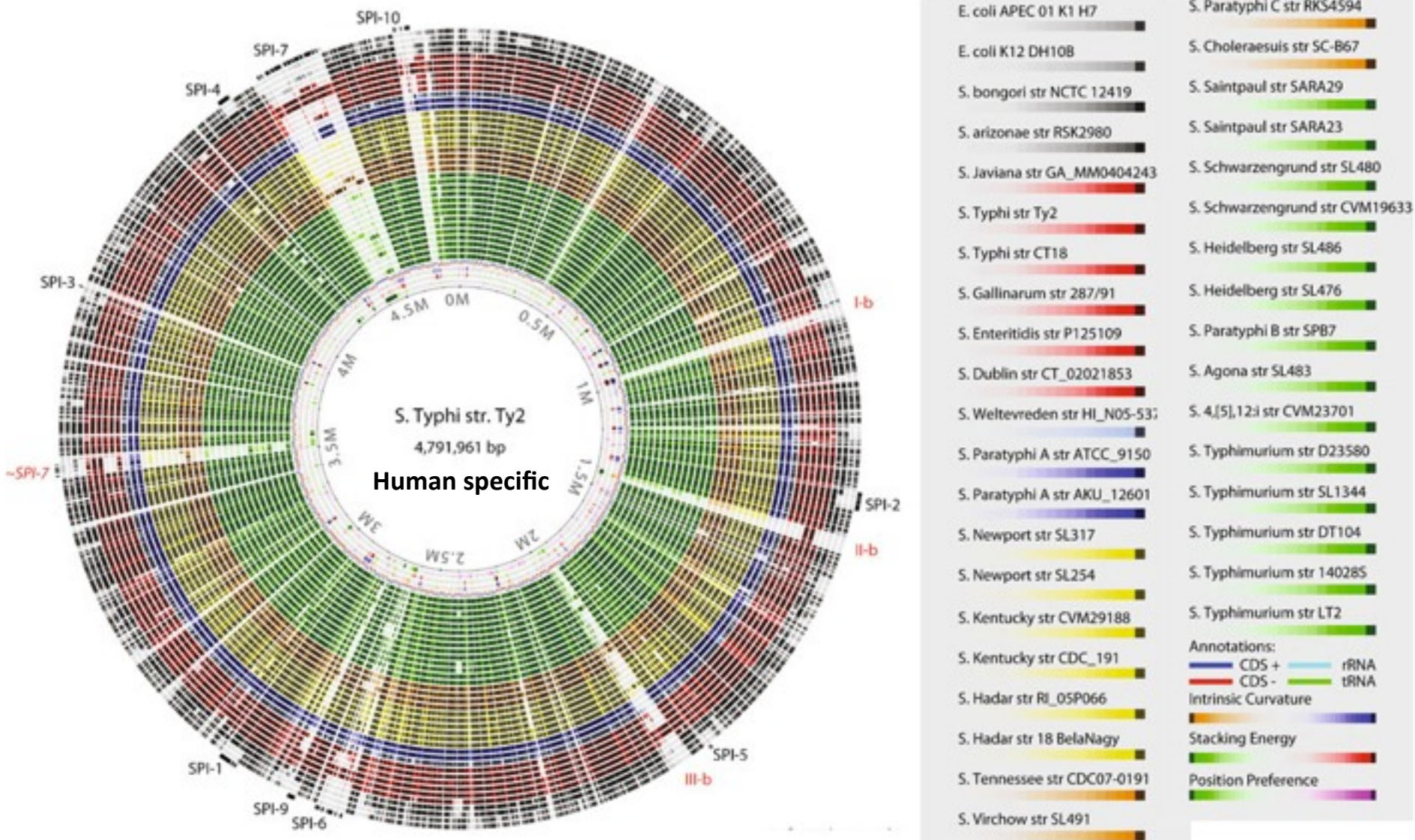
Flowerplot



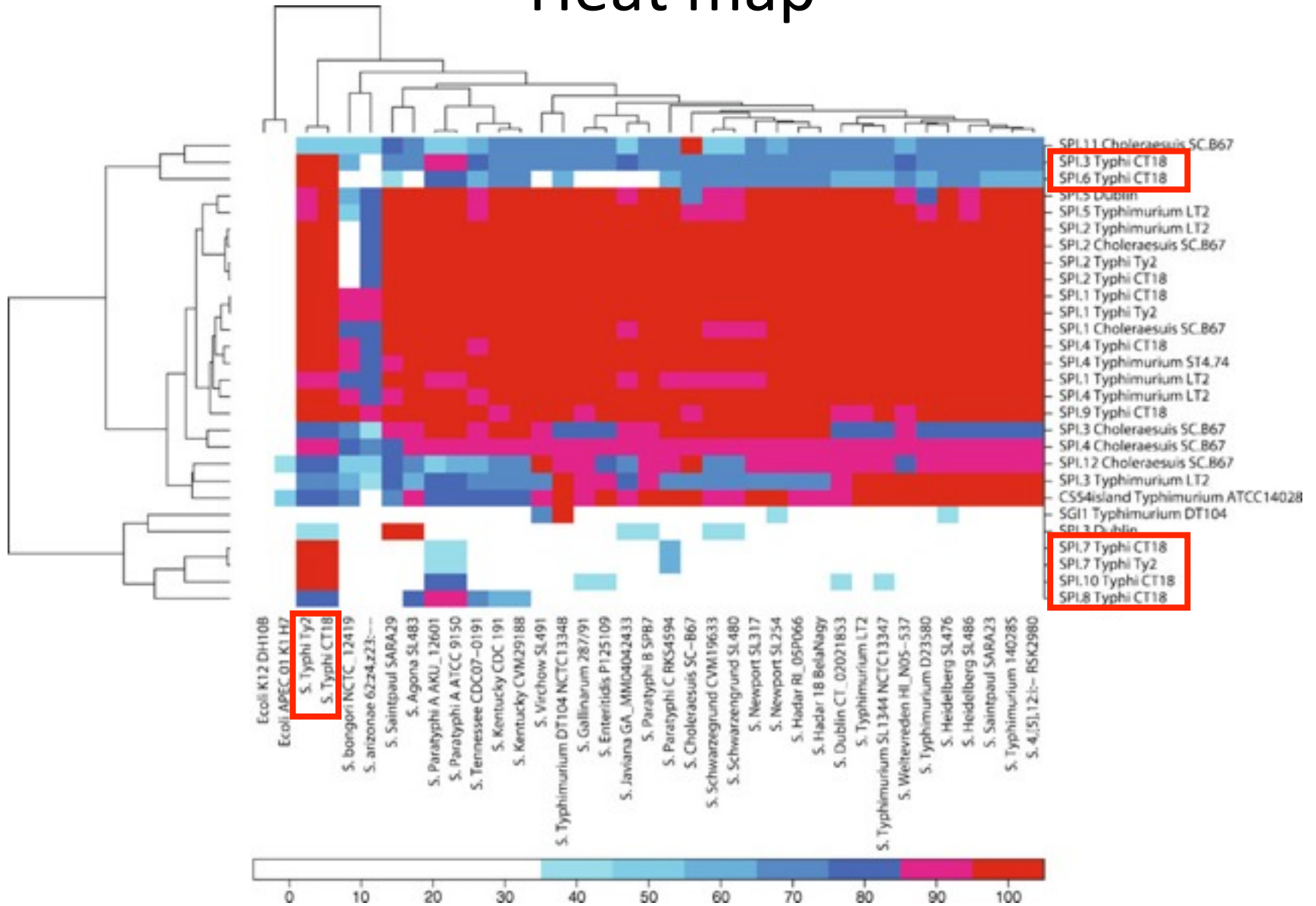
Blast atlas



Blast atlas



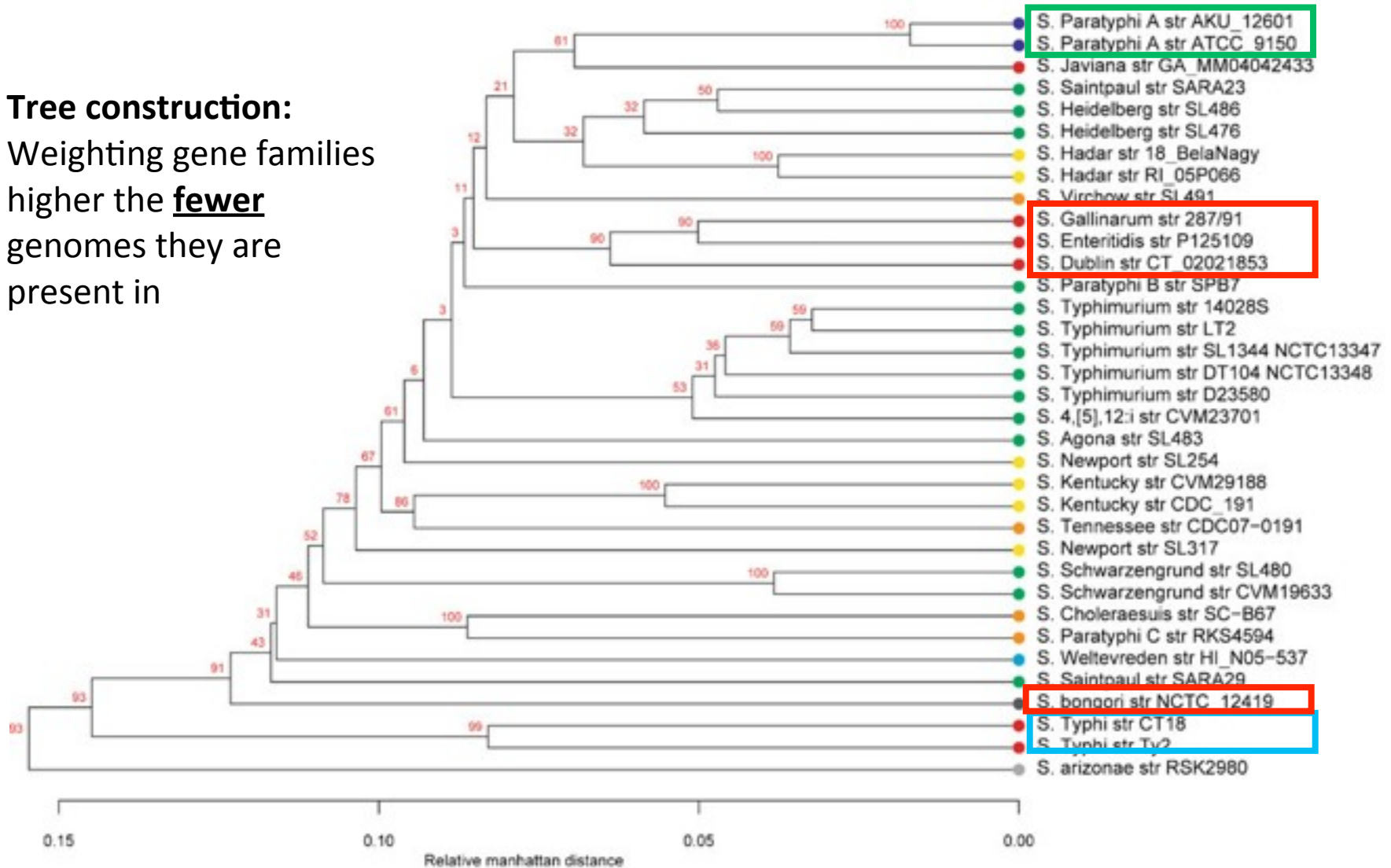
Heat map



Pan-genome family tree

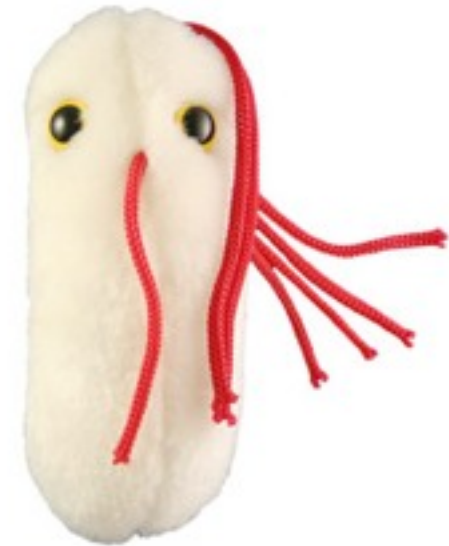
Tree construction:

Weighting gene families higher the **fewer** genomes they are present in



Concluding remarks

- Strains known for host specificity group together with strains having a broader range of hosts
- Relatively large core-genome and small pan-genome
- High conservation of SPI's
- Strain specific genome islands (e.g. SPI-7)



Critics

- Overview of the Salmonella pan-genome in general terms
- Better if higher quality sequenced genomes were used (fewer contigs)
- 16S rRNA tree made from SPI genes or other variable regions
- Presentation of BLAST atlas
- Broader representation of sub-species
- It would be nice with more future perspectives



Thursday, 15 September, 2011