

Bundesinstitut für Risikobewertung

### NGS analysis of rare *Campylobacter* isolates with ambiguous species differentiation

Marie-Theres Knüver, Julia Golz, Kerstin Stingl

NRL for Campylobacter of Germany

Species differenciation by real-time PCR

 $\rightarrow$  What do we find?

29 isolates with ambigous *mapA/ceuE* detection out of 2059 *C. jejuni* and 775 *C. coli* (Mayr et al. 2010 = Best et al. 2003 extended for *C. lari* detection; L06.32-1:2013, §64 German Feed and Food Law (LFGB))

 $\rightarrow$  around 1 % ambigous results

 $\rightarrow$  Van Rensburg et al. 2016 J. Clin. Microbiol.:

out of 1713 isolates 6 *C. coli* (0.3 % of all isolates) were ambigous in *mapA/ceuE* species differenciation by real-time PCR



Species differenciation by real-time PCR

→ Where do these isolates come from? Broiler meat, eggs, turkey meat/skin/cecum, duck meat

→ What is the real-time PCR phenomeon? "mixed" culture = both positive for mapA and ceuE (n=27) no Cj, Cc, Cl = negative for both mapA and ceuE (n=2)





### MiSeq and PacBio sequencing

	MiSeq (BfR)	PacBio (GATC)
principle	Massively parallel sequencing (MPS)	Single molecule real-time sequencing (SMRT)
Read length	2x 300 bp	~10-20 kb
DNA required	20-100 ng	~5 µg
Instrument time	2 days	2 hours
price	80 Euro	2000 Euro

- At BfR a platform is established, providing hardware (the sequencers), optimized protocols and knowledge for the NRLs
- currently the platform is working on a common pipeline for data analysis
- $\rightarrow$  team of Burkhard Malorny



### NGS analysis of Campylobacter

Illumina MiSeq of 26 of the Cj/Cc hybrids, 21 *C. coli* and 2 *C. jejuni* Several other control strains, double sequencing for defining QC Nextera XT Kit





NGS analysis of Campylobacter

Wet lab results: fastq.gz files

Ridom Seqsphere+ 5.1.0 software run under Ubuntu 16.04

Trimming reads to Phred score ≥ 30 Assembly via SPAdes 3.11.1

Ridom cgMLST (637 genes) Ridom accessory genes (958 genes)

Reference Sequence (default) is NC\_002163.1.gb (NCTC 11168)





### Quality control of the data

- ✓ Do reference sequences from NCBI match sequenced reference strains?
  - NCTC11168
  - 81-176
  - RM1221

- few allele differences
- Do PacBio sequences match MiSeq sequences?
  - 7 strains were analysed by both techniques
- ✓ Do they match each other if sequenced twice?
  - 28 strains were sequenced twice



Task Templates. C. jejuni/coli cgMLST 637 v1.3, C. jejuni/coli Accessory 958 v1.2, C. jejuni MLST Comparison Table Retrieval. CampyDB5, initial findings contains CJ/Cc. Projects: CampyDB5 (CampyIobacter jejuni/coli) Comparison Table created. Oct 1, 2018 IE16 PM (v5.1.0.[2018-063) C. jejuni/coli cgMLST 637 Cluster-Alert distance: 13 Ridom SeqSphere+ MST for 34 Samples based on 1602 columns, pairwise ignoring missing values, logarithmic scale Distance based on columns from C. jejuni/coli cgMLST 637 (637), C. jejuni/coli Accessory 958 (958), C. jejuni MLST (7) For citing correctly in publications the tools used for this analysis see menu Help | Citations. urknown 1595 5438 4148 832 1596 5468

0 #20

Ono group assigned



### Quality control of data



→ data might indicate that an av. coverage of 60 is sufficient for cgMLST analysis by Ridom Seqsphere+ (with the additional pipeline QC parameters)



### Quality control of the data

	av. Cov.	diff. % good
	MiSeq	targets
PacBio1	180	1,7
PacBio2	180	0,8
PacBio3	121	2,1
PacBio4	99	6,3
PacBio5	180	1
PacBio6	180	3,4
PacBio7	144	0,7



- although PacBio is superior in terms of gene order and assembly of 1 whole chromosome plus epichromosomal elements, insertions and deletions are frequently found
- → hybrid assembly with PacBio and MiSeq data (future task)





### Are the Cj/Cc hybrids phylogenetically linked?

Sampling location



- cgMLST + accessory genes
- pairwise ignore missing values → more differences possible!
- 3 "clusters" (à 2 strains); rest of the strains is phylogenetically unlinked
- → Multiple events as cause for ambigous species detection



page 10

BFR

## What species do the isolates with ambigous *mapA/ceuE* belong to?

Kmer Finder 3.0 by CGE genomics webpage (Kmer size 16)

**query\_coverage [%]:** is the percentage of input query Kmers that match the template

C. jejuni introgression	RT-PCR C. coli	RT-PCR Cj/Cc hybrid	oli
0%	8	1	ы С
< 3%	10	9	
3-6 %	2	1	. jejuni
>6 %	1	15	0



→ all isolates are C. coli with 0-14.4% C. jejuni introgression

### What happened at mapA?

	ultiple	Ali	900	nent (	ofta	rget	CAMPO	952																								
8 .+ +	4	80	M	9	9	9	Ø	4																								
[]		T	۵	0	0	-	tare .		0	Ø	B	T	91		0	1	1		0	10	0		I	(H)		. A. dataine	0		1	1		00
Consensus		TTO	ST G	AAAG	STCC	TGG	TGGTT	TTGA	AGCA	AAGA	TTA	AAGO	SGCT	TTT	ATA	CAT	TAGO	GAT	GTT	GGA	TTC	AAT	GTTO	TGCC	AATA	LAACG	CAC	TTTA	GACA	CTGG	TATTG	CTTTGA
BIRCAL 4216			c	G	T	A	TOUTTI		G	۸	C	1.4.6	C			TG	1.4.4.5		A	G	A	AT		20.00	1.0.7		٨	1111	T	G	GG	CC
BIRCA15282		r fin	100	AAA	1.0				OC.		T-T-	A.A.0.1	reer		-	2.	LAGO	G.R.T.	6 T T	0040	TIT	TAAT								100	1.5.2.2.	THE OAL
BfRCA1 4751	1 1		C.	G	To	· A :	TOSTI	TT SA	Galera	A	1 C	6641	C	111	ATA	TG		OAT.	AT T	GG		AAT		1001	AATV	AAA	AAC	TITA	ATA:	G	GG	CTC
BfRCA15395	e				ITCL							1.1.1.1		111	×1 ×	2011	10.00	HAT.	TT.	0.0.4	11.0	AAT	at ti	1 4 6 4	- 23						1.1.1.1.1	TITAA
BIRCA02511	DSM.	1.11	C.	G	T	A		DI NA	G	A . A	C		C		51 A	TG	r.all.C	UAT	A	GGG	A .	NAT	at i i	TOTA	1.1.1.2	AAA	A	P.T.T.M	T A	G	GG	CC
BfRCA09557	Pac.,	111	Ð	AAXI	1100	1.01	(1) 点前定于1	51 R.A.	G	A	- C -	A.A.313	C	3.2.7	~1 A	TG	1.0.25	1.62.	ATT.	G	1.	AAT!	6773	1 4 4 4	1.1.1.2.2		Λ	1.7.7.8	I AT A	G	GG	C C
BfRCA11057		110		4440	11,69	1.8.6	TRATIT	TTRA	G	- A	T C-	4.4.63	C	TTT	4.7.4	TG	1465	GeT.	ATT	G	A	AAT		TRO	CAT.Y	A	A	TTT	<b>TAT</b>	G	GG	TT CT C
8fRCA13971	8 (B		C.	G	T	I All	10.0111	0.94	GOOR	A	C		¢			TG	1899	947.	A T	0 0 G	A	0.010	(11)	1.904	ALC: N	A	A	1114	OF THE	G	GG	C C
BfRCA13264	ŧ., , , , ,	100	2	G	T	* <b>A</b> :	TARTI	CT HA	G	A 1.	· C	1621	C	3.3.3.	87.8	TG	1,649,5	0.0,73	A 1 11	G	•	PAT:	61.11	1.001	MAT 7	(444)	A	1114	T A	G	GG	CC
BfRCA14705	EQAS:	111	50	G	TIC	A	T O UTIT	TT II.	GORA	A .	c		C	111	12.4	TG	T.A.B.C	0.019	ATT	G	A			11-0.07	A.S.Y.A	Α.	Asc	TTTA.	The	G	GG	CT CT C
BIRCA14815	( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( )		5	8883	1.59	5.82	医温温剂 打	1728	Gaster	A	C	4.6.23	C	111	51.5	TG	1,495	OAT)	A	G	A			1.954	6877	A	A	1.1.1.8	to In	G	GG	C C
BIRCA15034			5	G	T.	A	TRATTI	TT BA	GOGA	A	C	1100	C	110	110	TG	1 4 4 5	9 1 1	AT T	G	•	MAT.	OT T I	100	5 A T /	A	A	TTTA	0 A 🖸 👌	G	GG	CC
BHRCA15062			5	G	11	10		1.1.1.4	Gode	1	ç	4401	e l		27.2	IG	t sin c	0.61	A 1	O O G	A	1		195	5 II Y I	A .	A	1114	na Is	G	GG	cc
BIRCA15077			8	G	110	100	1.0.01.1.1	1.1 8.5	GOCA	00.00	e e	5525	8		1.1	10	1 1 1 1 2	0.013	2	O O G	<b>A</b>	202		196	ALC: N	A	200	111.4	1.1	G	GG	C C
BTHCA14583			5	G		0		10.00	G	•	5		8	250	22.0	IG	10.05	0.417	1	G	-			1.001	A.A.T.S	A .	A .	111.4	1.15	G	GG	C C
BIRCALS3/1			×.	G	-	12	19411	1.00	G	A	2	1003	2	111	210	TC		0.414		G				1.1.1.1	6.17	A	2	1114		G	66	
DIRICAL 5403			2	6	4	2				002	2	22.15	2	1.51	22.2	To			2	6	2	203	2.1.1		2.000	A	2	1110	1040	G	6.6	22
BIPCALSO23			1	u.		0		1120		022	2		2	1000	210	TG	1012	2.5	2	6		201			221		~	2020	1010	G	6 6	2.2
DIDCA150AC				6					6		2	1111	ě.			TG		2.2	2	6									4	G	6 6	2.2
REPCATS978			2	G	÷.	- 2	7.0.07.1.1		G		è		č	220	210	TG	10.23		2	G		1011			2013	2	2	1111	÷.	G	GG	è è
RIBCAT 5301			ā.,	G	T				G		č		è.		242	TG				G			0Y V 1	1.1.1.1			Δ.		1.1.1	G	GG	C.C.
BfRCA13895	b:		26		-			-				ANGI	1000	1.7.7	47.3	2.00	taor	DRT.		0.0.4		TAAT	arr a	TOCI				-				
8fRCA15426	b:			AAAI									1007	÷	A.1.A	201					42	AAT	67 T I	7.0 03								TITOAR
BfRCA14731		111		AREL							1.	AAUS		111	ALA	CALL	AUE	0.41	11.1	0.0.4.	110	AAT	0111									T110A.
BfRCA13919		1.1	1.0	AAAO	1100						1 - 6	4.4.01	1001	TIL	27.4	CATI	TAGE	D.A.T.	1.7.0	COA.	tre	A & T	GTTO	TACE		44.00					TATTA	TTTOA
BfRCA15005	e i	1.1			IT CO							6841	10.01	TIT	AT A	CATI	TRAC		LTT.		117-0	AAT:	arra	1000	ANTA							TTTUAN
BfRCA14943	£	100		AAAI							114	A.A.111			A.I. (6	CA11	1 16 16 1	11.4.7		0.0.47		= AT	or in	TODO	A 1 1							COLUMN T
BfRCA15287		1.1.1		AAAI	(t)C					A.X. 17	3 I A	A.A. () (	So Ct	111	AT. A	CA13	t A U.C	0 AT	11.1	0.0.4	1.1.6	AAT!	61.1	11.000	A KYO				11.28			TIT TOWN
BfRCA14825	s 🔅	111		1223										111	82.8						17.5	DAT.		10.03	- 61.	ARCO	ma	anΔ	-R			TTTTUKA
BIRCA11586	6 3	110		2223			ma	nΛ	E			44.07	19.57	TIT	ATA		1:4.05	5.6.7	61.1	G-DLA.S	11.6	(1)A(1)	653 T #	计声音的	48.83	- Andrewson	пс	1pr				TITIGAL
BfRCA15396	b:	119		AAA	DI CT		1119	pA-		12.41	118	AAGI	10.07	110	GA	CAT	TA05	DAT	011	0.0.8	1.13	EAT!	0110	TOCI	ALT	4450	(CAG	LLLA	9 A G A	CO	14110	CTTTOAR
BfRCA15286		11.0			IT CT		1000	(1. K. K. )				AAQI	1.0 CT	TTT.	×1,4	SAT	t Kac		111	0.0 A	TTC	AAT.	6111	IT OCC								TTTTUA.
BIRCA13120				2.7.7.1	1100						1.14	AA () I		TT T.I.	5.5 K	1	1,8114	10.01	117	0.0.4.4		AAT.	arti	10.000	ALL							Q.1.1.0.0
BIRCA13188				AAA3	100						1.1	0.031	1951	2.2.2	83.6	CAL	1895	0.41	1.1	00A)	11.0	NAT:	0.024	1.923	111		2.4					TTT BAS
BTRCA13619															AL.		1995				UUS	10.0	ette	TREE								TT AAA
BIRCA14833				2001							1.0	20.01					10.05	0.01		CO.A.	1013	201	6111	1955								- 104
DIPICAL 5420	5			2223								20.21	963		6	202	26.85	242		0.0 % /	1.1.0	202	2111	1953								A T T OA
=/= - · · · · · ·	: S	-		22.00			o star to							Call and S	9	-		COLUMN 1		0.004.0			CL.L.		100		1.0.0					C.C.C.C.C.C.

 Adjacent genes also implicated in recombination event – in some isolates Cj sequence spans over 4 genes



### What happened at *ceuE*?

Multiple Alignment of target CAMP1271



 15489 has ceuE form C. jejuni; the 14973 has a ceuE allele below the threshold of Ridom



### Could the Cj/Cc hybrids be correctly identified via other PCRs?

Gene target	function	Species	Ridom scheme	Discrimi- natory in "hybrids"	references
glyA	serine hydroxymethyl- transferase	C. coli	core	yes	Wang 2002; LaGier 2004
hipO	hippurate hydrolase	C. jejuni	accessory	yes	Wang 2002; LaGier 2004, Toplak 2012
cpn60	chaperone (groEL)	C. jejuni/ C. coli	core	partial	Chaban 2009
cadF	Fibronectin-binding protein	C. jejuni/ C. coli	accessory	yes	Toplak 2012; Shams 2016
lpxA	N-actylglucosamine transferase	C. jejuni/ C. coli	accessory	yes	Klena 2004
ccoN	cytC subunit	C. jejuni/ C. coli	core	partial	Toplak 2012
mapA	Outer membrane protein (MOMP)	C. jejuni	accessory	no	Best 2003; Mayr 2010
ceuE	enterochelin transporter, substrate-binding	C. coli	accessory	no	Best 2003; Mayr 2010

- Results on phylogenetic tree based on gene sequence; oligo annealing sites have to be checked
- Wang et al. 2002 multiplex-PCR identifies *C. coli* (*glyA*) in wet lab



### Are other core *C. jejuni* genes, never found in *C. coli*, recombined in the Cj/Cc hybrids?

- 48 genes were defined which belong to the *C. jejuni* core and are never found in *C. coli* (Lefébure et al., 2010; Meric et al., 2014)
- we could detect introgression of *C. jejuni* in 8 of these genes
- 15 of the Cj/Cc hybrids and 2 of the *C. coli* harboured core *C. jejuni*

	core (L)	core (L)	core (L)	core (M)	core (L/M)	core (M)	core (L)	core (M)
Initial_findings	Cj0004c	Cj0005c	Cj0037c	Cj0135	Cj0327	Cj1062	Cj1589	Cj1714
C. jejuni	6	18	10	1	1	9	12	1
C. jejuni	4	23	76	1	7	36	? (failed)	5
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	8	? (not fou	? (failed)	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
C. coli	2	15	? (not fou	? (not fou	? (not foun	56	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	29	1	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	1
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	1
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	16	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	3	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	13	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	18	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	1
Cj/Cc-Hybrid	? (not fou	? (not fou	21	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	3	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not four	? (not fou	? (not fou	? (not fou	? (not found	? (not four	? (not four	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	23	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	23	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not four	? (not fou	16	? (not four	? (not found	? (not four	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid	? (not fou	? (not fou	? (failed)	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid*	? (not fou	? (not fou	? (not fou	? (not fou	? (not foun	? (not fou	? (not fou	? (not found)
Cj/Cc-Hybrid*	? (not fou	? (not fou	22	? (not fou	? (not foun	13	? (not fou	? (not found)



### **Conclusions and perspective**

- mapA/ceuE species differentiation is ambigous for ~1 % of Cj/Cc isolates but does not provide false identification
- NGS analysis revealed them to be *C. coli*, which was cofirmed by Wang et al. 2002 (*glyA*)
- $\rightarrow$  Good decision to add both PCR options as Annex to ISO 10272
- QC analysis suggested that an average coverage of ≥60 might be suitable (Ridom pipeline)
- PacBio data should be "cured" for insertion/deletions by MiSeq data (future task to use hybrid assembly such as Unicycler)
- Further characterization of the *C. coli* with hudge amount of *C. jejuni* introgression
  - Are they special or just found by chance?
  - Do they have extended DNA uptake capacity?
  - Is there a trigger for DNA exchange by natural transformation?



<u>Thanks to...</u> the Federal Laender laboratories for isolating *Campylobacter* 

BfR NGS platform Maria Borowiak & Burkhard Malorny

<u>NRL team:</u> Christiane Buhler Britta Kraushaar Maja Thieck Petra Vogt Imke Wulsten

# Thank you for your attention

**Kerstin Stingl** 

Phone +49 30 - 184 12 - 2135 kerstin.stingl@bfr.bund.de • www.bfr.bund.de/en





Bundesinstitut für Risikobewertung

