Il sequenziamento genico e le nuove biotecnologie in Microbiologia e Virologia

Olivier Jousson

Responsabile Laboratorio Microbial Genomics, CIBIO, Università di Trento
Coordinatore Corso di Laurea in Scienze e Tecnologie Biomolecolari, Università di Trento

CIBIO

UNIVERSITÀ DEGLI STUDI DI TRENTO

Many molecular diagnostic techniques and targets…

Table 1. Main molecular techniques and specific targets used for the identification of *C. neoformans* and *C. gattii*.

Technique	Targets	Advantages	Disadvantages	Reference
Identification				
Hybridization	Repetitive and polymorphic DNA	High sensitivity and specificity	Expensive and laborious	Spitzer and Spitzer 1992
PCR	ITS and 5.6S rRNA	High sensitivity and specificity; quickness and feasibility	Presence of contaminants in extraction and reaction phases	Paschoal et al. 2004
Nested PCR	ITS rDNA	High sensitivity and specificity	Presence of contaminants in reaction	Rappelli et al. 1998
Multiplex PCR	Serotype specific	Amplification of 2 or more loci in only 1 reaction; small amounts of extracted DNA	Interference due to the presence of polymorphism; reagent competition; possible nonspecific products	Leal et al. 2008
Real-time PCR	18S/28 rRNA	High sensitivity and specificity; detection of gene expression levels; quickness	Interference in the final analyze due to contamination with genomic DNA; requires technical ability and support; expensive	Levy et al. 2008
Typing				
PCR fingerprinting	Microsatellite $(GACA)_4$	Previous knowledge of target sequences is not necessary; use of short primers; polymorphism detection	Standardization of the technique according to the conditions of each laboratory	Hafner et al. 2005
RAPD	Minisatellite (M13)	Previous knowledge of target sequences is not necessary; use of short primers; polymorphism detection	Standardization of the technique according to the conditions of each laboratory	Capoor et al. 2008
PCR–RFLP	Urease	Specificity; hybridization phase can be skipped	Decreased sensitivity in case of isolated mutations	Meyer et al. 2003
AFLP	Capsule	High sensitivity and specificity; high resolution and sampling power; detection of genetic variability	Greater number of phases, larger number of reagents, expensive	Enache-Angoulvant et al. 2007
MLST	IGS, capsule, laccase, u ease, phospholipase	Reproducible and unambiguous; completely automated analysis, simultaneous analysis of multiple loci	Limitations in the differentiation of strains when genes are very conserved	Chen et al. 2008

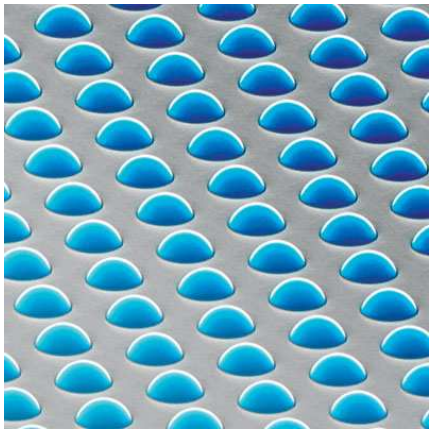
Sidrim et al. Can J Microbiol, 2010

Pros e Cons of molecular diagnostic methods

DILEMMAS

Not highly sensitive Low contamination risk	◢	Highly sensitive High contamination risk
Not highly specific Wide identification range	◢	Highly specific False negatives
Based on single or few markers = incomplete view Low cost and fast	◢	Genome-wide or Metagenome-wide identification High cost, time-consuming

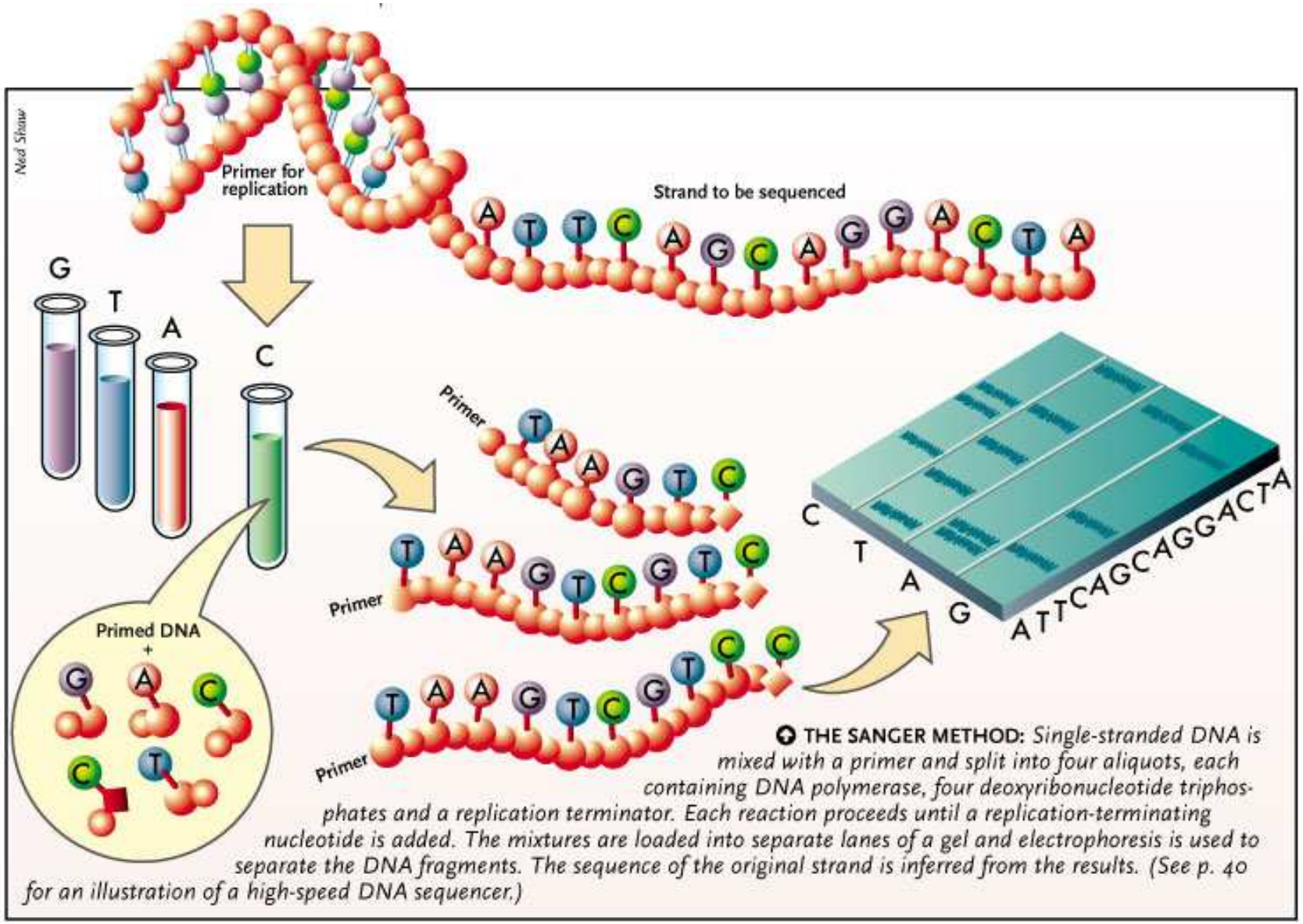
… the most revolutionary effect of *Next Generation Sequencing* is likely to be creation of a novel sequence-based, culture-independent **diagnostic microbiology** that incorporates microbial community profiling, metagenomics and single-cell genomics.



We should prepare for the coming 'technological singularity' in sequencing, when this technology becomes so fast and so cheap that it threatens to out-compete existing diagnostic and typing methods in microbiology.

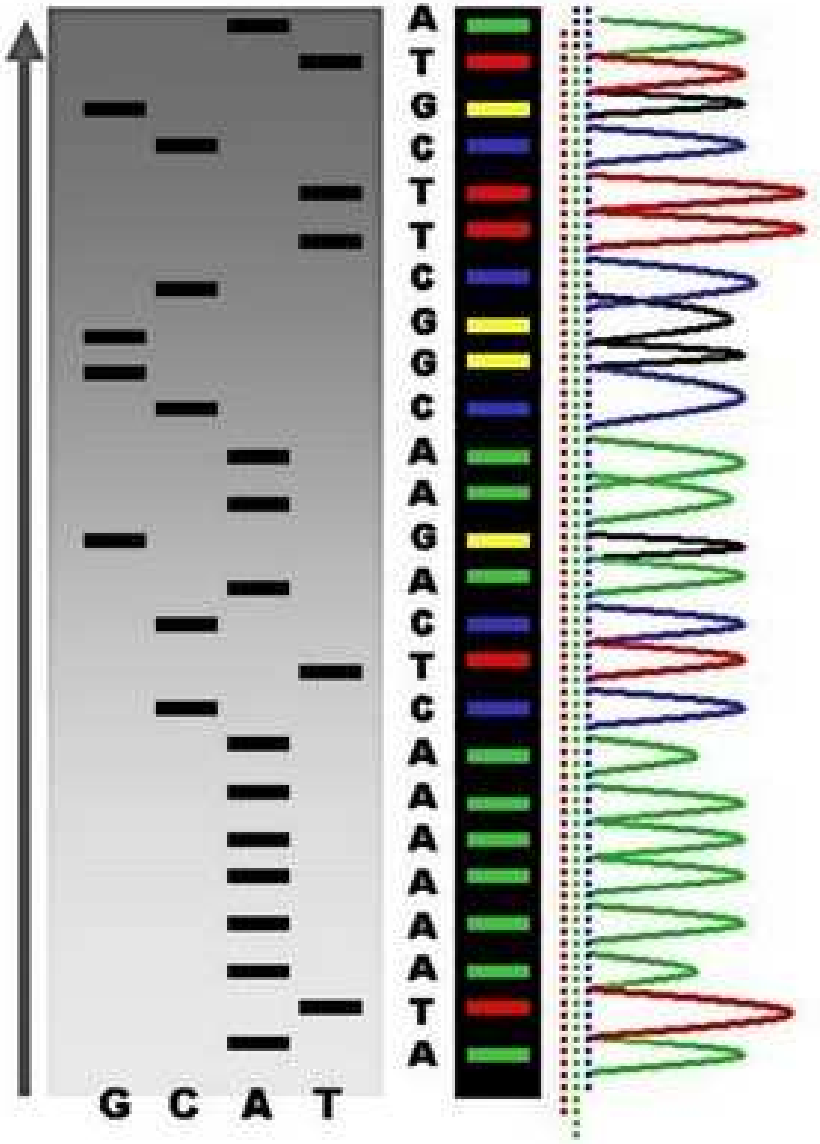
Pallen et al., Curr Opin Microbiol, 2010

Sequencing methods: Sanger sequencing



THE SANGER METHOD: *Single-stranded DNA is mixed with a primer and split into four aliquots, each containing DNA polymerase, four deoxyribonucleotide triphosphates and a replication terminator. Each reaction proceeds until a replication-terminating nucleotide is added. The mixtures are loaded into separate lanes of a gel and electrophoresis is used to separate the DNA fragments. The sequence of the original strand is inferred from the results. (See p. 40 for an illustration of a high-speed DNA sequencer.)*

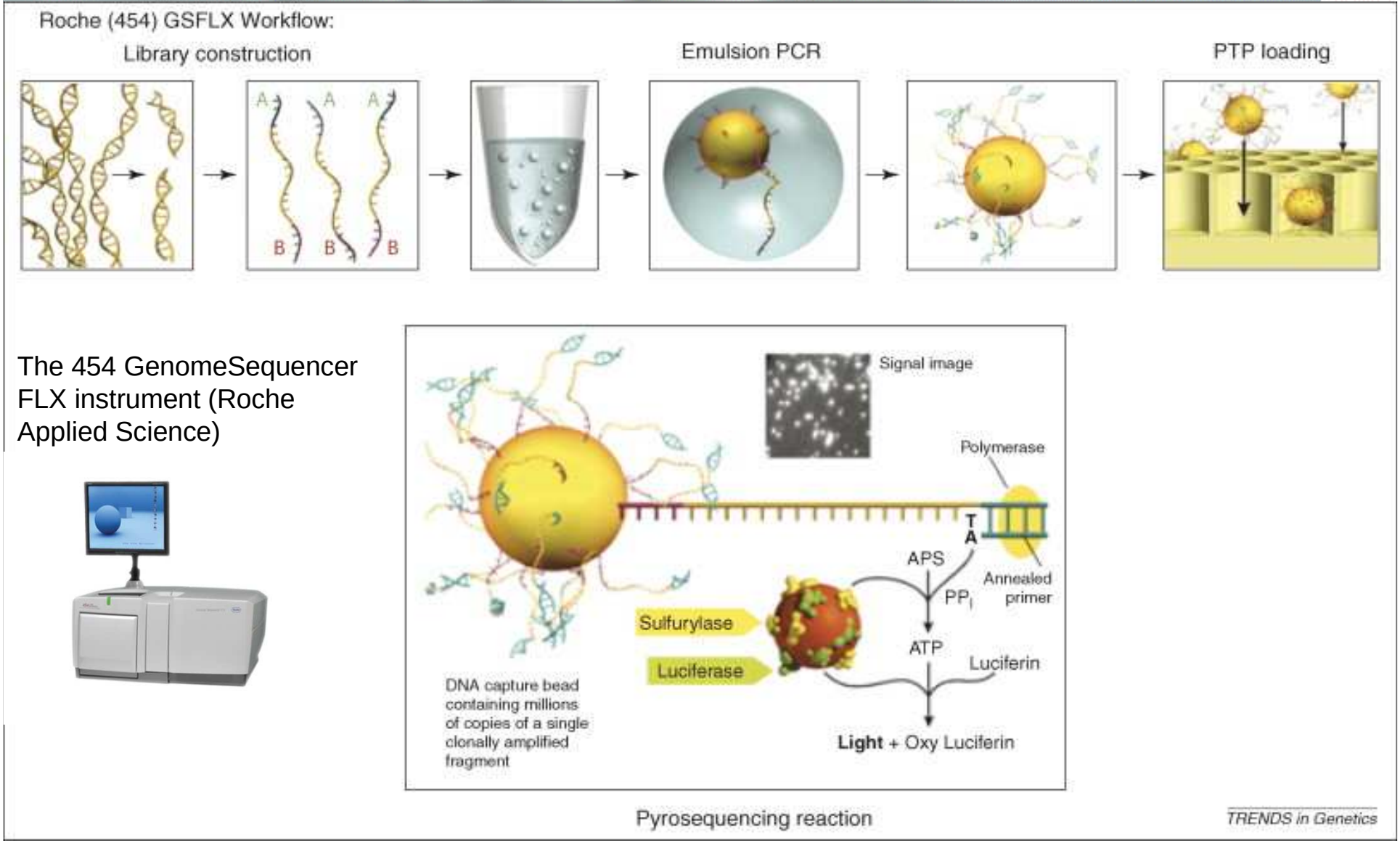
Sequencing methods: Sanger sequencing



Throughput: 1 sample / lane;
read lenght about 1 kb

Radioactive versus fluorescently labelled dideoxynucleotides

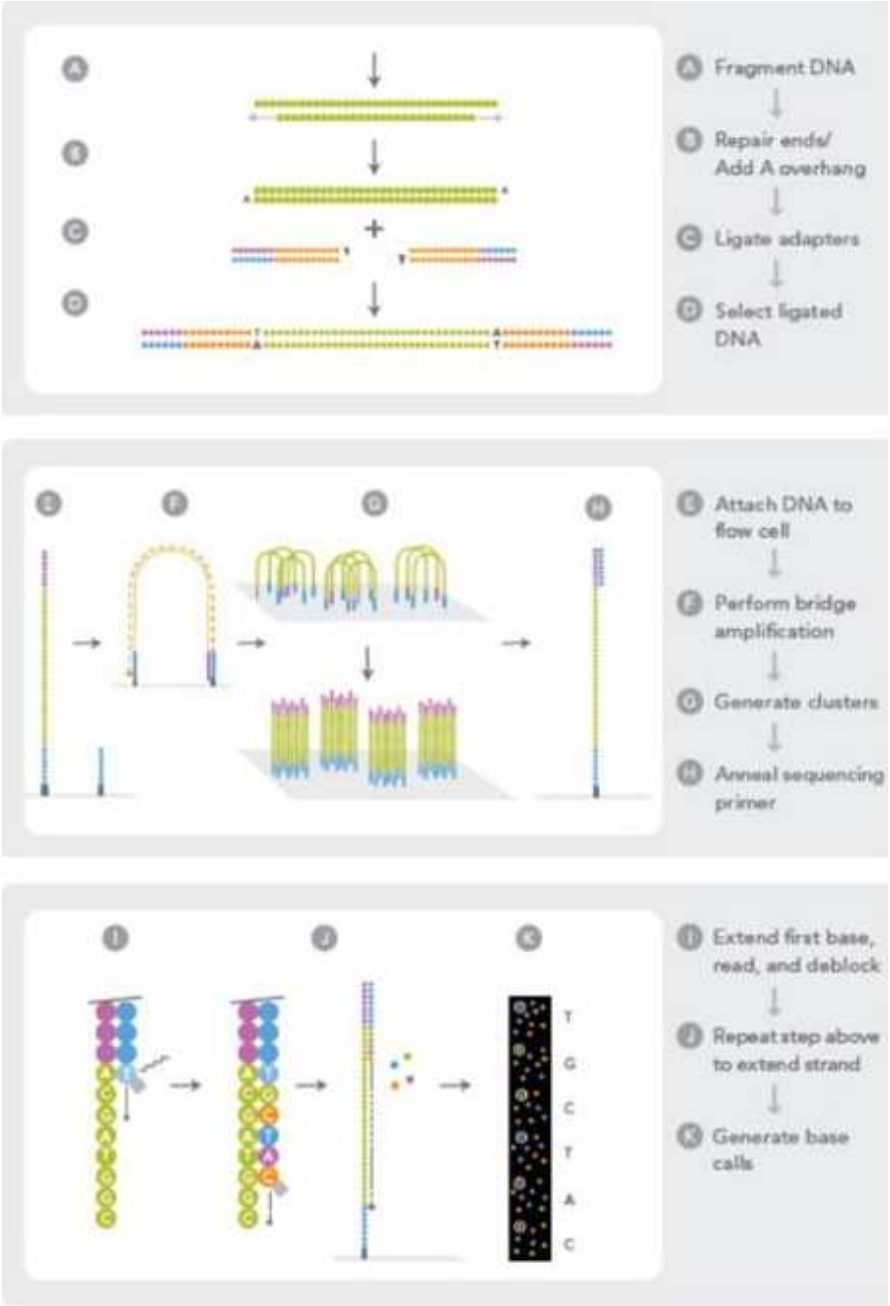
Next-generation sequencing technologies



The 454 GenomeSequencer FLX instrument (Roche Applied Science)

Library construction (I) ligates 454-specific adapters to DNA fragments (indicated as A and B) and couples amplification beads with DNA in an emulsion PCR to amplify fragments before sequencing (II). The beads are loaded into the picotiter plate (III). (B) Schematic illustration of the pyrosequencing reaction which occurs on nucleotide incorporation to report sequencing-by-synthesis.

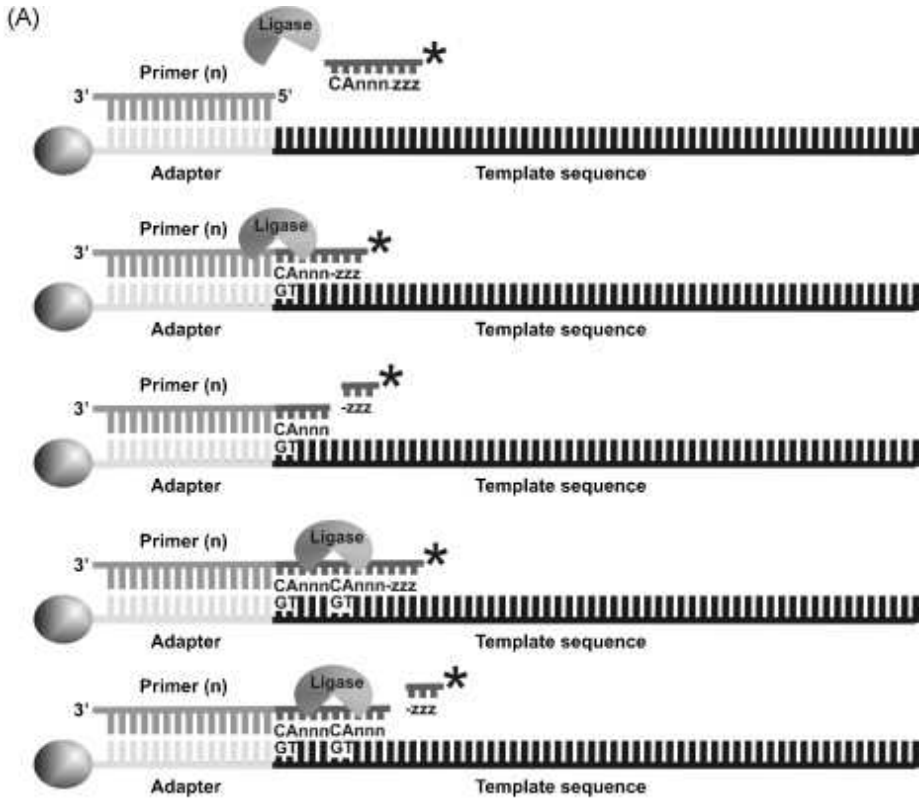
Next-generation sequencing technologies

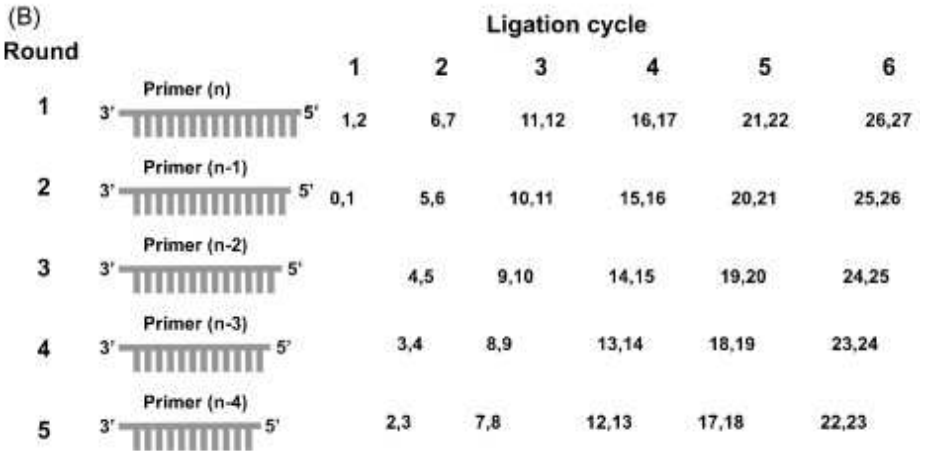


The Illumina (Solexa) Genome Analyzer

Outline of the Illumina Genome Analyzer workflow. Similar fragmentation and adapter ligation steps take place, before applying the library onto the solid surface of a flow cell. Attached DNA fragments form ‘bridge’ molecules which are subsequently amplified via an isothermal amplification process, leading to a cluster of identical fragments that are subsequently denatured for sequencing primer annealing. Amplified DNA fragments are subjected to sequencing-by-synthesis using labelled nucleotides.

Next-generation sequencing technologies





The Applied Biosystems ABI SOLiD system

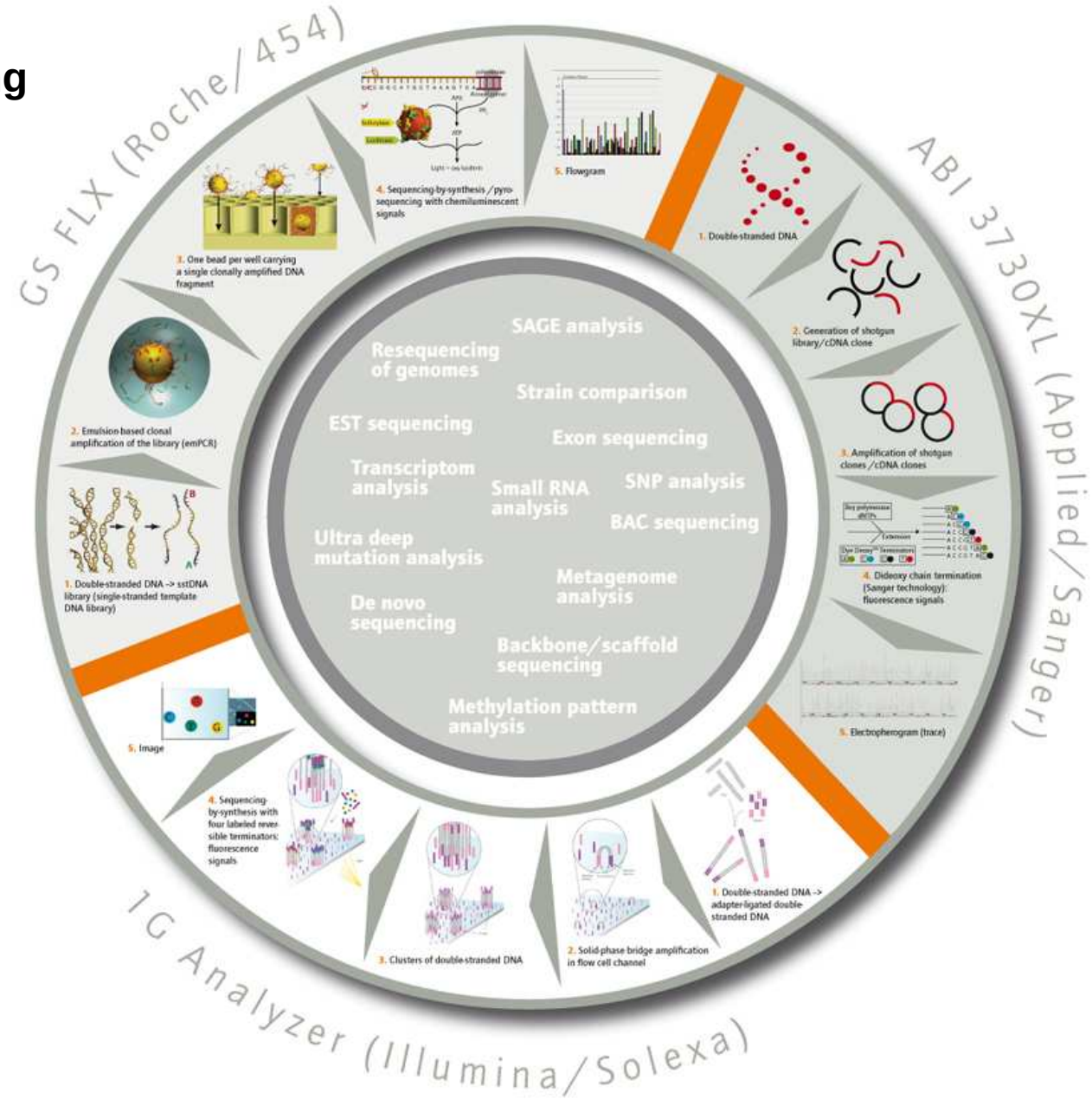
Sequencing-by-ligation, using the SOLiD DNA sequencing platform. **(A)** Primers hybridise to the P1 adapter within the library template. A set of four fluorescence-labelled di-base probes competes for ligation to the sequencing primer. These probes have partly degenerated DNA sequence (indicated by *n* and *z*) and for simplicity only one probe is shown (labelling is denoted by asterisk). Specificity of the di-base probe is achieved by interrogating the first and second base in each ligation reaction (CA in this case for the complementary strand). Following ligation, the fluorescent label is enzymatically removed together with the three last bases of the octamer. **(B)** Sequence determination by the SOLiD DNA sequencing platform is performed in multiple ligation cycles, using different primers, each one shorter from the previous one by a single base. The number of ligation cycles (six for this example) determines the eventual read length, whilst for each sequence tag, six rounds of primer reset occur [from primer (*n*) to primer (*n* − 4)]. The dinucleotide positions on the template sequence that are interrogated each time, are depicted underneath each ligation cycle and are separated by 5-bp from the dinucleotide position interrogated in the subsequent ligation cycle.

Comparing metrics and performance of next-generation DNA sequencers

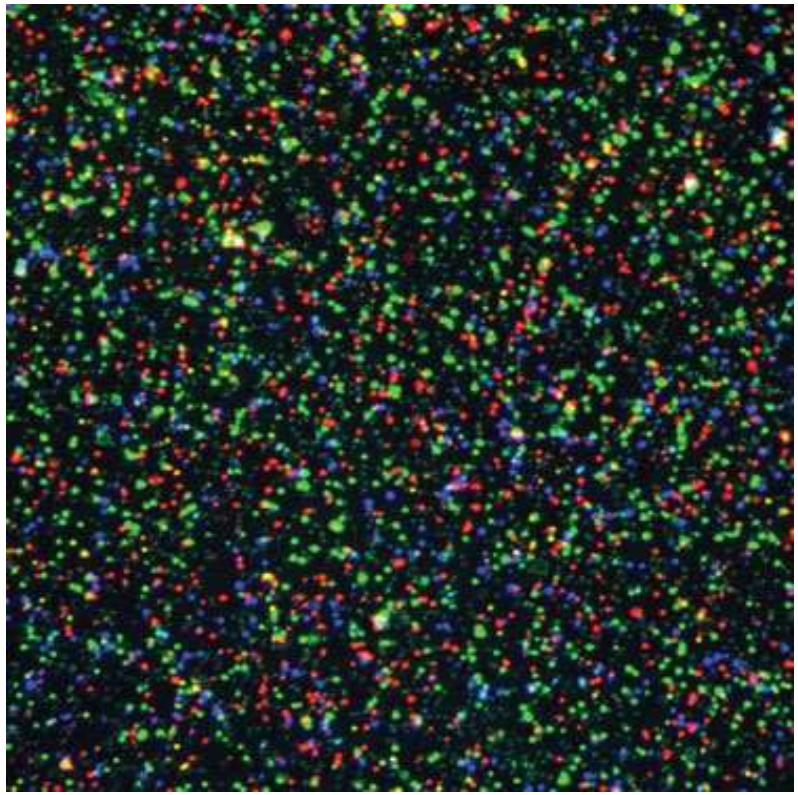
	Platform Roche(454)	Illumina	SOLiD
Sequencing chemistry	Pyrosequencing	Polymerase-based sequencing-by-synthesis	Ligation-based sequencing
Amplification approach	Emulsion PCR	Bridge amplification	Emulsion PCR
Mb/run	100 -500 Mb	1300 Mb – 3000 Mb	3000 Mb -10000 Mb
Time/run	7 h	4 days	5 days
Read length	250 - 500 bp	32–40 bp	35 bp
Cost per run	$8439	$8950	$17 447
Cost per Mb	$84.39	$5.97	$5.81

Next-generation sequencing applications

- SAGE analysis
- Strain comparison
- Exon sequencing
- SNP analysis
- Small RNA analysis
- BAC sequencing
- Metagenome analysis
- Methylation pattern analysis
- De novo sequencing
- Ultra deep mutation analysis
- Transcriptome analysis
- EST sequencing
- Resequencing of genomes

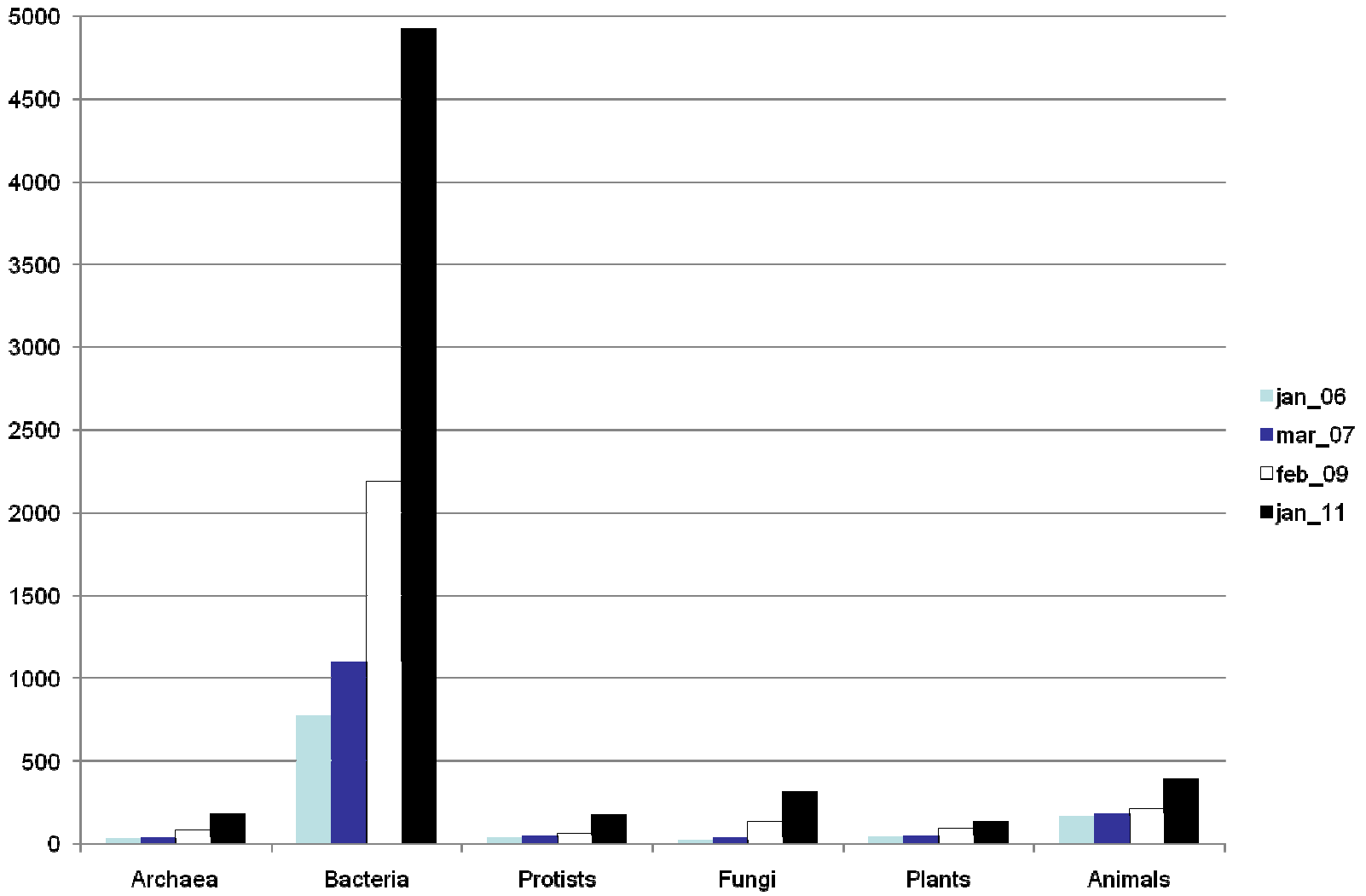


The Data Problem



"If the data problem is not addressed, ABI's SOLiD, 454's GS FLX, Illumina's GAII or any of the other deep sequencing platforms will be destined to sit in their air-conditioned rooms like a Stradivarius without a bow."

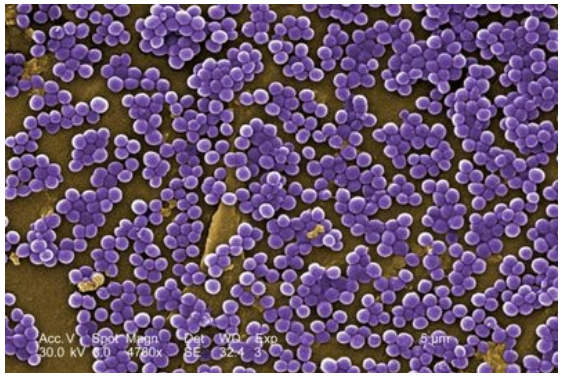
Sequenced genomes



Evolution of MRSA During Hospital Transmission and Intercontinental Spread

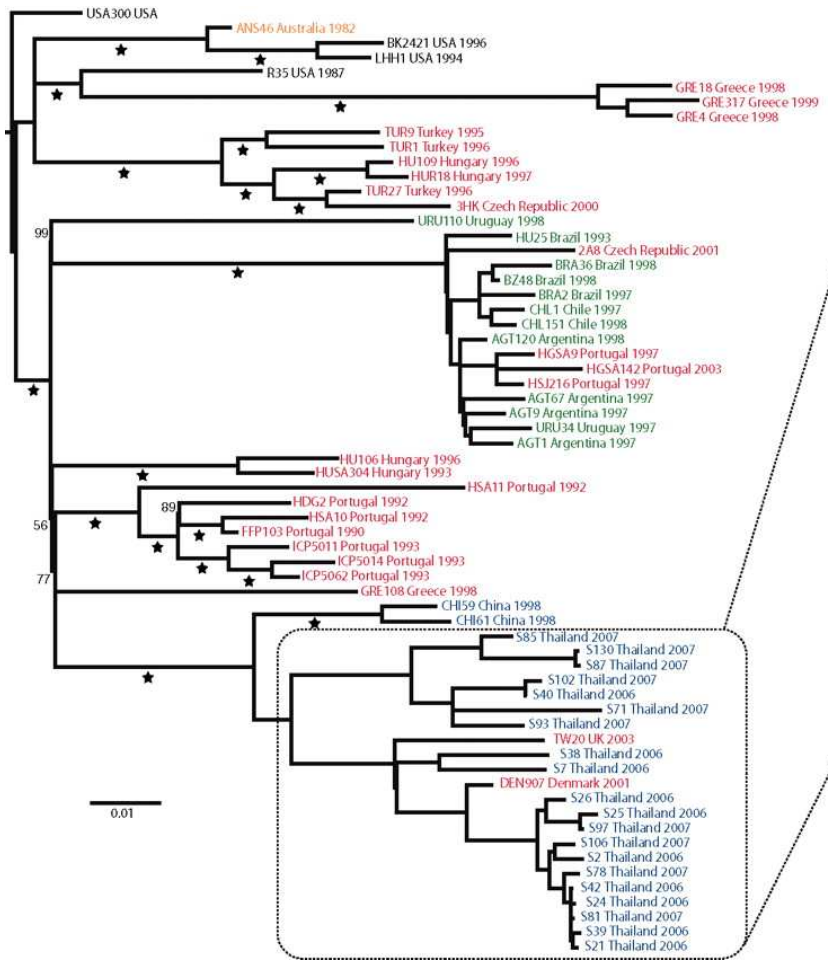
-a high-throughput genomics approach that provides a high-resolution view of the epidemiology and microevolution of a dominant strain of methicillin-resistant *Staphylococcus aureus* (MRSA)

-defines the global geographic structure within the lineage, intercontinental transmission through four decades, and the potential to trace person-to-person transmission within a hospital environment

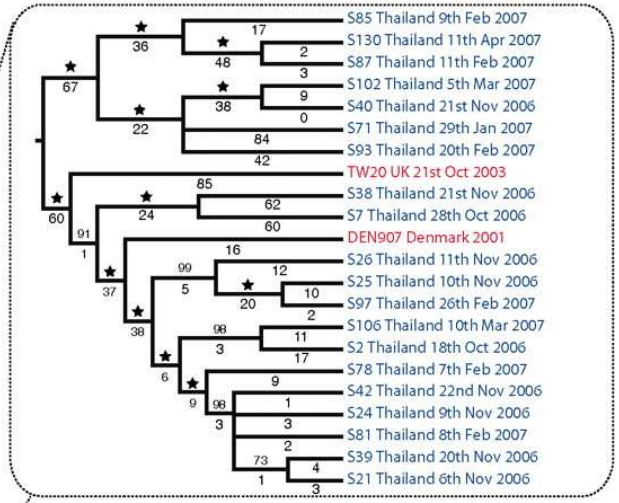


Harris et al, Science, 2010

Evolution of MRSA During Hospital Transmission and Intercontinental Spread



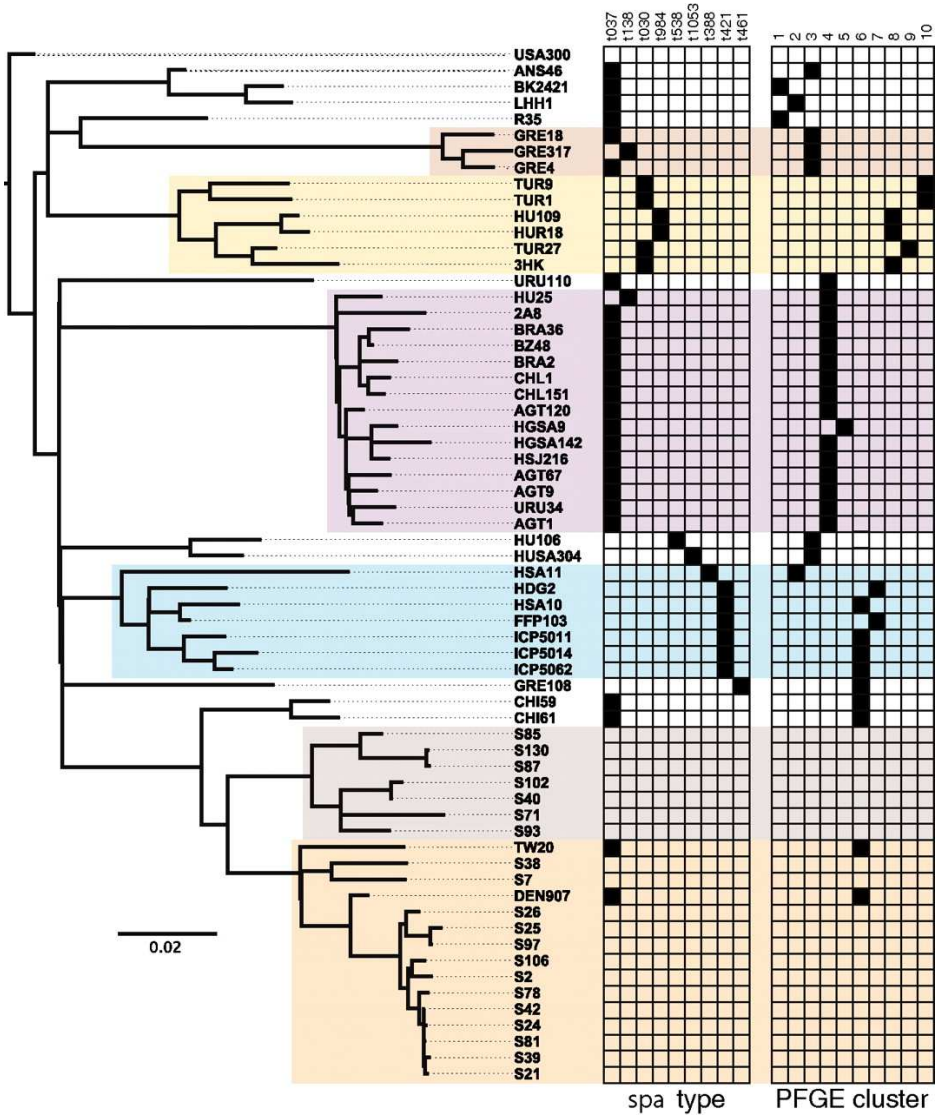
-Barcode sequencing identified 6714 SNPs (Single Nucleotide Polymorphisms) in two samples of 63 and 20 isolates



Phylogenetic evidence for intercontinental spread and hospital transmission of ST239 isolates

Harris et al, Science, 2010

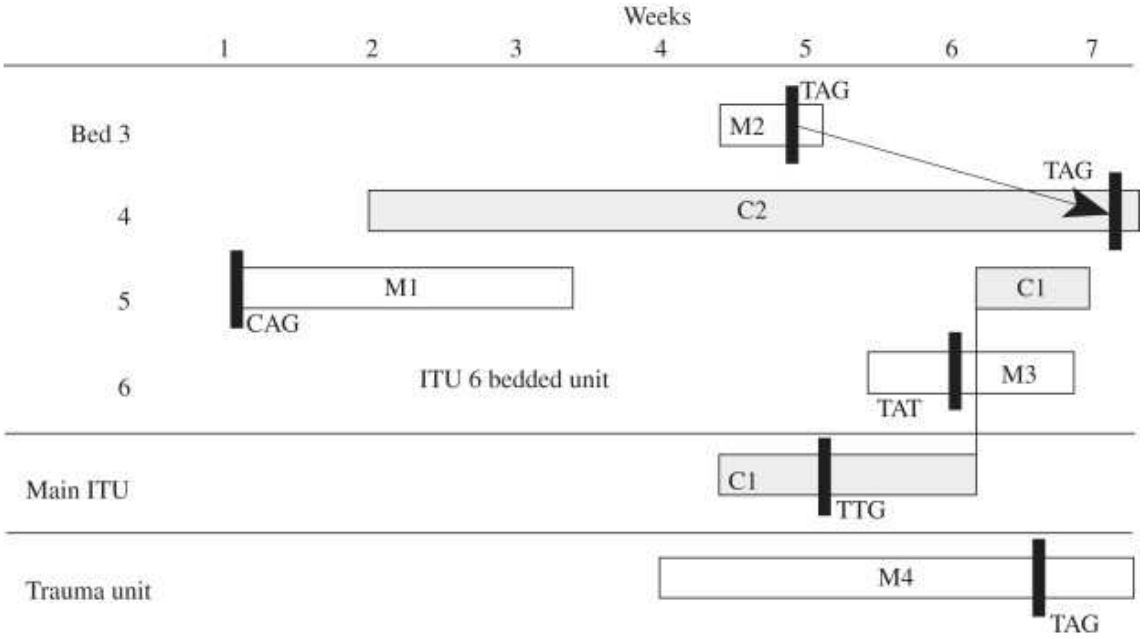
Evolution of MRSA During Hospital Transmission and Intercontinental Spread



Comparison of phylogeny with traditional typing techniques. There is a high consistency between techniques but a much higher resolution of genome-wide SNPs analysis.

Harris et al, Science, 2010

High-throughput whole-genome sequencing to dissect the epidemiology of *Acinetobacter baumannii* isolates from a hospital outbreak



Timeline showing bedspaces of individual patients while in the 6-bedded bay of the critical care unit. Vertical bars indicate a positive MDR-Aci isolate from the patient and their corresponding single nucleotide polymorphism genotype.

Epidemiological scenarios

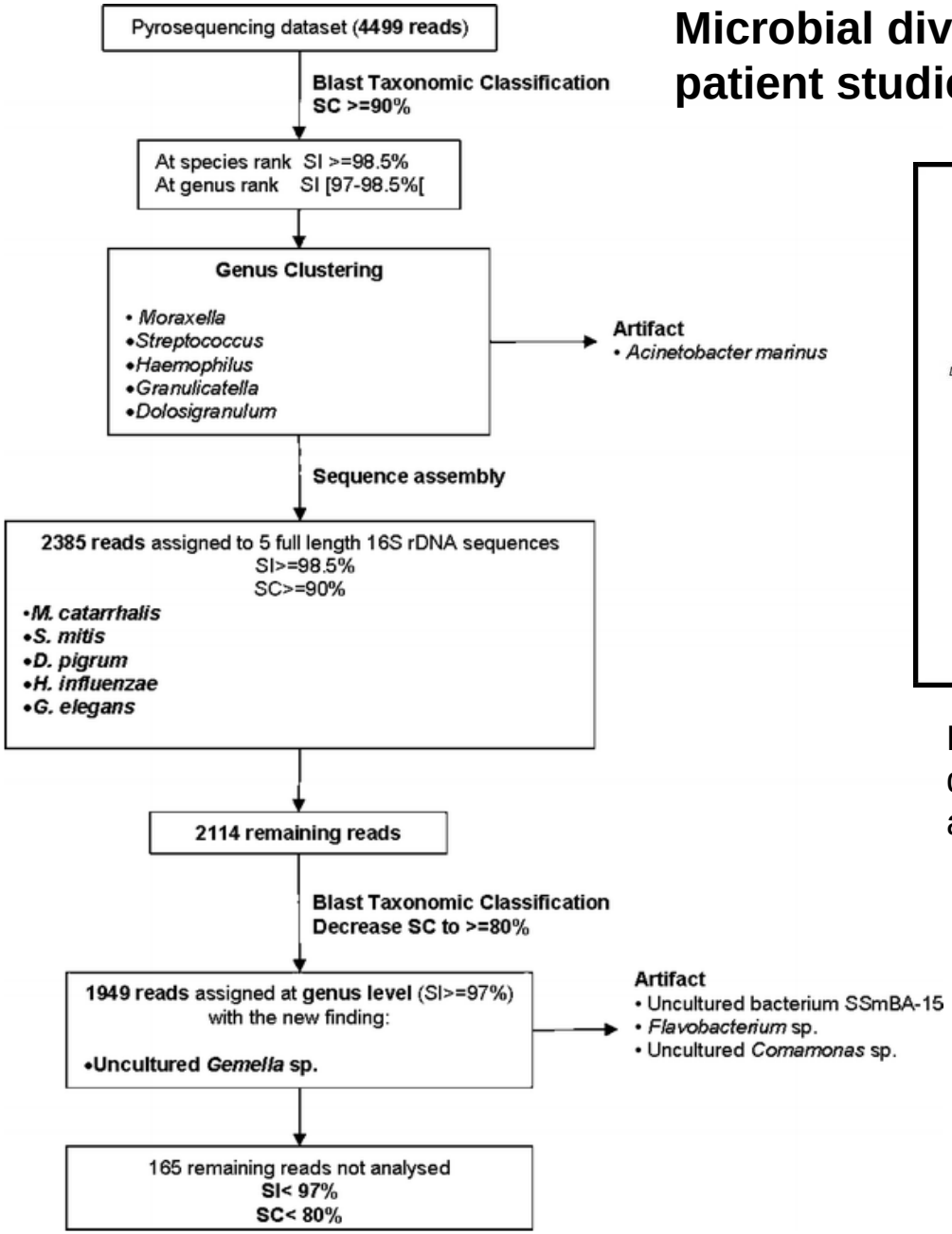
1. Transmission from M1. C2 was nursed in a bed next to M1 during week 2.
2. Transmission from M2. C2 was nursed in a bed next to M2 during week 4.
3. Transmission from M3, who occupied a nearby bed space in the six-bedded unit in the two weeks before MDR-Aci was first isolated from C2.
4. Transmission from C1, who occupied a nearby bed space in the six-bedded unit in the week before MDR-Aci was first isolated from C2.
5. Acquisition from an unknown source, such as an environmental reservoir or an unidentified patient or health worker.

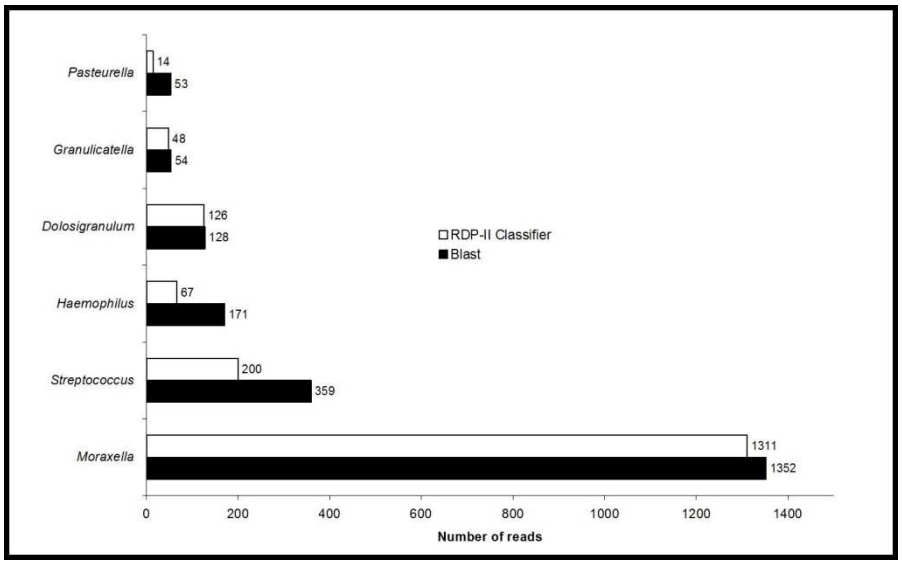
Shared care of military and civilian patients has resulted in transmission of multidrug-resistant *Acinetobacter baumannii* (MDR-Aci) from military casualties to civilians. In a recent hospital outbreak in Birmingham, six patients were colonised with MDR-Aci isolates indistinguishable using standard techniques. Whole-genome sequencing was used to identify single nucleotide polymorphisms in these isolates, allowing to discriminate between alternative epidemiological hypotheses in the setting.

Lewis et al, J Hosp Inf, 2010

Community profiling

Microbial diversity in the sputum of a cystic fibrosis patient studied with 16S rDNA pyrosequencing



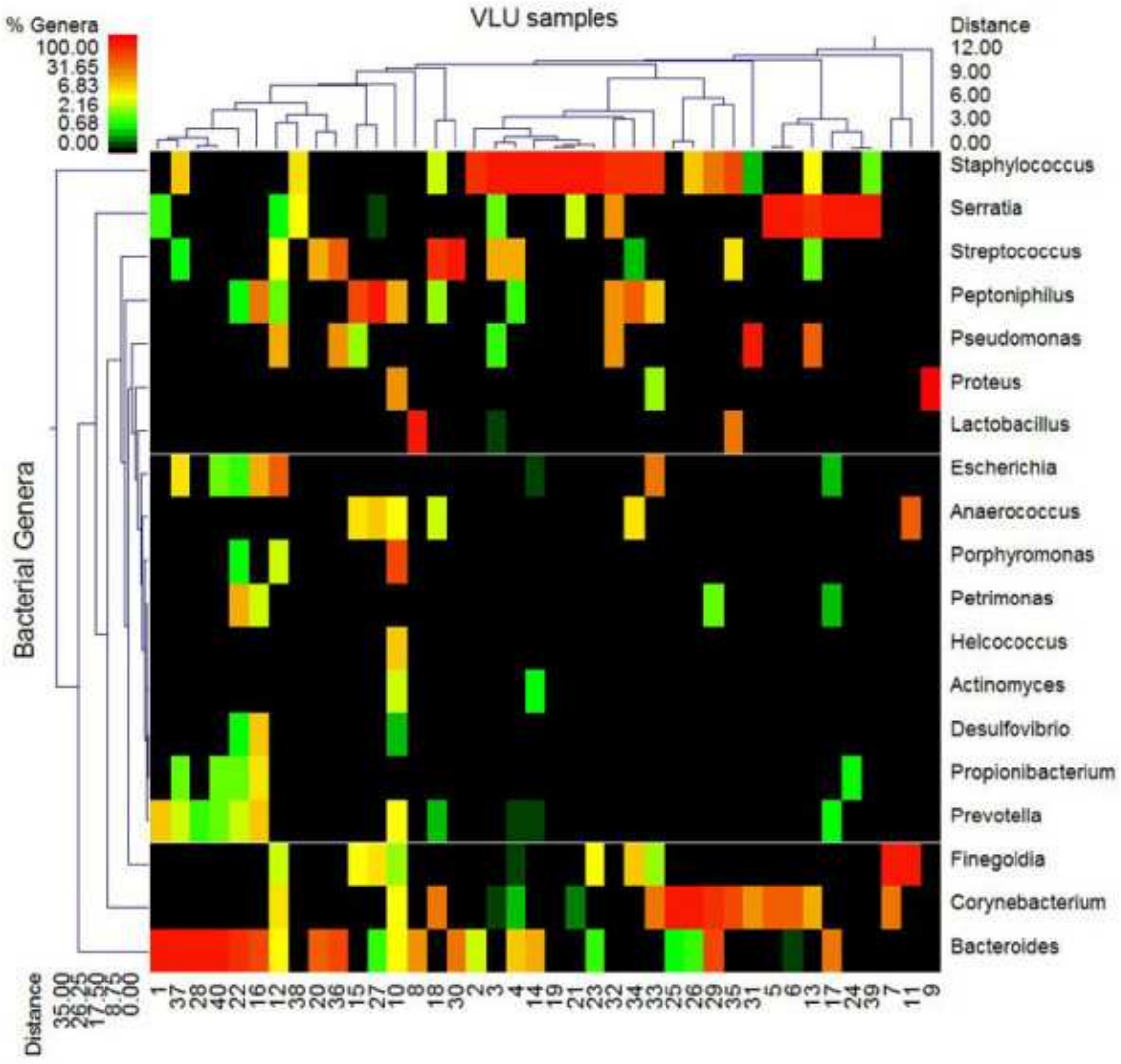


High number of reads provide information on qualitative (species diversity) and quantitative aspects (species abundance).

Armougom et al, Eur J Clin Microbiol Infect Dis, 2009

Community profiling

Evaluation of the bacterial diversity among and within individual venous leg ulcers using bacterial tag-encoded FLX and Titanium amplicon pyrosequencing and metagenomic approaches



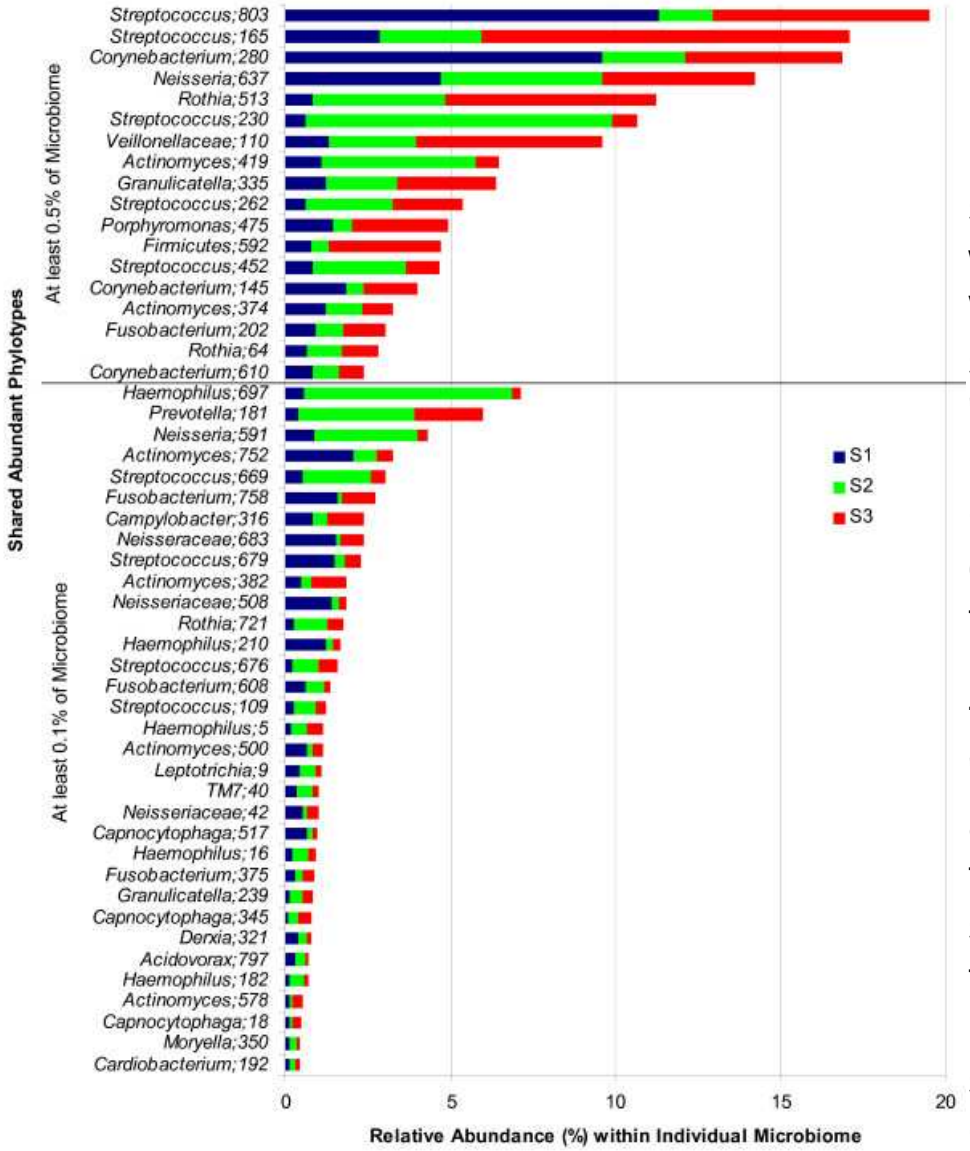
VLU infections are polymicrobial with no single bacterium colonizing the wounds (62 genera!). The most ubiquitous and predominant organisms include a previously uncharacterized bacteroidales, various anaerobes, *Staphylococcus, Corynebacterium, and Serratia.*

Topological analysis of VLU show some notable differences in bacterial populations across the surface of the wounds highlighting the importance of sampling techniques during diagnostics.

Metagenomics provide a preliminary indication that there may be protozoa, fungi and possibly an undescribed virus associated with these wounds.

Wolcott et al, BMC Microbiol, 2009

Community profiling



Shared abundant phylotypes in three oral microbiomes and their relative abundance

Defining the healthy "core microbiome" of oral microbial communities

Microbiomes from several intraoral niches (dental surfaces, cheek, hard palate, tongue and saliva) were sequenced in three healthy individuals. Within an individual oral cavity, over 3600 unique sequences and over 500 different OTUs or "species-level" phylotypes were found.

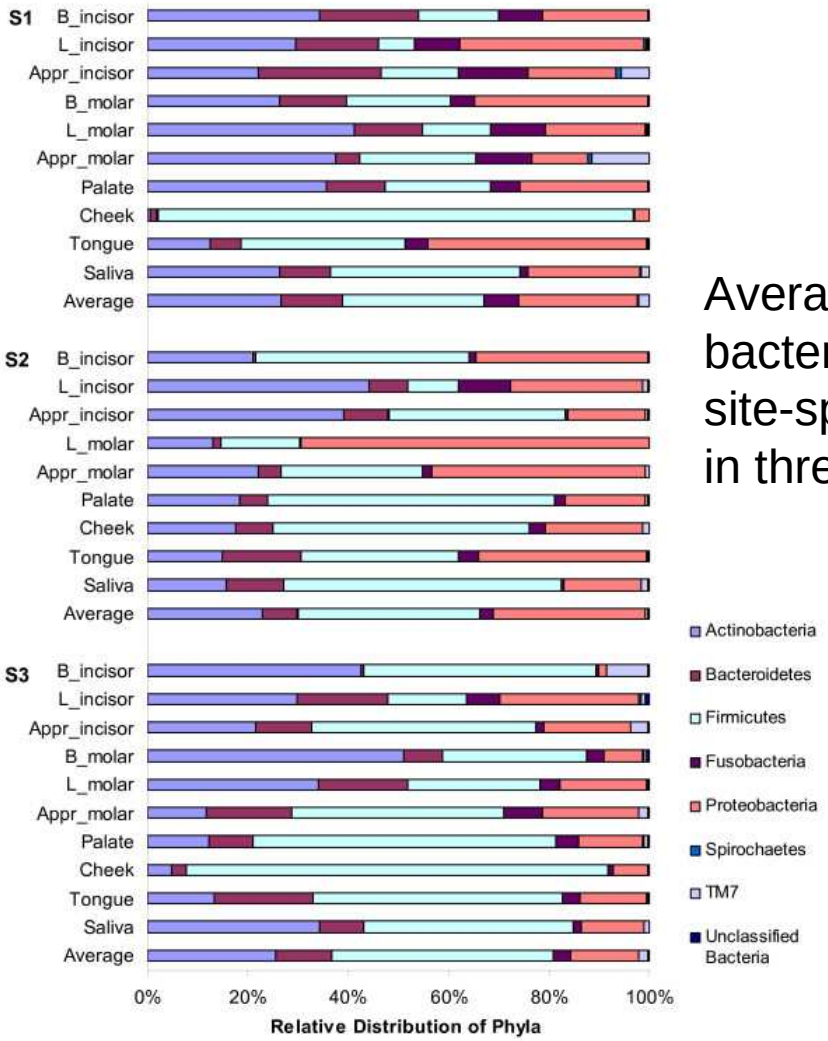
Cheek samples are the least diverse and the dental samples from approximal surfaces shows the highest diversity.

There was a large overlap in the higher taxa, "species-level" phylotypes and unique sequences among the three microbiomes: 84% of the higher taxa, 75% of the OTUs and 65% of the unique sequences were present in at least two of the three microbiomes. The three individuals shared 1660 of 6315 unique sequences. These 1660 sequences (the "core microbiome") contributed 66% of the reads.

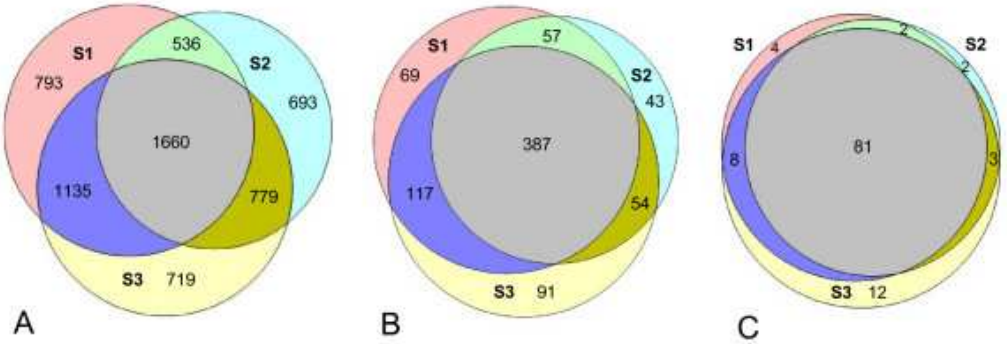
Zaura et al, BMC Microbiol, 2009

Community profiling

Defining the healthy "core microbiome" of oral microbial communities



Average and site-specific relative distribution of bacterial phyla in three individuals. Average and site-specific relative distribution of bacterial phyla in three individuals



The extent of overlap of oral microbiome between three individuals. The extent of overlap between subjects S1 (pink circle), S2 (light blue circle) and S3 (yellow circle) at the level of A) unique sequences, B) OTUs clustered at 3% difference and C) higher taxa (genus or more inclusive taxon).

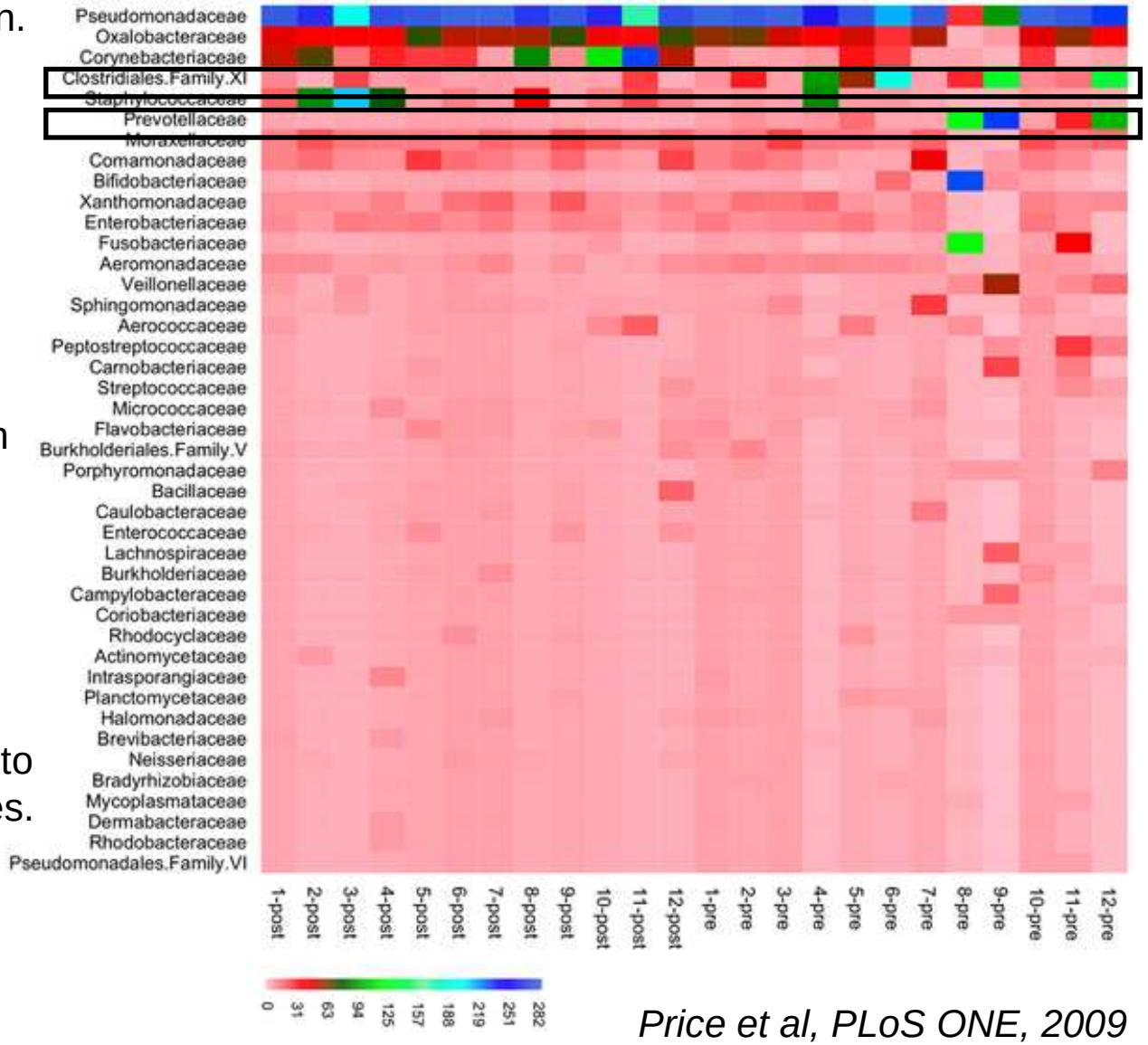
Zaura et al, BMC Microbiol, 2009

Community profiling

Circumcision is associated with significant reductions in HIV, HSV-2 and HPV infections among men. The penile (coronal sulci) microbiota in 12 HIV-negative Ugandan men was assessed before and after circumcision. 42 unique bacterial families were identified. Circumcision was associated with a significant change in the overall microbiota and with a significant decrease in putative anaerobic bacterial families.

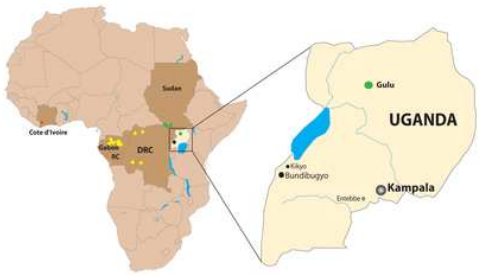
The anoxic microenvironment of the subpreputial space may support pro-inflammatory anaerobes that can activate Langerhans cells to present HIV to CD4 cells in draining lymph nodes. Thus, the reduction in putative anaerobic bacteria after circumcision may play a role in protection from HIV and other sexually transmitted diseases.

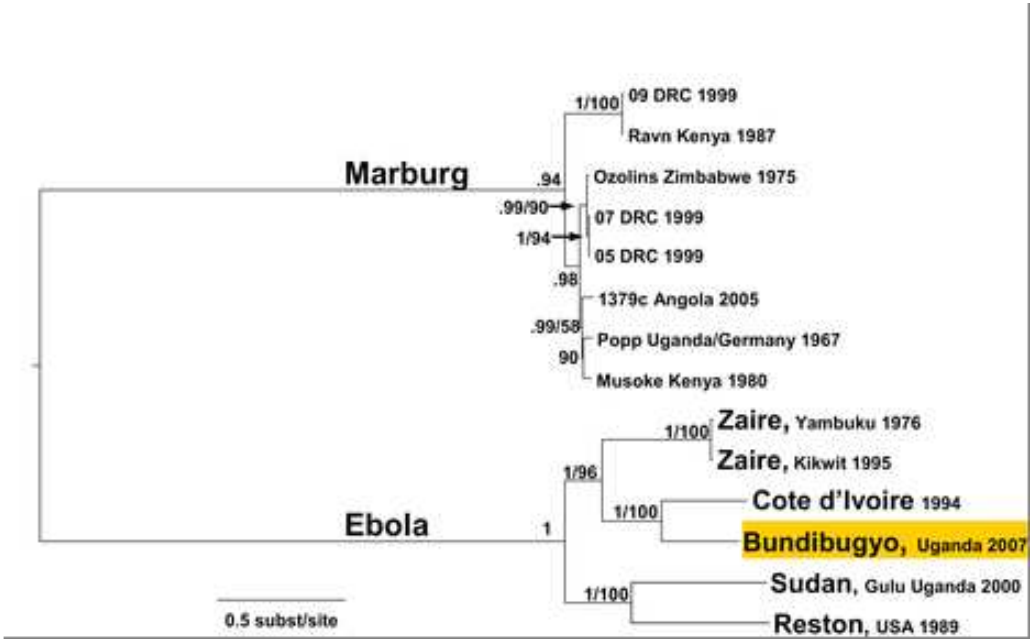
The Effects of Circumcision on the Penis Microbiome



Price et al, PLoS ONE, 2009

Newly Discovered Ebola Virus Associated with Hemorrhagic Fever Outbreak in Uganda





Phylogenetic relationships of representative filoviruses. Phylogenetic tree comparing full-length genomes of ebolavirus and Marburgvirus.

Over the past 30 years, Zaire and Sudan ebolaviruses have been responsible for large hemorrhagic fever (HF) outbreaks with case fatalities ranging from 53% to 90%.

Laboratory investigation of the initial 29 HF cases in Uganda (2007) using a pyrosequencing approach quickly identified this to be an Ebola HF outbreak associated with a newly discovered ebolavirus species (Bundibugyo ebolavirus).

Due to the sequence divergence (about 32%) of this new virus relative to all previously recognized ebolaviruses, these findings have important implications for design of future diagnostic assays to monitor Ebola HF disease in humans and animals, and ongoing efforts to develop effective antivirals and vaccines.

Towner et al, PLoS Pathogens, 2008

Pathogen discovery

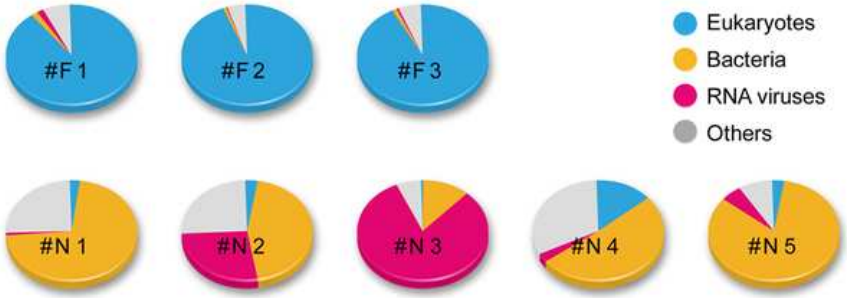
Direct Metagenomic Detection of Viral Pathogens in Nasal and Fecal Specimens Using an Unbiased High-Throughput Sequencing Approach

Unbiased high-throughput sequencing of these 3 nasal and 5 fecal samples yielded 15,298–32,335 reads.

In nasopharyngeal samples, 20–460 Flu-reads were detected, which was sufficient for subtype identification.

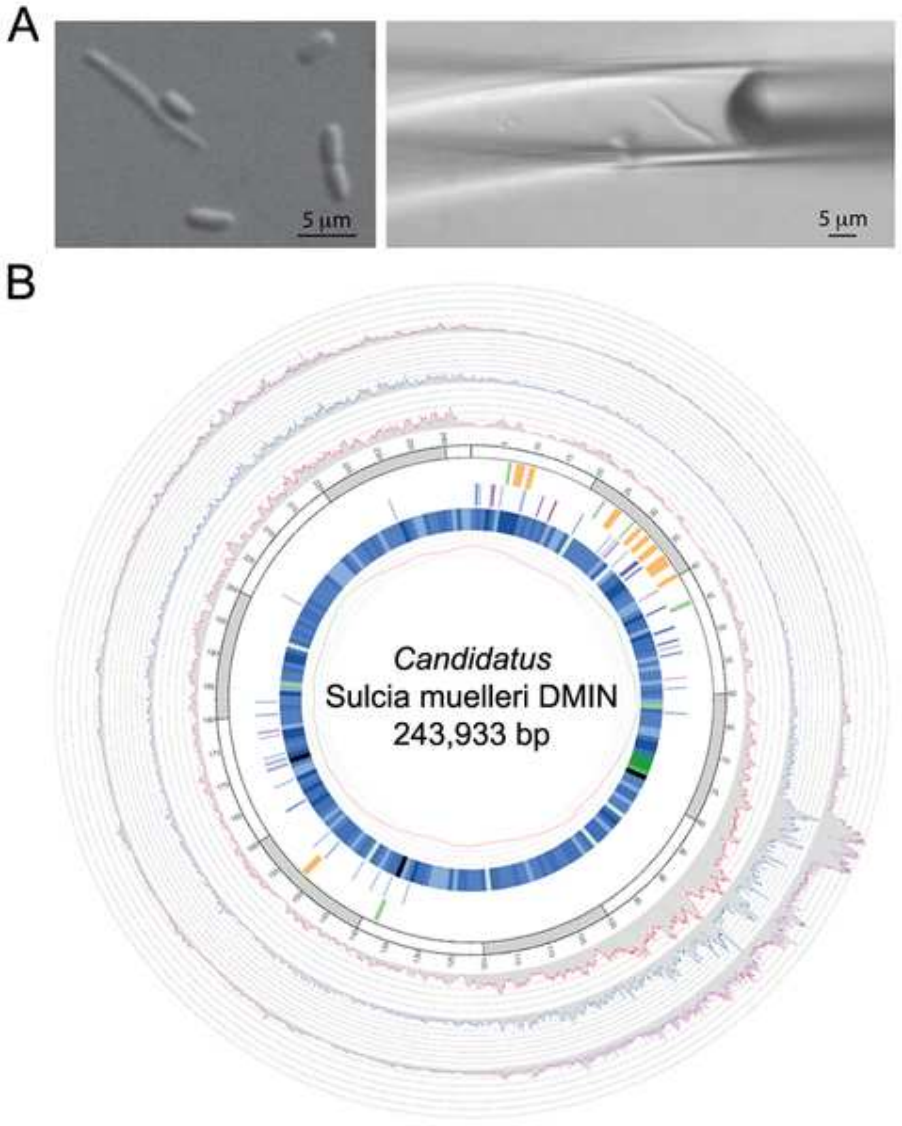
In fecal samples, bacteria and host cells were removed by centrifugation, resulting in gain of 484–15,260 reads of norovirus sequence.

These results suggest that such unbiased high-throughput sequencing approach is useful for directly detecting pathogenic viruses without advance genetic information



Sample	Age	Read	Virus
#F1	3	460	Influenza A virus (H3N2)
		3	Human endogenous retrovirus HCML-ARV
#F2	7	20	Influenza A virus (H1N1)
#F3	5	107	Influenza A virus (H3N2)
		7	WU Polyomavirus
#N1	62[a]	7	Norovirus (GII/4)
#N2	82[b]	7,304	Norovirus (GII/4)
#N3	92[b]	15,272	Norovirus (GII/4)
		813	Kyuri green mottle mosaic virus
		7	Citrus tristeza virus
		3	Enterobacteria phage phiK
#N4	3[c]	484	Norovirus (GII/4)
		14	Human coronavirus HKU1
		3	Phage phiV10
		3	Human endogenous retrovirus K
#N5	44[d]	762	Pepper mild mottle virus
		611	Norovirus (GII/4)
		17	Crucifer tobamovirus
		2	Tobacco mosaic virus

Nakamura et al, PLoS ONE, 2008

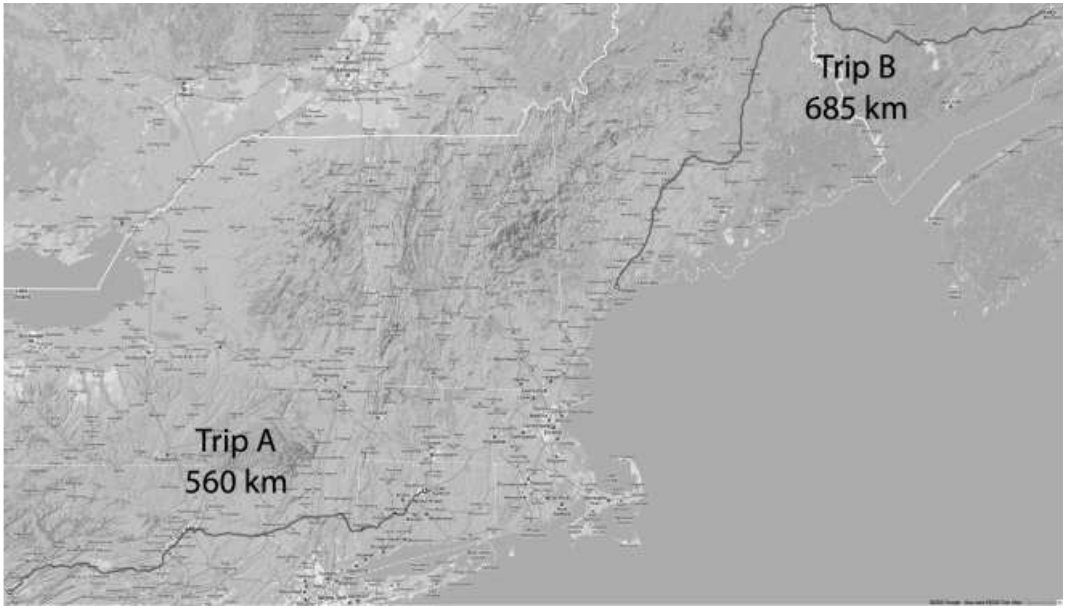


One bacterial cell, one complete genome

Single cell genomics is a novel culture-independent approach, which enables access to the genetic material of an individual cell. Digital PCR on single cells allowed to assess that this symbiotic bacterium is polyploid with genome copies ranging from approximately 200–900 per cell, making it a most suitable target for single cell finishing efforts. For single cell shotgun sequencing, an individual Sulcia cell was isolated and whole genome amplified by multiple displacement amplification. This study demonstrates the power of single cell genomics to generate a complete, high quality, non-composite reference genome within an environmental or clinical sample.

Woyke et al, PLoS ONE, 2010

Absurd community profiling



Windshield splatter analysis with the Galaxy metagenomic pipeline

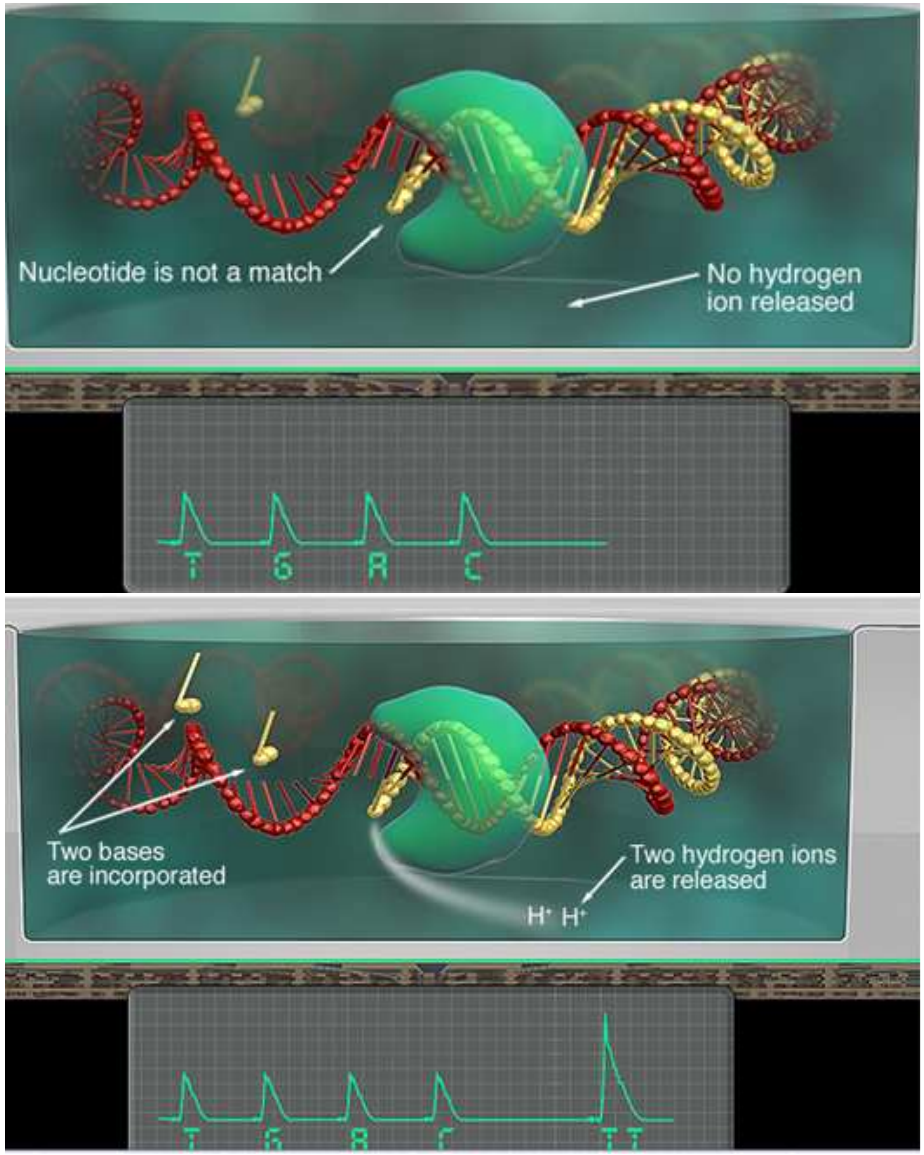
Kosakovsky et al, Genome Res, 2009

How many species inhabit our immediate surroundings? The windshield of a moving vehicle is subjected to numerous insect strikes and can be used as a collection device for representative sampling. In this study, organic matter was collected by a moving vehicle to design and test a comprehensive pipeline for phylogenetic profiling of metagenomic samples that includes all steps from processing and quality control of data generated by next-generation sequencing technologies to statistical analyses and data visualization.

This is one of the first metagenomic study inventorying most of living phyla: in addition to bacteria the samples also contained sequences from plants, mollusks, nematodes, fungi, echinoderms, arthropods, cnidaria, and mammals…

Next-next-generation sequencing

Ion Torrent Technology



High-throughput sequencing is already transforming the research landscape in microbiology. However, it is only a matter of time before it will also transform the practice of clinical microbiology in the reference and routine laboratory setting.
We have to ask how we can embed these technologies and associated workflows into the health care environment, so that health care professionals can understand and exploit the results they produce?
And most importantly, we have to persuade the managers of diagnostic and reference laboratory facilities that the future will look nothing like the past.
Microbiologists of the world, be warned—the sequencing singularity is coming!

Pallen et al., Curr Opin Microbiol, 2010

Grazie per l'attenzione!

ERROR: undefined
OFFENDING COMMAND: JOUSSON_Il

STACK:

(3)
/Title
()
/Subject
(D:20110128103724+01’00’)
/ModDate
()
/Keywords
(PDFCreator Version 0.9.5)
/Creator
(D:20110128103724+01’00’)
/CreationDate
(5317841)
/Author
-mark-