Welcome to Day 2 of the SciLifeLab in **METAGENOMICS**

Schedule

This course is given at Navet (E10) at BMC. Click the heading of a topic to see the lecture slides or lab instructions.

Tuesday 24th

Room: E10:1308/1309 (BMC, entrance C11)

09.00-09.30 Lecture: Introduction to metagenomics (Anders Andersson)
09.30-12.00 Tutorial (incl. coffe break) Metagenomics (Johannes Alneberg, Anders Andersson, Sophie Charvet, Thijs Ettema, Yue Hu, Luisa Hugerth, John Larsson)
12.00-12.45 Lunch
12.45-17.00 Tutorial (incl. coffe break) Metagenomics (Johannes Alneberg, Anders Andersson, Sophie Charvet, Thijs Ettema, Yue Hu, Luisa Hugerth, John Larsson)
18.00 Dinner at restaurant Koh Pangan

Wednesday 25th

Room: E10:1308/1309 (BMC, entrance C11)

09.00-09.40 Lecture: Single cell genomics (Thijs Ettema)
09:40-10:00 Lecture: Single cell genome assembly using SPAdes (Kasia Zaremba-Niedzwiedzka)
10:00-10:20 Coffee break
10:20-12:00 Tutorial Single cell genome assembly (Anders Andersson, Thijs Ettema, Felix Homa, Kasia Zaremba-Niedzwiedzka)
12.00-12.45 Lunch
12.45-17.00 Tutorial Single cell genome assembly (Anders Andersson, Thijs Ettema, Felix Homa, Kasia Zaremba-Niedzwiedzka)

Single cell genomics in a nutshell

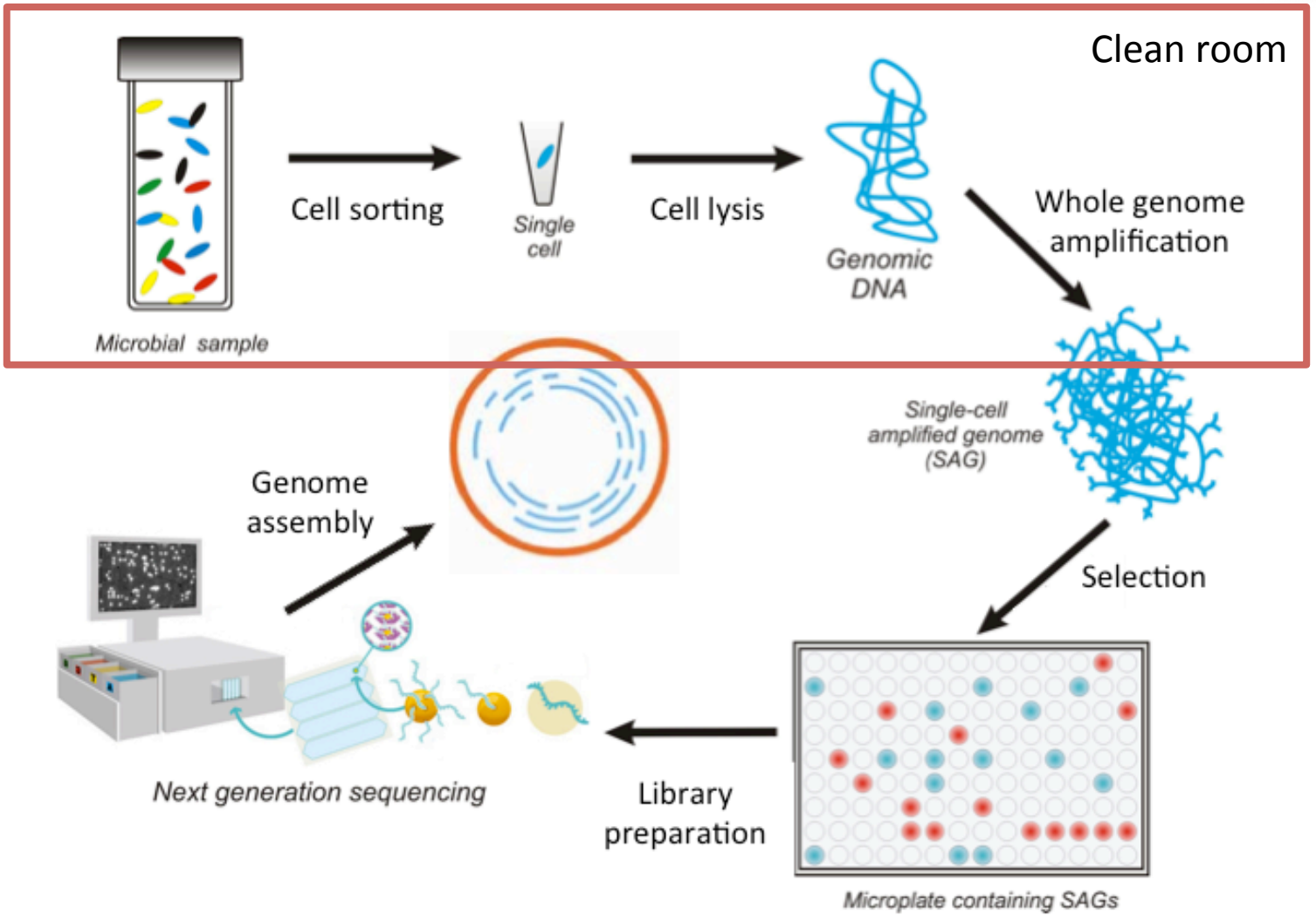
Courtesy of Carrie Jackson

Thijs Ettema

Single cell genomics in a nutshell

- Single cell sorting
- Whole Genome Amplification (WGA)
- Library preparation and NGS
- SCG example study
- The Pros and Cons of SCG
- Future directions of SCG
- SiCell: the SciLife platform for Single Cell Genomics

Single cell genomics in a nutshell

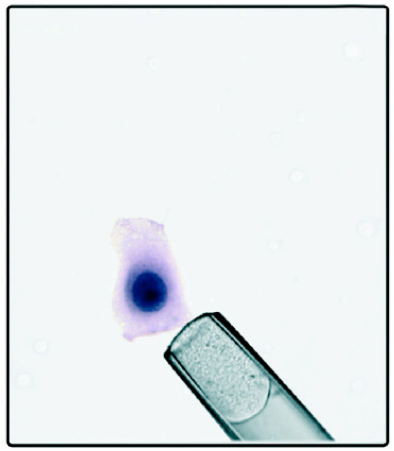


Single cell sorting

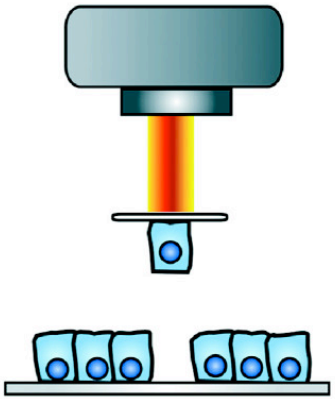
Cell material, some considerations

- Fresh material = better
- Intact cells (do not freeze without cryo-protectant)
- Cells should not be aggregated

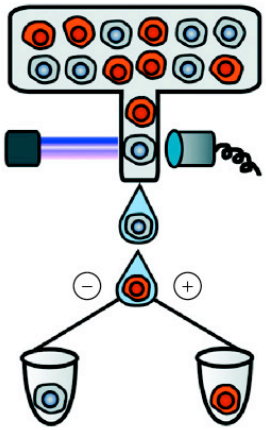
Different approaches



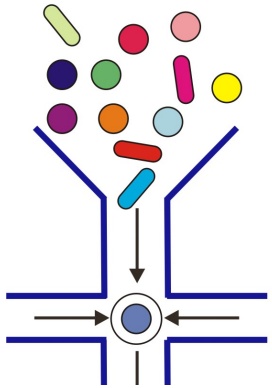
Micro-manipulation



Light-capturing microscopy



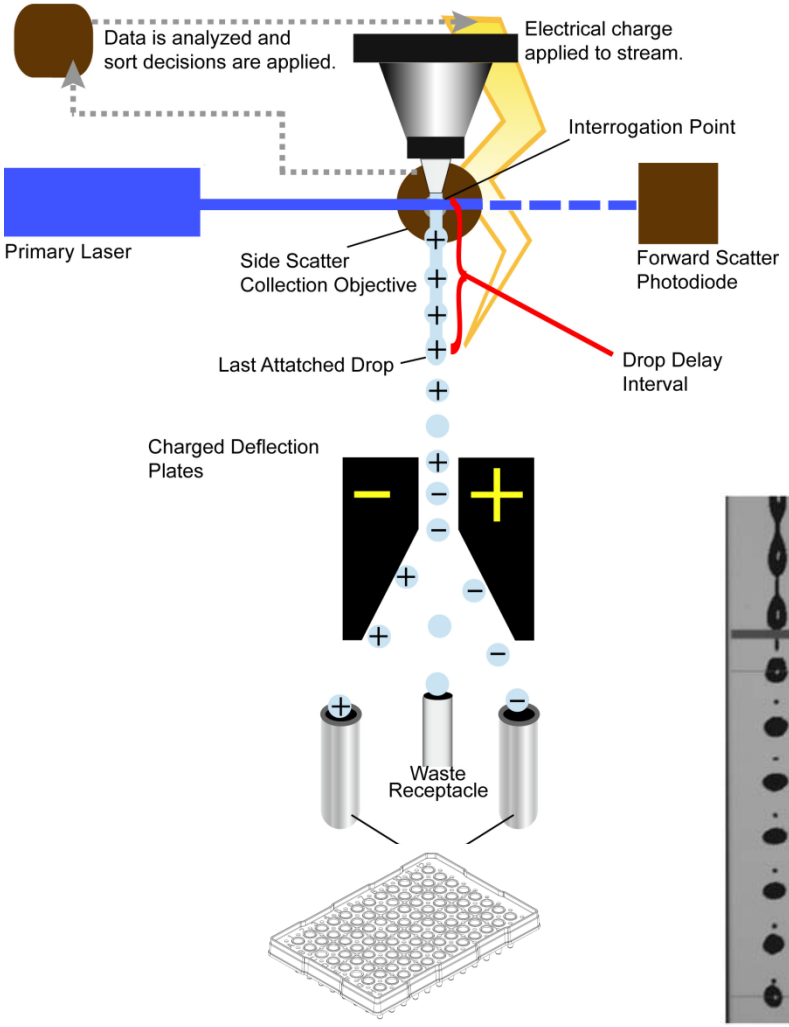
Fluorescence-activated cell sorting (FACS)

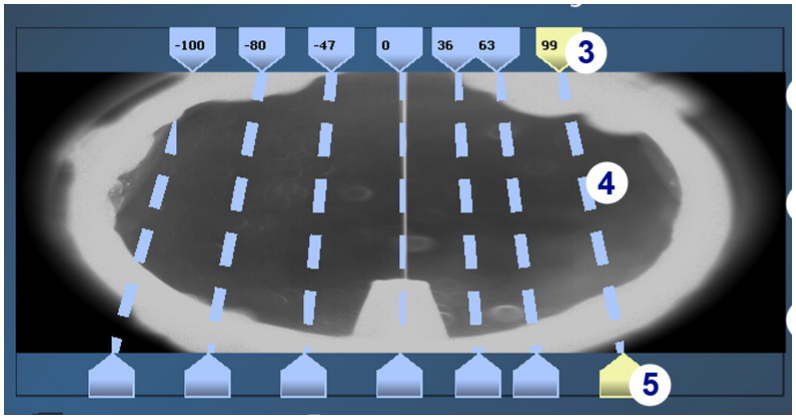


Microfluidics

Single cell sorting

Fluorescence-assisted Cell sorting (FACS)

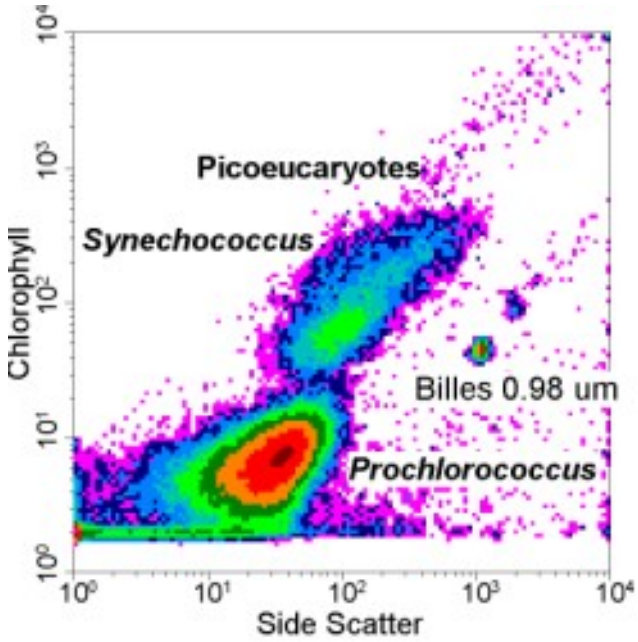




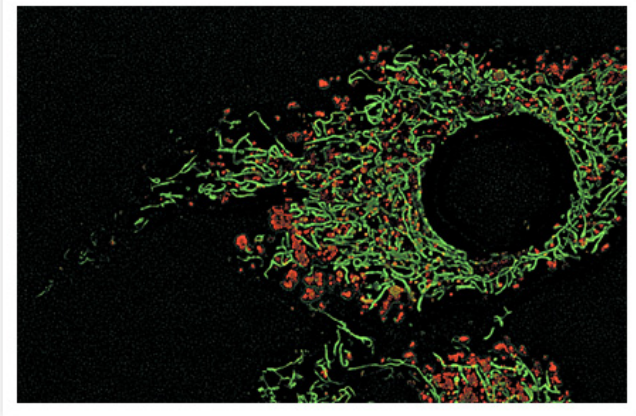
- Advanced Cell sorters can do 6-way sorting, i.e. sorting in different quadrants.
- Cells 0.2 – 100 um in diameter
- Typically in multiwell plates, e.g. 96, 384 or 1536-well plates

Single cell sorting

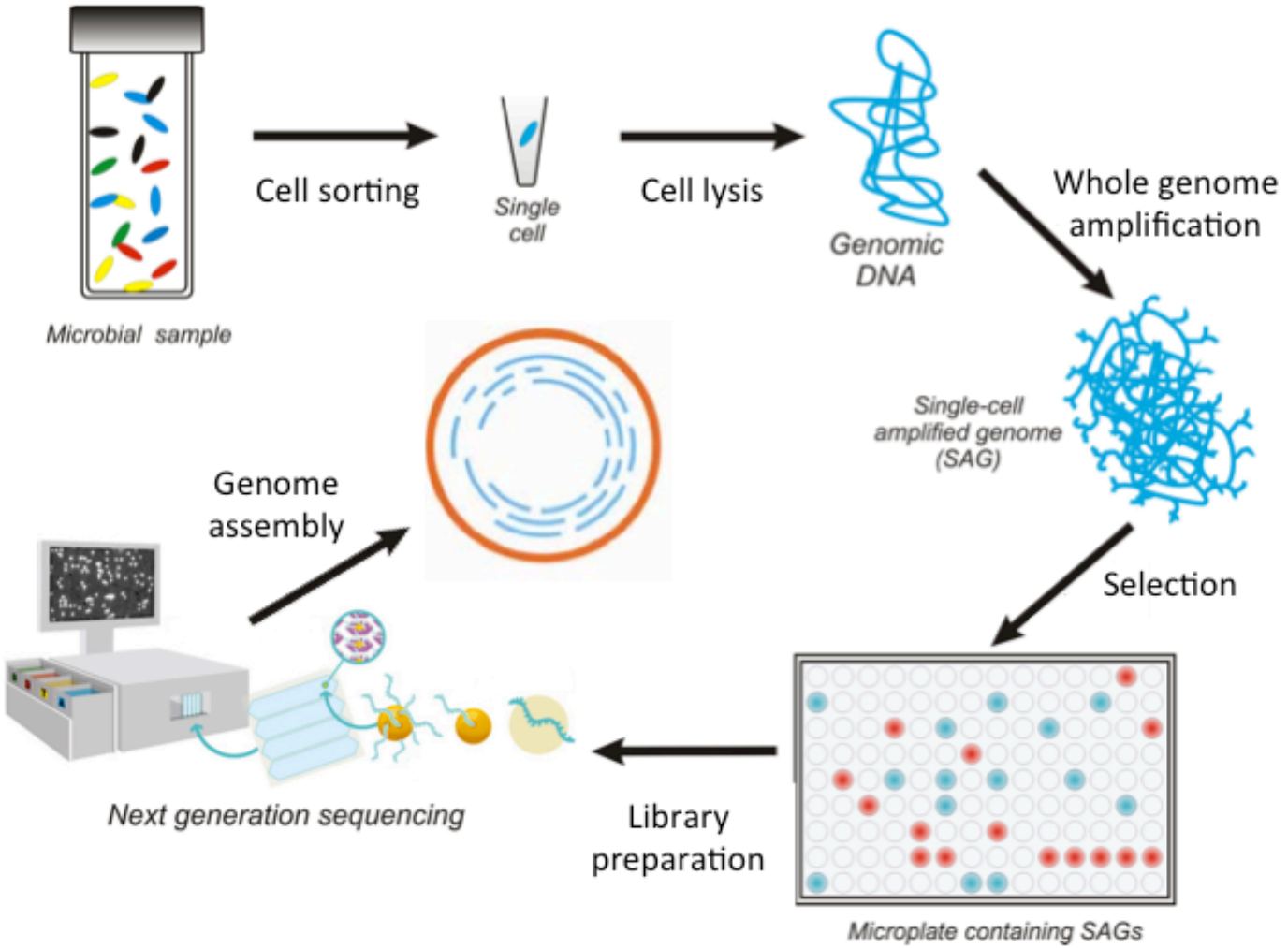
- Sorting subpopulations
  - ’Endogenic’ markers (fluorophores)
  - Cell size (Side scatter)
  - Specific probes



1. Total, High and Low Nucleic Acids (staining with SYTO-9)
2. Prokaryotes with active electron transfer system (probing with CTC)
3. Prokaryotes with intact cell membranes (staining with SYTO-9 and propidium iodide)
4. Prokaryotes with esterase activity (probing with CFDA)
5. Mitotracker and Lysotracker dyes for staining of mitochondria resp. lysosomes in eukaryotic cells

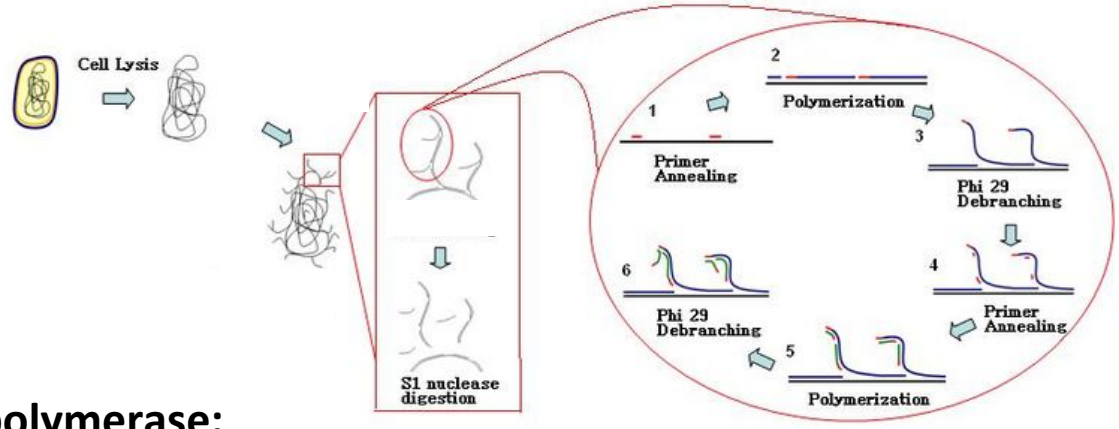


Single cell genomics in a nutshell



Whole Genome Amplification

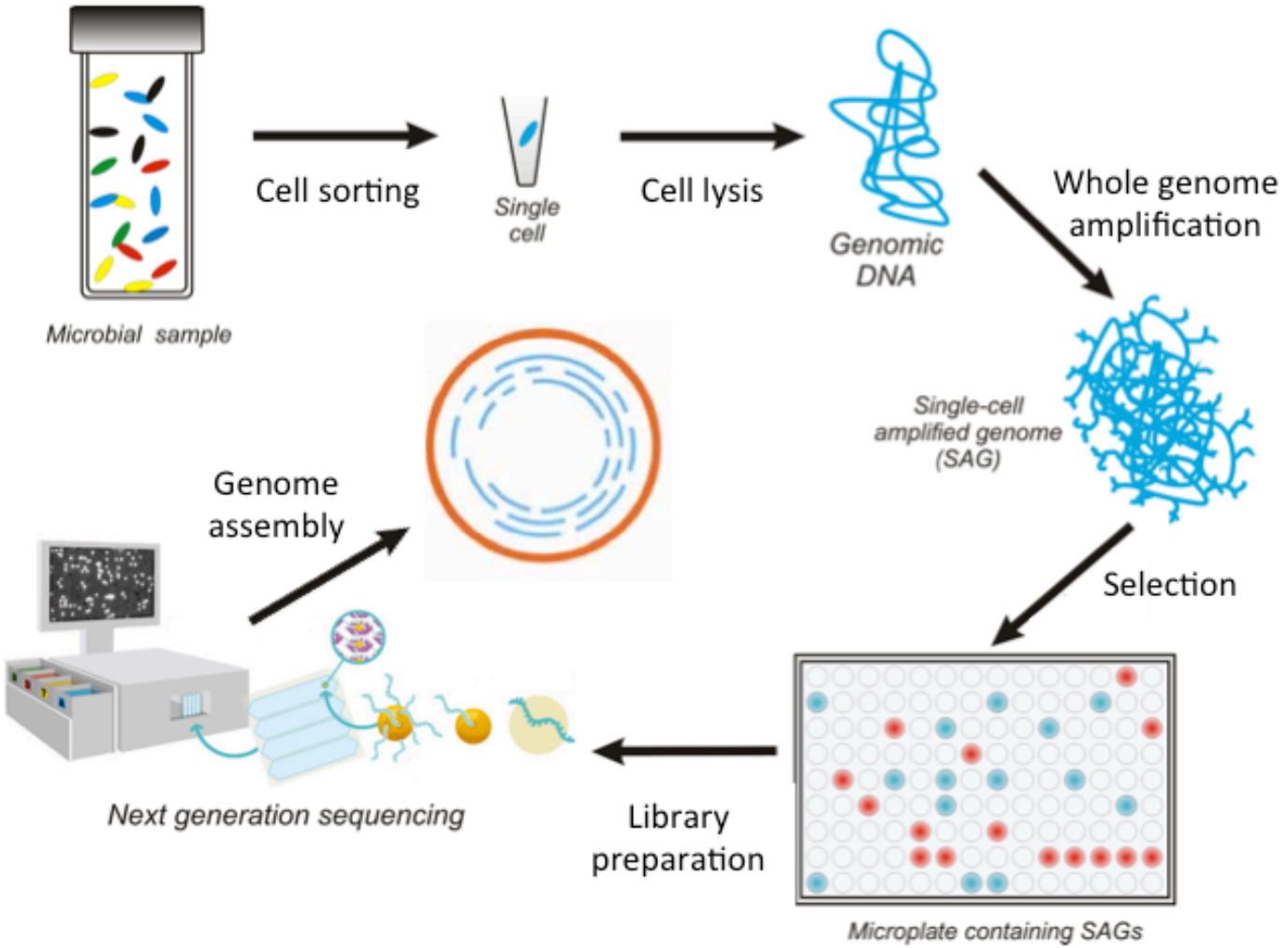
Multiple-strand Displacement Amplification (MDA)



Phi29 DNA polymerase:

- Highly processive (>70 kb DNA fragments)
- Strand-displacement activity allows for efficient isothermal (30 C) DNA amplification
- 3'→5' exonuclease (proofreading) activity:
  - 3'-modified primers recommended
  - High accuracy (~1 error in 10^6–10^7 bases)
- Displacement activity causes chimeric DNA fragments (1 chimeric junction per ~5 kb)
  - Large insert libraries not recommended
- Amplification is not uniform: severe amplification bias
- Amplification is not complete (typically 30 – 90%)

Single cell genomics in a nutshell

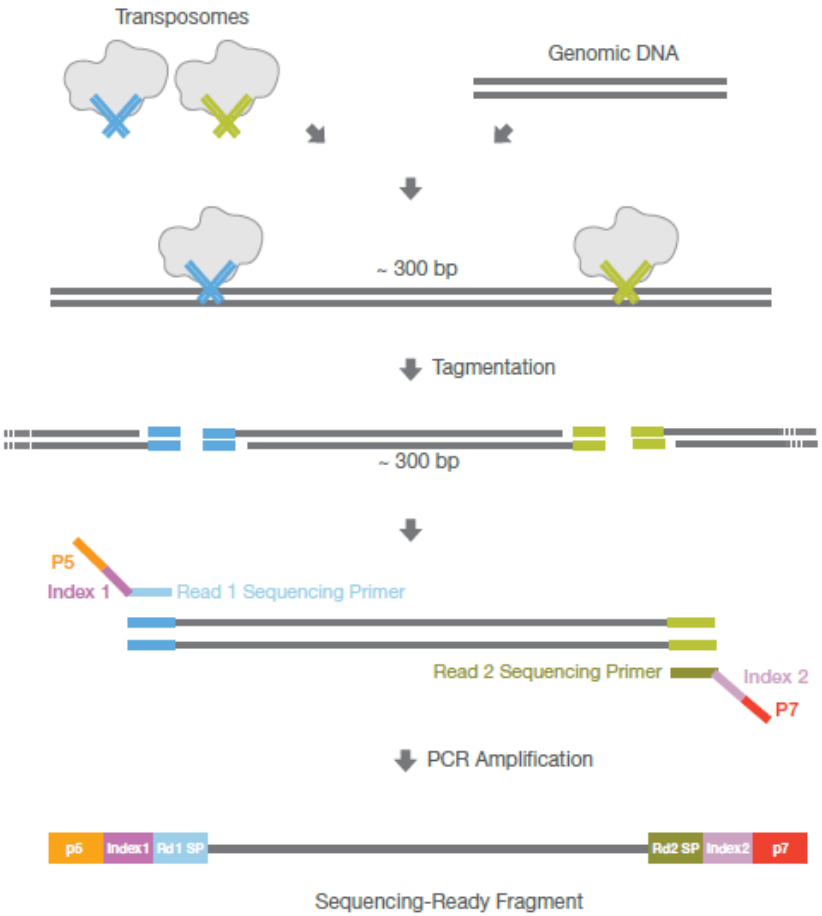


NGS-based SAG sequencing

- Select SAGs or random
- Library creation:
  - Re-MDA (expect more redundancy!)
  - Alt. Use low-input library kits
- Chimera formation
  - Constraints regarding creating LIPE libraries
- Illumina HiSeq2000: sequence 96 SAGs per lane (200-300X per SAG)

Low-input library kits

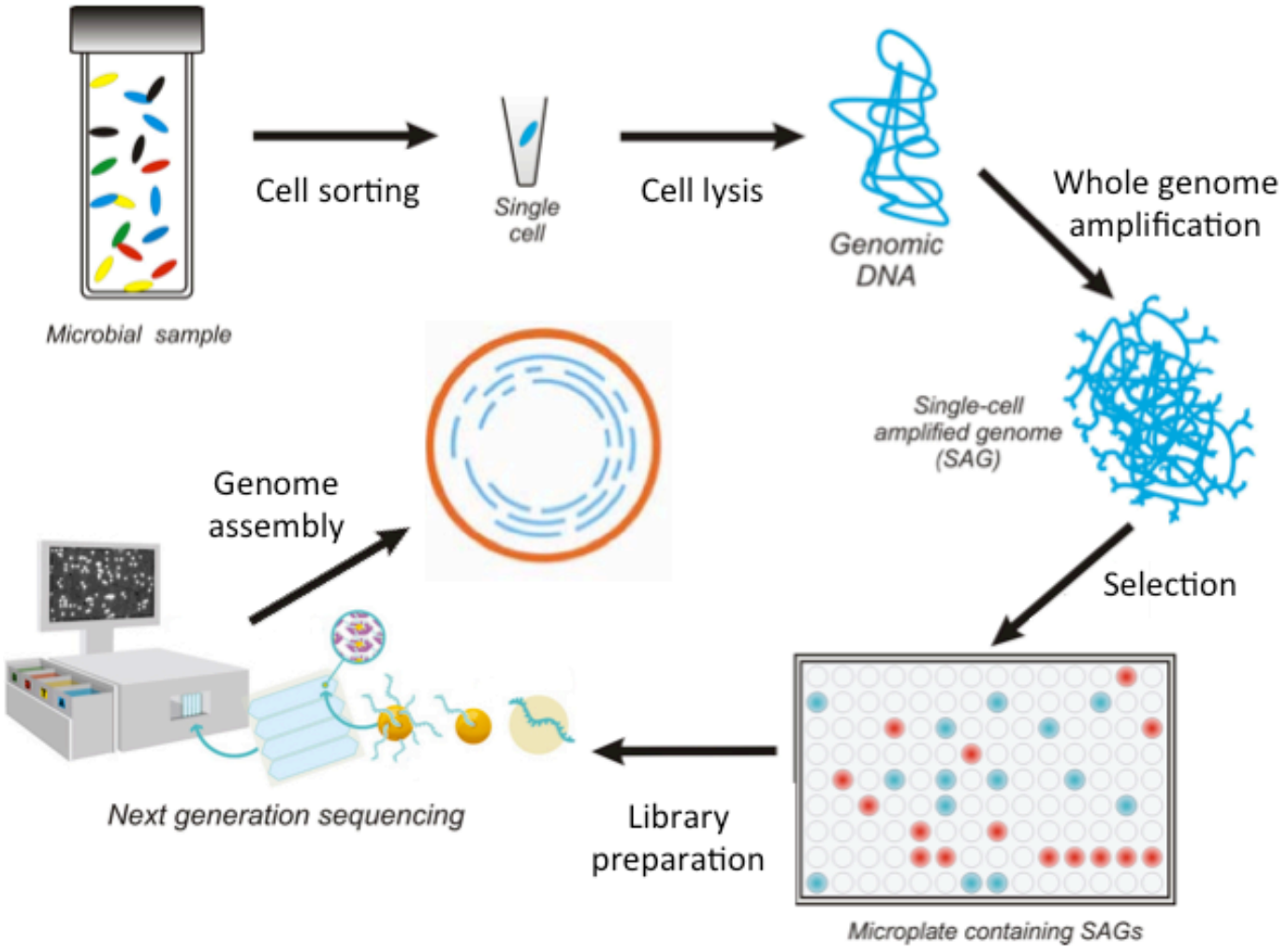
Illumina NextEra library kit



- Transposon-based DNA "tagmentation"
- NextEra: input = 50 ng ($80/library)
- NextEra XT: input = 1 ng ($40/library)
- Insert size ~300 bp
- Multiplexing: Currently 96 barcode combinations available (8X12)

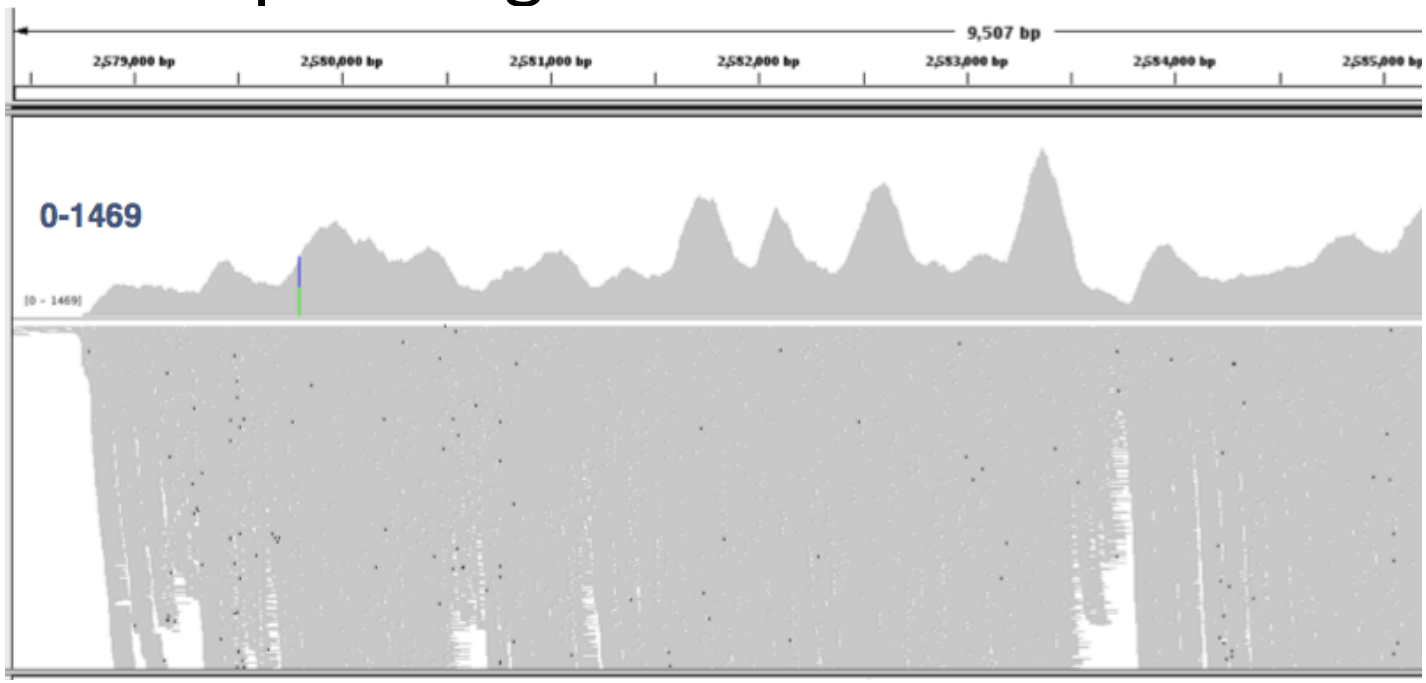
Alternative low-input libraries available from New England Biolabs (NEBNext-Ultra) and several others

Single cell genomics in a nutshell



Assembly (more details lecture Kasia)

- Uneven coverage is bottleneck in assembly of SC sequencing data



Available SC-assemblers:

- Velvet-SC (Chitsaz *et al.* (2011) Nat. Biotech. 29, 915-21)
- SPADes (Bankevich *et al*. (2012) J Comput Biol. 19, 455-77)
- IDBA-UD (Peng *et al.* (2012) Bioinformatics 28, 1420-8)

Assembly (more details lecture Kasia)

- Dedicated SC assemblers outperform 'normal' assemblers
  - contigs
  - IDBA-UD: iterate *k*-values from small to large to build a more efficient de Bruijn graph
  - SPAdes: Error-correction, Paired de Bruijn graphs (PDBGs), k-bimer adjustment

Assembly	NG50	# contigs	Largest	Total length	MA	MM	IND	GF (%)	# genes
Single-cell *E. coli*									
A5	14399	745	101584	4441145	8	12.01	**0.17**	89.880	3444
ABySS	68534	**179**	178720	4345617	6	3.32	1.68	88.268	3704
CLC	32506	503	113285	4656964	2	5.53	1.42	92.291	3768
EULER-SR	26662	429	140518	4248713	17	10.87	35.67	84.898	3416
Ray	45448	361	210820	4379139	17	6.29	2.83	88.372	3636
SOAPdenovo	1540	1166	51517	2958144	1	**1.87**	**0.11**	57.672	1766
Velvet	22648	261	132865	3501984	2	2.19	1.23	73.765	3080
E+V-SC	32051	344	132865	4540286	2	2.35	0.73	91.744	3771
IDBA-UD contigs	98306	244	**284464**	**4814043**	8	5.09	0.27	**95.210**	**4045**
IDBA-UD scaffolds	**109057**	229	**284464**	**4813609**	8	5.14	0.77	**95.199**	**4052**
SPAdes2.5 contigs	**109825**	250	268493	**4799522**	1	2.97	0.50	**94.917**	**4032**
SPAdes2.5 scaffolds	**112342**	240	268493	**4801710**	1	3.54	1.11	**94.939**	**4039**

The pros and cons of SCG (*cf*. metagenomics)

- Pros
  - Obtaining genomic information of individual cell ("Who is doing what?", within population variation)
  - Assessment of rare cell types via targeted sorting
  - Low sequencing costs, low computational demands
- Cons
  - Current throughput is limited
  - High contamination risk
  - Formation of chimeric artefacts
  - (Incomplete genomes)

SCG example study

ARTICLE

OPEN

doi:10.1038/nature12352

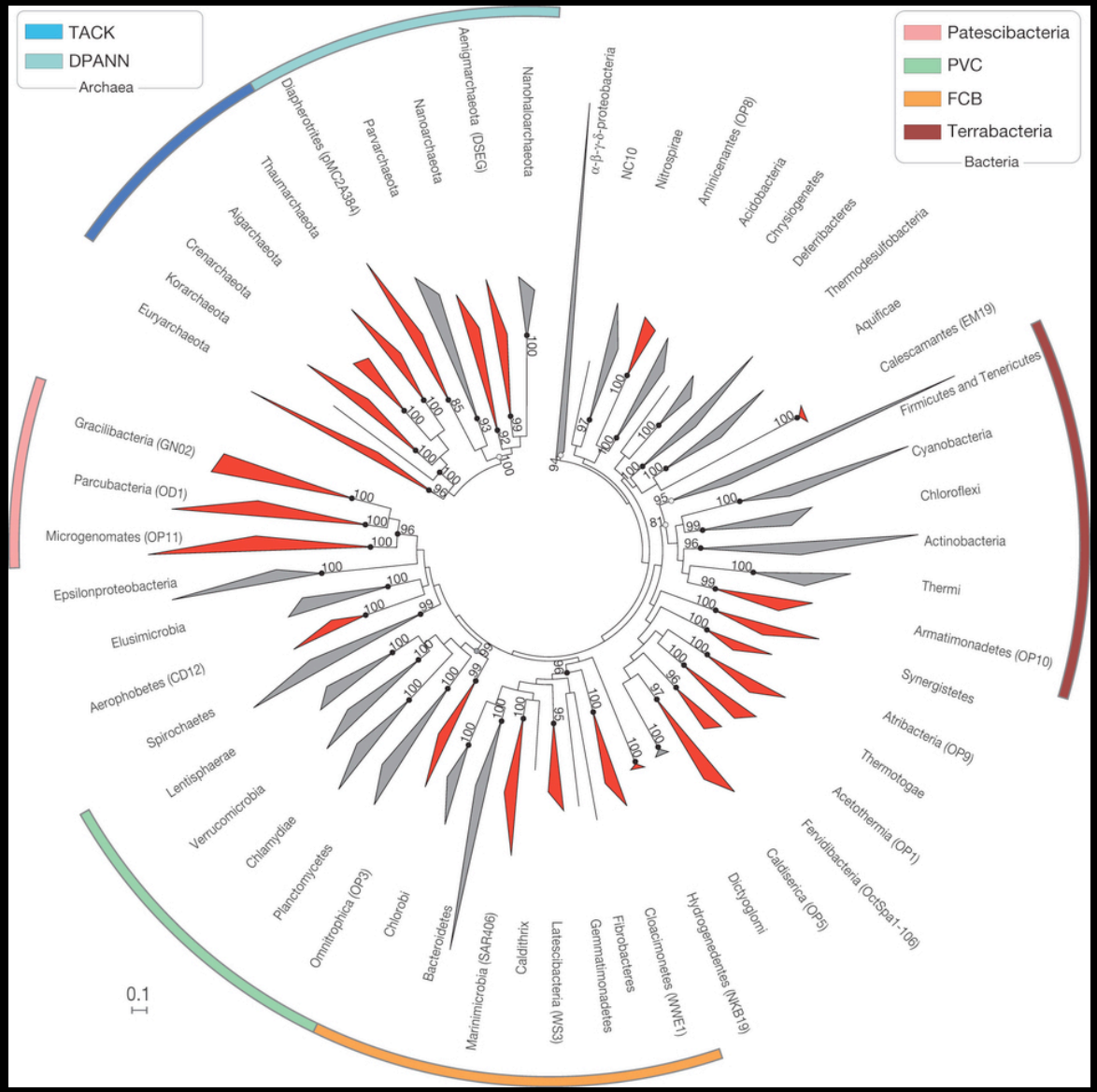
Insights into the phylogeny and coding potential of microbial dark matter

Christian Rinke[1], Patrick Schwientek[1], Alexander Sczyrba[1,2], Natalia N. Ivanova[1], Iain J. Anderson[1]‡, Jan-Fang Cheng[1], Aaron Darling[3,4], Stephanie Malfatti[1], Brandon K. Swan[5], Esther A. Gies[6], Jeremy A. Dodsworth[7], Brian P. Hedlund[7], George Tsiamis[8], Stefan M. Sievert[9], Wen-Tso Liu[10], Jonathan A. Eisen[3], Steven J. Hallam[6], Nikos C. Kyrpides[1], Ramunas Stepanauskas[5], Edward M. Rubin[1], Philip Hugenholtz[11] & Tanja Woyke[1]

Genome sequencing enhances our understanding of the biological world by providing blueprints for the evolutionary and functional diversity that shapes the biosphere. However, microbial genomes that are currently available are of limited phylogenetic breadth, owing to our historical inability to cultivate most microorganisms in the laboratory. We apply single-cell genomics to target and sequence 201 uncultivated archaeal and bacterial cells from nine diverse habitats belonging to 29 major mostly uncharted branches of the tree of life, so-called 'microbial dark matter'. With this additional genomic information, we are able to resolve many intra- and inter-phylum-level relationships and to propose two new superphyla. We uncover unexpected metabolic features that extend our understanding of biology and challenge established boundaries between the three domains of life. These include a novel amino acid use for the opal stop codon, an archaeal-type purine synthesis in Bacteria and complete sigma factors in Archaea similar to those in Bacteria. The single-cell genomes also served to phylogenetically anchor up to 20% of metagenomic reads in some habitats, facilitating organism-level interpretation of ecosystem function. This study greatly expands the genomic representation of the tree of life and provides a systematic step towards a better understanding of biological evolution on our planet.

Rinke et al, Nature 499, 431–437

SCG example study



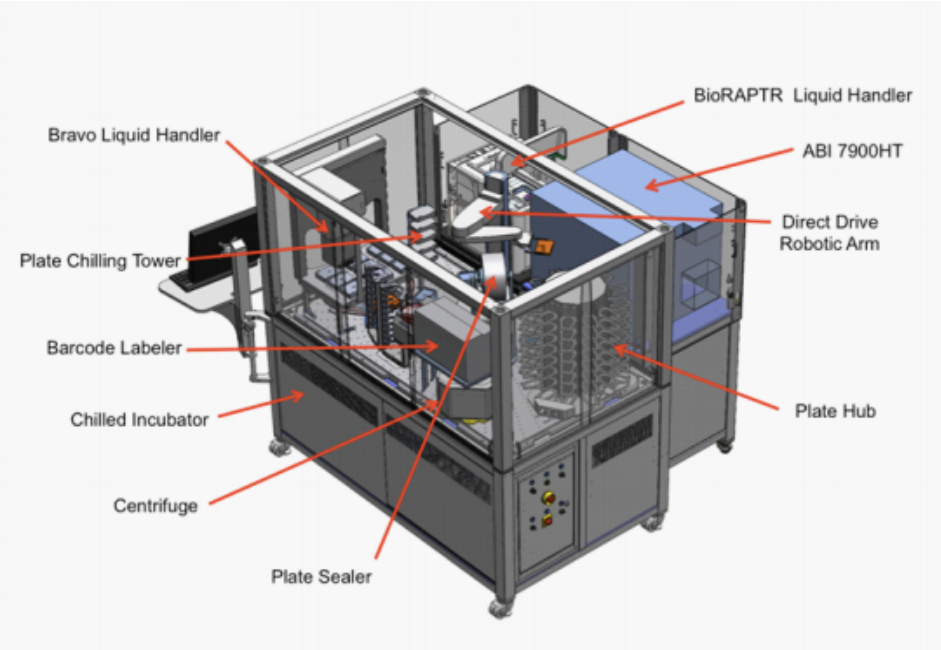
Highlights

- Sequencing of 201 SAGs across the Tree of Life
- Several representatives of proposed novel bacterial and archaeal phyla
- Several newly proposed superphyla:
  - Terrabacteria (B)
  - Patescibacteria (B)
  - DPANN (A)
- Several cases of HGT across the Domains of life
- Bacteria with alternative genetic code (opal stop codon)

Rinke et al, Nature 499, 431–437

Future directions of SCG

- Fully automated-based approaches

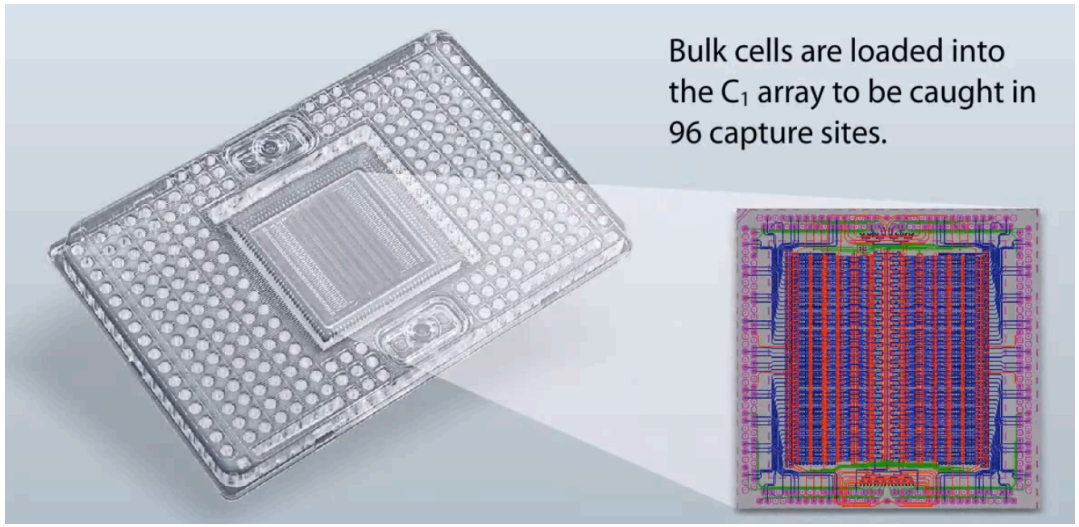


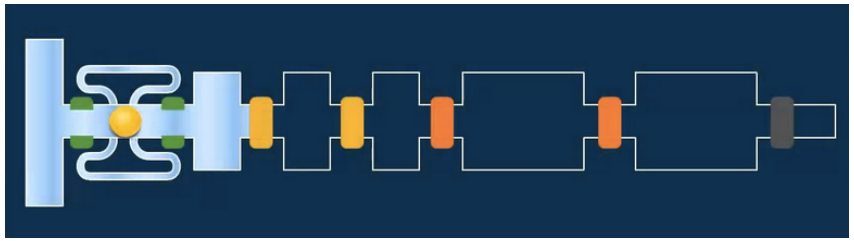
JCVI

- Custom-made SCG robotics setup
- "Normal" FACS-based single cell sorting
- All steps fully automated, except FACS and qPCR screen
- Throughput 5000 SAGs/week

Future directions of SCG

- Microfluidics-based approaches

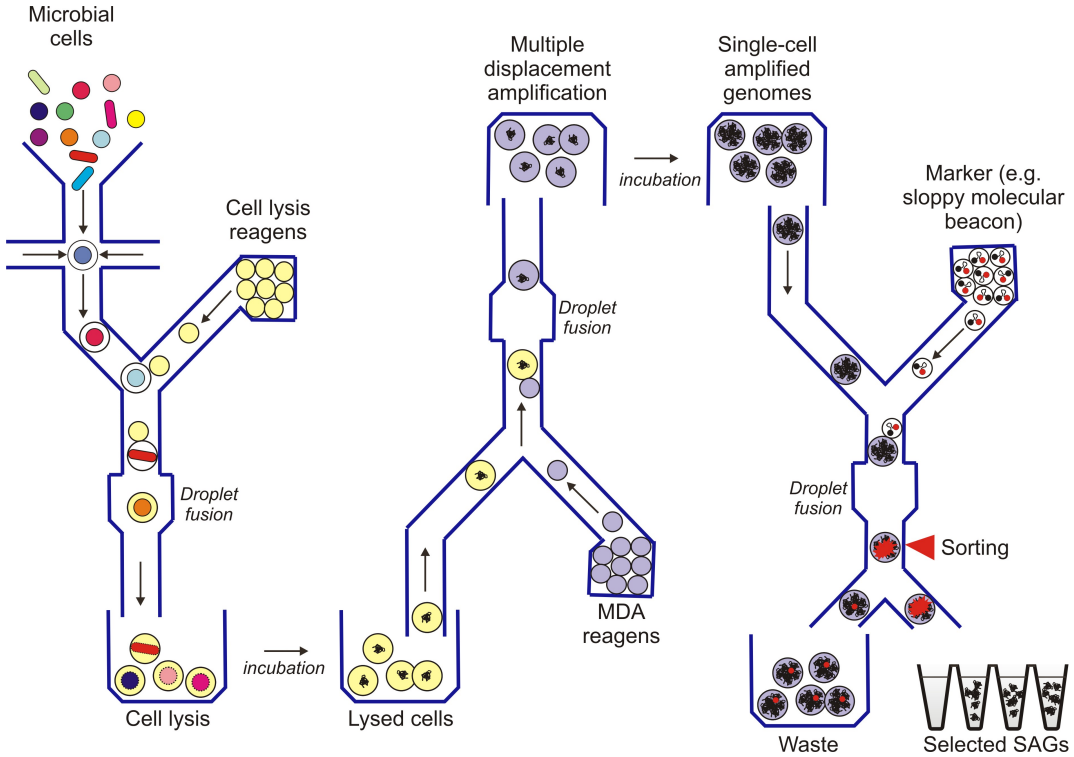




- Fluidigm C1: Integrated solution for single cell WGS and RNA-seq
- On-chip cell capture, lysis, (cDNA synthesis) and amplification.
- Limited cell sizes (diameter ~5-25 um)
- Up to 96 wells per chip

Future directions of SCG

- Microfluidics-based approaches



Picoliter volume droplets (contamination less of an issue)
Droplet based cell sorting, lysis, amplification and screening
Throughput: Millions of droplets per experiment

SciLifeLab SiCell

Platform for Single Cell Genomics @ Uppsala University

Stefan Bertilsson
Director

Thijs Ettema
Platform manager

Claudia Bergin
Scientist

Anna-Maria Divne
Scientist

- Open as a service-based platform
- National SciLife Platform in 2015