Future Microbial NGS Developments and Genomic Nomenclature

Dag Harmsen
University Münster, Münster, Germany
dharmsen@uni-muenster.de

Commercial Disclosure

Dag Harmsen is co-founder and partial owner of a bioinformatics company (Ridom GmbH, Münster, Germany) that develops software for DNA sequence analysis. Recently Ridom and Ion Torrent/Thermo Fisher (Waltham, MA) partnered and released SeqSphere^{+} software to speed and simplify whole genome based bacterial typing.

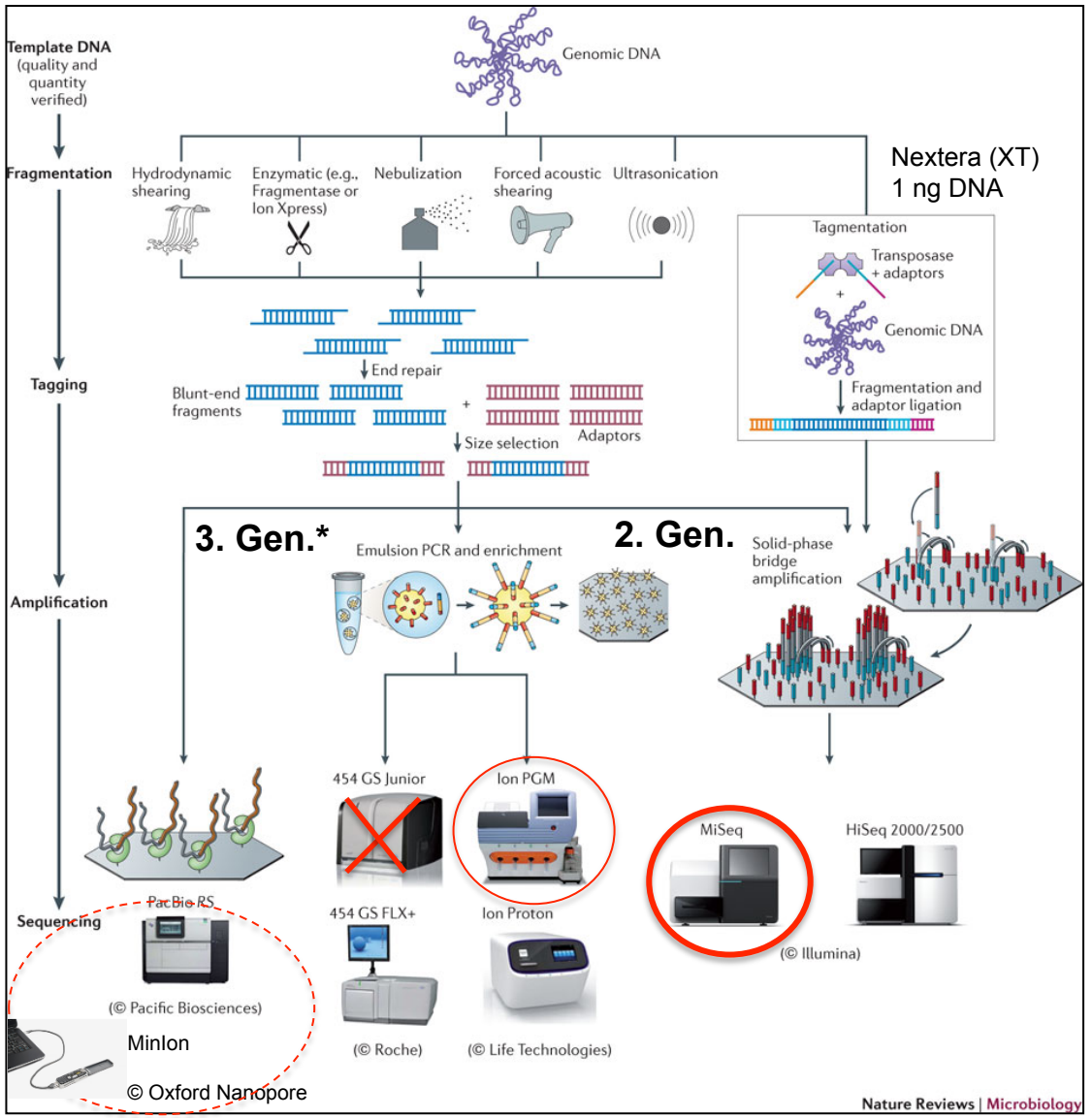
High-throughput Sequencing Workflow and Platforms

1. **DNA Extraction**

2. **Library Preparation**

3.1. ~~Template Amplification~~

3.2. **High-throughput Sequencing**



***Single Molecule Sequencing**
- larger quantities of DNA input (1-5 µg)
- longer reads
- higher read error rate (≈ at random)

Clinical microbiologist platforms – desktop / benchtop machines as speed really matters

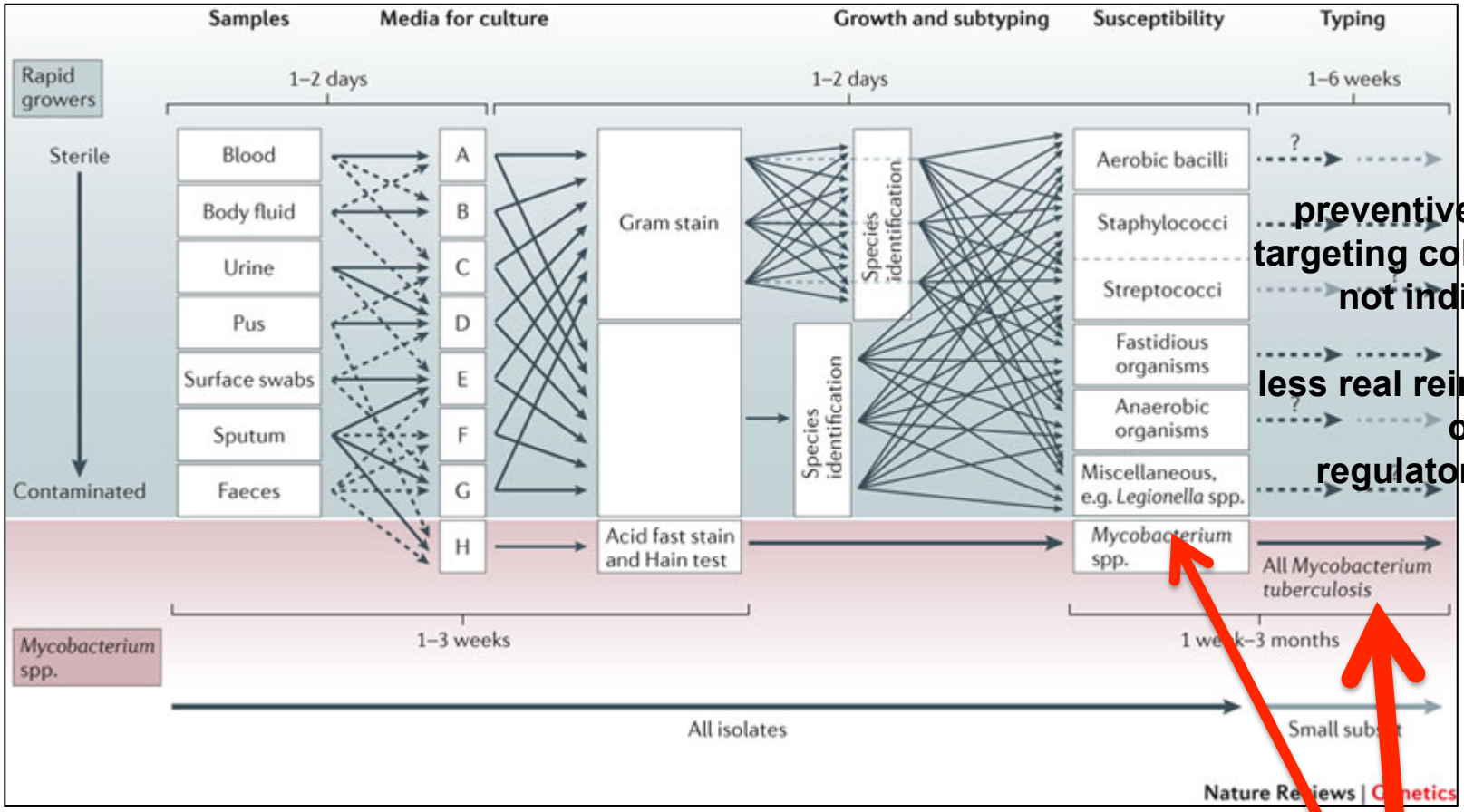
4. **Run, Raw Data QC, Bioinformatics**

Loman *et al.* (2012). *Nature Rev. Microbiol.* **10:** 599 [PubMed].

Diagnostic Workflow in Clinical Bacteriology/Mycology

Meta-genomics | Genomics

Cultivation **Identification** **Susceptibility** **Genotyping**



Didelot *et al*. (2012). *Nature Rev. Genet*. **13:** 601 [PubMed].

preventive in nature targeting collectives and not individuals

less real reimbursement or regulatory issues

number of samples / isolates

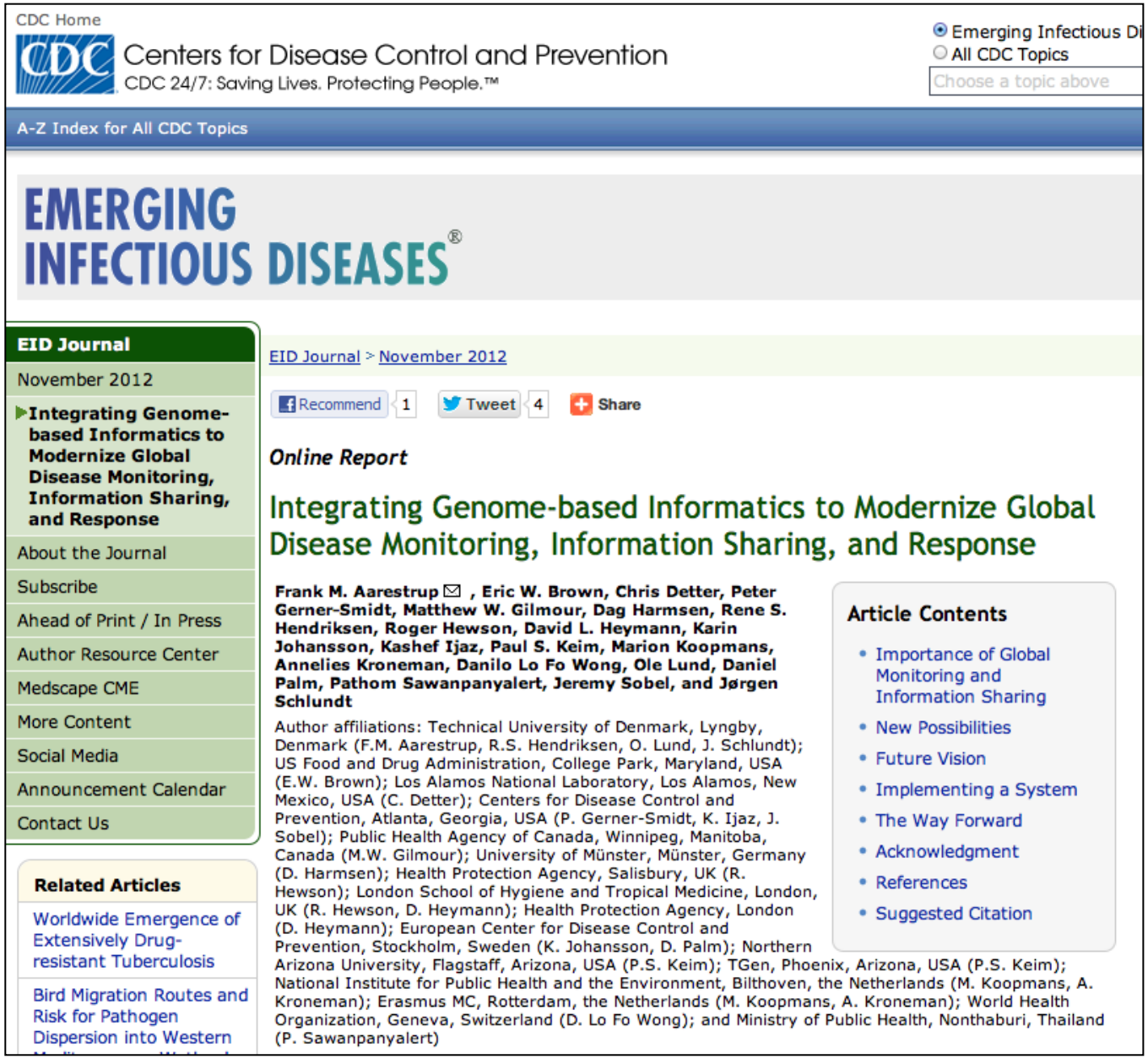
NGS* WGS**

*NGS, next generation sequencing; WGS, whole genome (shotgun) sequencing.

***Claydon** *et al.* (1996). *Nat. Biçotechnol.* **14:**1584 [PubMed],

Mellmann *et al*. (2008). *J. Clin. Microbiol*. 4**6:**1946 [PubMed].

Global Microbial Identifier initiative



CDC Home

Centers for Disease Control and Prevention

CDC 24/7: Saving Lives. Protecting People.™

Emerging Infectious Di
All CDC Topics
Choose a topic above

A-Z Index for All CDC Topics

EMERGING INFECTIOUS DISEASES®

EID Journal
November 2012
Integrating Genome-based Informatics to Modernize Global Disease Monitoring, Information Sharing, and Response
About the Journal
Subscribe
Ahead of Print / In Press
Author Resource Center
Medscape CME
More Content
Social Media
Announcement Calendar
Contact Us

Related Articles
Worldwide Emergence of Extensively Drug-resistant Tuberculosis
Bird Migration Routes and Risk for Pathogen Dispersion into Western

EID Journal > November 2012

Recommend 1 Tweet 4 Share

Online Report

Integrating Genome-based Informatics to Modernize Global Disease Monitoring, Information Sharing, and Response

Frank M. Aarestrup ✉, Eric W. Brown, Chris Detter, Peter Gerner-Smidt, Matthew W. Gilmour, Dag Harmsen, Rene S. Hendriksen, Roger Hewson, David L. Heymann, Karin Johansson, Kashef Ijaz, Paul S. Keim, Marion Koopmans, Annelies Kroneman, Danilo Lo Fo Wong, Ole Lund, Daniel Palm, Pathom Sawanpanyalert, Jeremy Sobel, and Jørgen Schlundt

Author affiliations: Technical University of Denmark, Lyngby, Denmark (F.M. Aarestrup, R.S. Hendriksen, O. Lund, J. Schlundt); US Food and Drug Administration, College Park, Maryland, USA (E.W. Brown); Los Alamos National Laboratory, Los Alamos, New Mexico, USA (C. Detter); Centers for Disease Control and Prevention, Atlanta, Georgia, USA (P. Gerner-Smidt, K. Ijaz, J. Sobel); Public Health Agency of Canada, Winnipeg, Manitoba, Canada (M.W. Gilmour); University of Münster, Münster, Germany (D. Harmsen); Health Protection Agency, Salisbury, UK (R. Hewson); London School of Hygiene and Tropical Medicine, London, UK (R. Hewson, D. Heymann); Health Protection Agency, London (D. Heymann); European Center for Disease Control and Prevention, Stockholm, Sweden (K. Johansson, D. Palm); Northern Arizona University, Flagstaff, Arizona, USA (P.S. Keim); TGen, Phoenix, Arizona, USA (P.S. Keim); National Institute for Public Health and the Environment, Bilthoven, the Netherlands (M. Koopmans, A. Kroneman); Erasmus MC, Rotterdam, the Netherlands (M. Koopmans, A. Kroneman); World Health Organization, Geneva, Switzerland (D. Lo Fo Wong); and Ministry of Public Health, Nonthaburi, Thailand (P. Sawanpanyalert)

Article Contents
- Importance of Global Monitoring and Information Sharing
- New Possibilities
- Future Vision
- Implementing a System
- The Way Forward
- Acknowledgment
- References
- Suggested Citation

GMI
Global Microbial Identifier

http://www.globalmicrobialidentifier.org/

Aarestrup *et al*. (2012). Emerg. Infect. Dis. **18:**e1 [PubMed].

Microbial Genomic Epidemiology

'CDC collaborated with international partners to identify the **core genome** and create a whole-genome multilocus sequence typing (wgMLST) scheme for *Lm*.'

Jackson *et al.* (2016). *Clin. Infect. Dis.* **62:** pii: ciw242 [PubMed].

Challenges Towards Personalized Microbiology NGS

- **Scientific** challenges involve developing robust evidence of predictive utility. While the richness of NGS data expands predictive potential, it also increases the complexity of showing analytical and predictive validity.

- **Clinical** challenges involve providing robust evidence of clinical utility. Innovators must go beyond establishing analytical utility and show how the test can be used effectively in clinical practice to change actual treatment decisions and improve clinical outcomes. *QC/QA must be operational.*

- **Regulatory** challenges require innovators to navigate a fluid and increasingly demanding regulatory environment, including evaluating the pros and cons of different potential pathways to market (*e.g.*, laboratory-developed tests [**LDT**] vs. FDA-approved IvD kit) and associated evidence demands, costs, and risks.

- **Reimbursement** challenges await those products that successfully navigate the challenges above. Increasingly, innovators must show that the new test provides value to payers: that its clinical value drives economic value for payers.

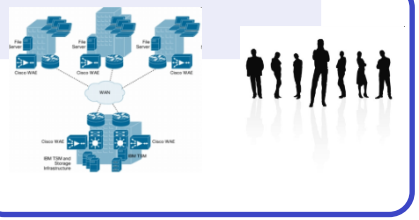
www.analysisgroup.com; **fall 2014**, **Diagnostics and next-generation genetic sequencing.**

Current Clinical NGS Workflow Bottlenecks

Sample Processing

NGS Platforms

Bioinformatics, IT infrastructure



NGS Machine Manufactures Try to Become IVD Companies
Qiagen

genomeweb

Illumina Sues Qiagen for Patent Infringement

Jun 01, 2016 | a GenomeWeb staff reporter

NEW YORK (GenomeWeb) – Illumina has sued Qiagen, alleging that the company's GeneReader next-generation sequencing instrument infringes on a patent Illumina holds related to sequencing-by-synthesis technology, according to documents filed with the US District Court of the Northern District of California.

Illumina alleges that Qiagen's GeneReader infringes on US Patent No. 7,566,537, titled "Labelled Nucelotides," which describes a method of labeling nucleotides as part of the sequencing-by-synthesis technology underlying Illumina's instruments.

Bio.

NGS Machine Manufactures Try to Become IVD Companies
Illumina

genomeweb

Strengthening its Regulatory Strategy Support, Illumina Buys Myraqa

July 16, 2014

Strengthening its Regulatory Strategy Support, Illumina Buys Myraqa

By a GenomeWeb staff reporter

NEW YORK (GenomeWeb) – Illumina today announced it has acquired regulatory and quality consulting firm Myraqa for an undisclosed amount.

development and general manager of its newly formed enterprise informatics unit.

Computing Companies Try to Participate Too

genomeweb

SAP Starts Early Access for Clinical, Genomic Data Integration Software; Announces PHEMI HIV Pact

February 28, 2014

SAP Starts Early Access for Clinical, Genomic Data Integration Software; Announces PHEMI HIV Pact

By Uduak Grace Thomas

SAP has unveiled an early version of a new standalone application based on its SAP Hana in-memory database technology called Medical Insights that integrates and analyzes clinical and genomic data to help oncologists make better treatment decisions for patients.

Computing Companies Try to Participate Too

Pathogens **2014**, *3*, 437-458; doi:10.3390/pathogens3020437

OPEN ACCESS

pathogens

ISSN 2076-0817

www.mdpi.com/journal/pathogens

Review

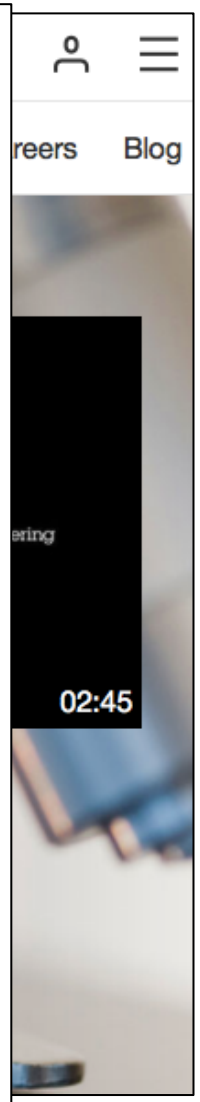
WGS Analysis and Interpretation in Clinical and Public Health Microbiology Laboratories: What Are the Requirements and How Do Existing Tools Compare?

Kelly L. Wyres [1,*], Thomas C. Conway [1], Saurabh Garg [1], Carlos Queiroz [1], Matthias Reumann [2], Kathryn Holt [3] and Laura I. Rusu [1]

[1] IBM Research—Australia, Level 5, 204 Lygon Street, Carlton, Victoria 3053, Australia; E-Mails: tconway@au1.ibm.com (T.C.C.); saurabh.kr.garg@gmail.com (S.G.); caxqueiroz@gmail.com (C.Q.); laurusu@au1.ibm.com (L.I.R.)

[2] IBM Research—Zuerich, Säumerstrasse 4, Rüschlikon 8803, Switzerland; E-Mail: MRE@zurich.ibm.com

[3] Bio21 Institute, University of Melbourne, 30 Flemington Road, Melbourne, Parkville, VIC 3052, Australia; E-Mail: kholt@unimelb.edu.au



w.ibm.com

IVD Companies Try to Integrate (Meta-)Genomics

BIOMÉRIEUX

genomeweb

Becton Dickinson Acquires Single-Cell Genomics Startup Cellular Research

Aug 25, 2015 | a GenomeWeb staff reporter

NEW YORK (GenomeWeb) – Becton Dickinson said after the close of the market Tuesday that it has acquired San Francisco Bay Area single-cell genomics startup Cellular Research for an undisclosed amount.

POWERED BY illumina

Regulatory NGS Challenges

Although the FDA is now developing regulatory frameworks for NGS-based tests, **NGS test developers face significant uncertainty** that has the potential to undermine product economics. Historically, a significant proportion of diagnostic tests in the United States have been conducted in certified labs as LDTs, rather than with the use of 'kits' designed for the purpose by a diagnostics manufacturer. Unlike manufacturer-developed kits, LDTs have not traditionally been subject to FDA approval. This means that late-stage opportunists could market **'copycat' tests** that draw on the same underlying pattern and mode of detection as FDA-approved tests, without taking on any of the cost or risk necessary to secure regulatory approval.

genomeweb

FDA Outlines Ways to Assess Analytical, Clinical Performance of NGS ahead of February Workshop

Jan 06, 2015 | a GenomeWeb staff reporter

NEW YORK (GenomeWeb) – The US Food and Drug Administration is considering a standards-based approach for assessing the analytical performance of next-generation sequencing diagnostic tests and using centralized curated databases to evaluate their clinical performance, according to a recent paper published by the agency.

The preliminary discussion paper, posted on the FDA's website last month, outlines possible ways the agency might regulate NGS tests in the future and provides discussion points for a one-day public workshop entitled "Optimizing FDA's Regulatory Oversight of Next Generation Sequencing Diagnostic Tests" that the FDA will hold at the National Institutes of Health campus in Bethesda, Md. on February 20.

www.analysisgroup.com; **fall 2014. Diagnostics and next-generation genetic sequencing.**

In **Europe** neither FDA-like **510(k) clearance*** (only **CE labeling**; *i.e.*, 'self-declaration') nor **LDT oversight**. Therefore, even American companies frequently **enter clinical market first** in Europe. Most clinical microbiology NGS **scientific evidence** currently emerge also from here!

*This will change somewhat as the EU has published a new IVD Regulation which will replace the current Directive 98/79/EC on In-Vitro Diagnostic Medical Devices (IVDD) from around 2017.

Assuring Quality of NGS in Clinical Laboratory Practice

genomeweb

CAP Publishes Accreditation Checklist for NGS in Clinical Labs

August 01, 2012

CAP Publishes Accreditation Checklist for NGS in Clinical Labs

By Monica Heger

The College of American Pathologists has published a checklist specific to next-generation sequencing for clinical lab accreditation. The NGS-specific checklist is part of CAP's revised molecular pathology checklist for accrediting clinical laboratories, released this week.

the analytical phase of testing	•Synthetic DNA and electronic reference data files may serve as RMs for rare or challenging sequence variations. •Efforts should be undertaken to establish a suitable NGS RM and the sequence of the RM should be refined as the technology changes. Such a RM should be annotated to indicate regions of high and low sequence reliability.

[a]See **Supplementary Guidelines** for complete recommendations. RM, reference material.

http://www.cap.org/

Gargis *et al.* (2012). *Nature Biotechnology* **30 (11):** 1033 [PubMed].

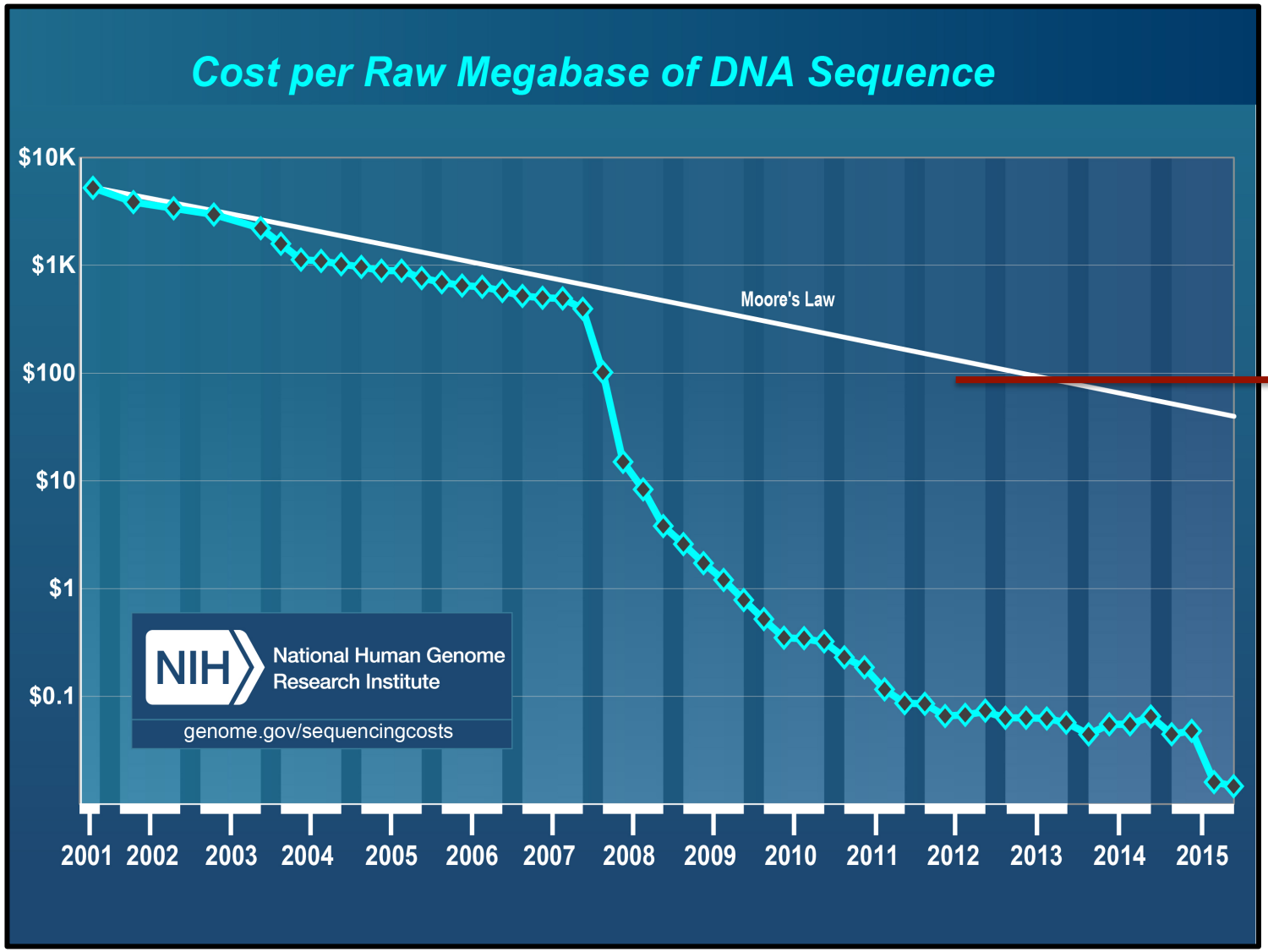
Reimbursement NGS Challenges

The move away from code-stacking (*i.e.*, utilizing miscellaneous codes and/or process step codes) has increased the importance of obtaining new **CPT** (current procedural terminology) codes for novel diagnostics. Recognizing the potential of these NGS techniques, the American Medical Association (**AMA**) has proposed in February 2014 new CPT codes *(for oncology, human genetics)* to cover a broad range of NGS-based tests, which are under consideration for introduction.

Historically, however, **securing new CPT codes** has been a **multi-year proposition**, and the effectiveness of changes in the AMA's process remains unknown.

www.analysisgroup.com; **fall 2014**, **Diagnostics and next-generation genetic sequencing.**

Cost of NGS in Clinical Microbiology



Consumable cost per bacterial genome (about 5 Mb; Illumina MiSeq)

www.genome.gov

Pacific Biosciences

genomeweb

PACBIO

PRODUCTS + SERVICES | RESEARCH FOCUS | APPLICATIONS | SMRT SCIENCE | SUPPORT | COMPANY

PRODUCTS + SERVICES

SEQUEL SYSTEM

PACBIO SYSTEMS

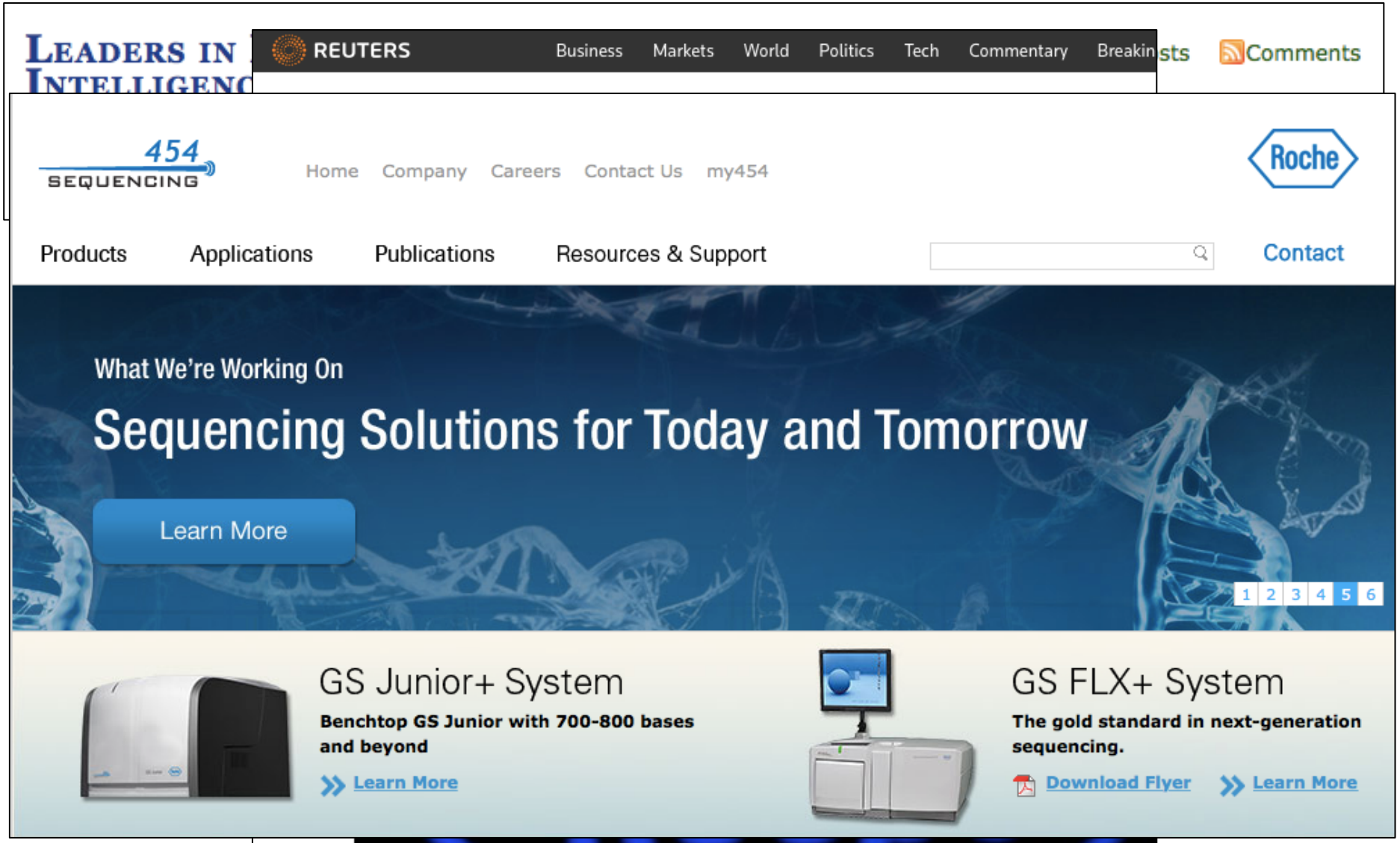
- Sequel System
  - Sequel Software
- PacBio RS II
  - PacBio RS II Software

THE SCALABLE PLATFORM FOR SINGLE MOLECULE, REAL-TIME (SMRT) SEQUENCING

Accelerate your research with the most comprehensive view of genomes, transcriptomes, and epigenomes from our higher-throughput system. Reduce project costs and timelines as you create higher-quality whole genome assemblies and explore the full size-spectrum of genetic variation.

Not really a desktop machine! However, closed genomes at a cost per bacterial genome that is competitive with Illumina!

www.pacb.com



First heard about and seen NGS [illegible] SM [illegible] al Meeti [illegible] 005 [illegible] S.

Roche tablets are seen positioned in front of a displayed Roche logo in this photo illustration shot in Zenica, Bosnia a Herzegovina, January 22, 2016.
REUTERS/DADO RUVIC

Commercial Nanopore Sequencing / Genia

Genia Publishes Proof of Principle Study, Says Commercial System Will Serve Clinical Market

Apr 19, 2016 | Monica Heger

Premium

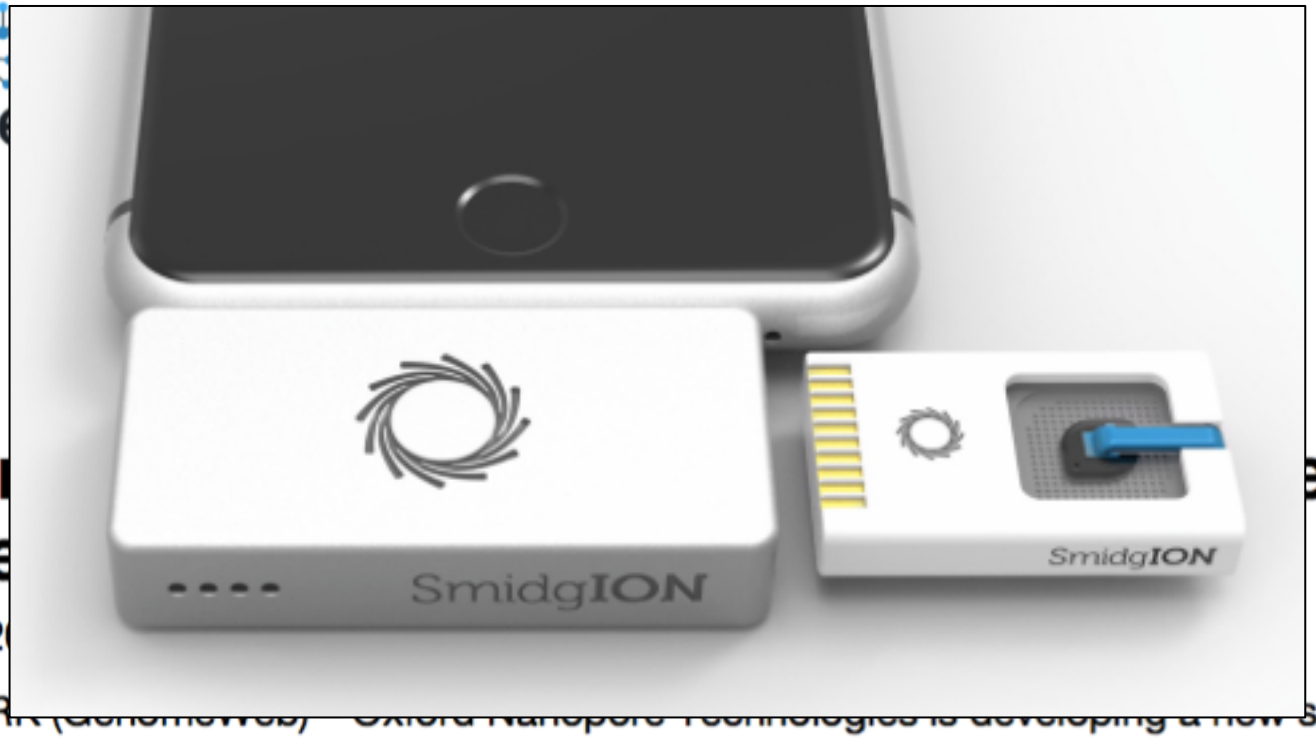
NEW YORK (GenomeWeb) – Genia has published a proof of principle study of its nanopore-based sequencing-by-synthesis technology, although the final commercial product will look substantially different and be suitable for clinical diagnostics, according to Genis CEO Stefan Roever.

Roever did not provide a timeline for when the firm, which Roche acquired in 2014, would launch an instrument. In 2013, he had said that Genia planned to launch by the end of 2014, but after the acquisition, Roche "revisited what they thought the specs would need to be to launch a sequencer into the clinical sequencing market," Roever told GenomeWeb. Roche had a "substantially more stringent set of requirements" than the targets Genia had previously set as goals for launching a viable product, he said.

In the new study, which was published this week in the *Proceedings of the National Academy of the Sciences*, Genia researchers, in collaboration with Jingyue Ju's lab at Columbia University, George Church's group at Harvard Medical School, and researchers from the National Institute of Standards and Technology, demonstrated that they could design a 264-nanopore array and use a tagging technique to discriminate the four nucleotides and sequence DNA.

Wikipedia.org

Oxford Nanopore Technologies – Point-of-Care NGS

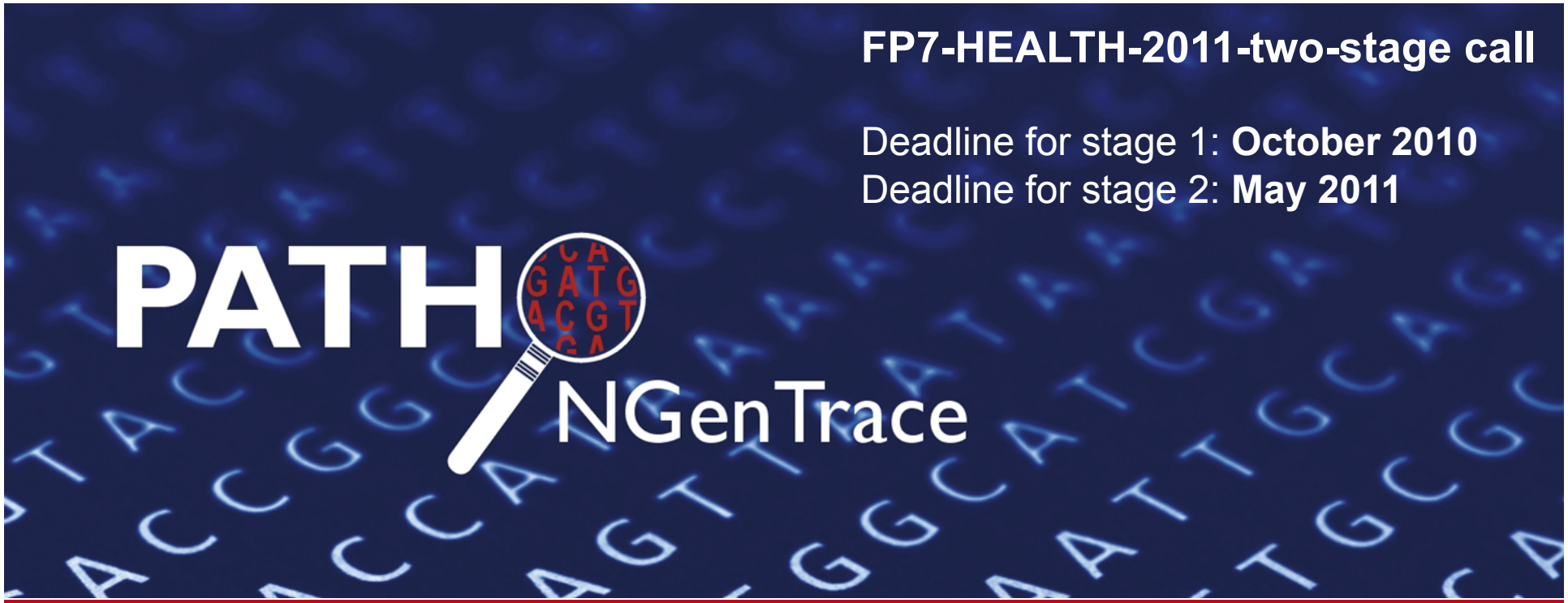


Smidgion is in early development and is expected to be available **in late 2017**.

www.nanoporetech.com

Nature of and Requirements for Microbial Genomic Nomenclature

- is always a **reduction of complexity/information** and serves mainly for **human communication** to quickly group/compare isolates
- is **artificial**, *i.e.*, it is man made for humans and NOT for bugs
- must be **public**
- must be **stable**
- must be **sustainable**
- must be **automatic expandable** at **low compute cost**



Bacterial genome detectives

Next Generation Genome Based High Resolution Tracing of Pathogens

Grant agreement number:
278864-2

EC contribution:
5.995.267 €

Duration:
54 months (**01/01/2012** - 30/06/2016)

Funding scheme:
SME-targeted Collaborative Project

URL:
http://www.patho-ngen-trace.eu/

Scientific Advisory Board

Marc J. Struelens
European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden

Rene S. Hendriksen
Technical University of Denmark - National Food Institute, Lyngby, Denmark

Stephen H. Gillespie
University of St Andrews, St Andrews, Scotland UK

Gary Van Domselaar
National Microbiology Laboratory Public Health Agency of Canada

PATH NGenTrace

Consortium

Stefan Niemann
Coordinator, FZ Borstel

Claudia Giehl
Eurice GmbH

Dag Harmsen
Universität Münster

Martin C.J. Maiden
University of Oxford

Bruno Pot
Applied Maths NV

Philip Supply
Genoscreen

Ronald Burggrave
Piext BV

Jörg Rothgänger
Ridom GmbH

Associated Partners

Alexander Mellmann
Univ. Münster, Germany

Roland Diel
Univ. Kiel, Germany

Joao Carrico
Univ. Lisbon, Portugal

OpGEN®

life technologies™

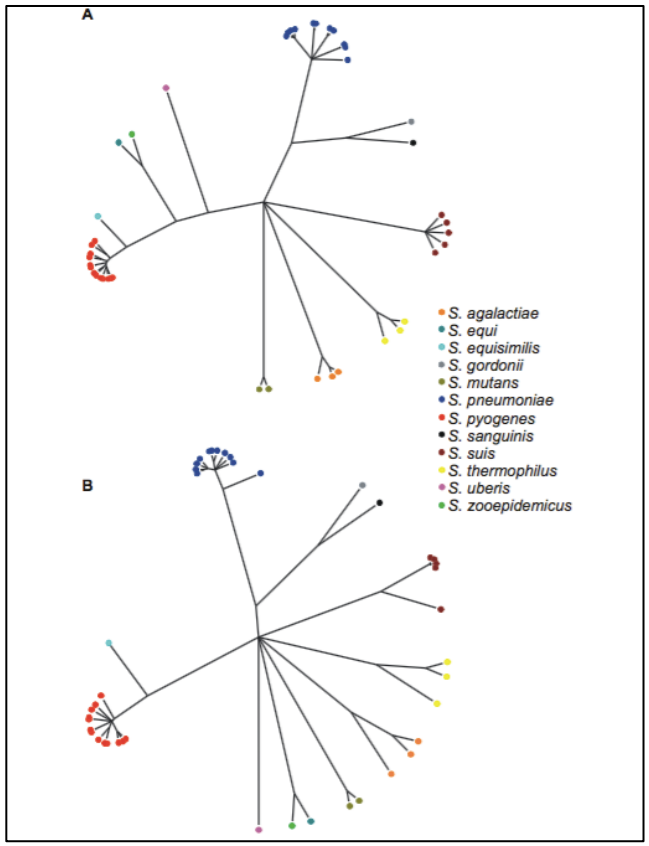
Jolley and Maiden *BMC Bioinformatics* 2010, **11**:595
http://www.biomedcentral.com/1471-2105/11/595

BMC Bioinformatics

SOFTWARE Open Access

BIGSdb: Scalable analysis of bacterial genome variation at the population level

Keith A Jolley*, Martin CJ Maiden



Jolley & Maiden (November, 2010). *BMC Bioinformatics*. **11:** 595 [PubMed].

ClonalFrame trees were generated from 43 streptococcal genome sequences, i.e., from **concatenated sequences**, using **A** seven MLSA gene fragment loci and **B** 77 complete genes found to be present throughout the genus identified by BIGSdb.

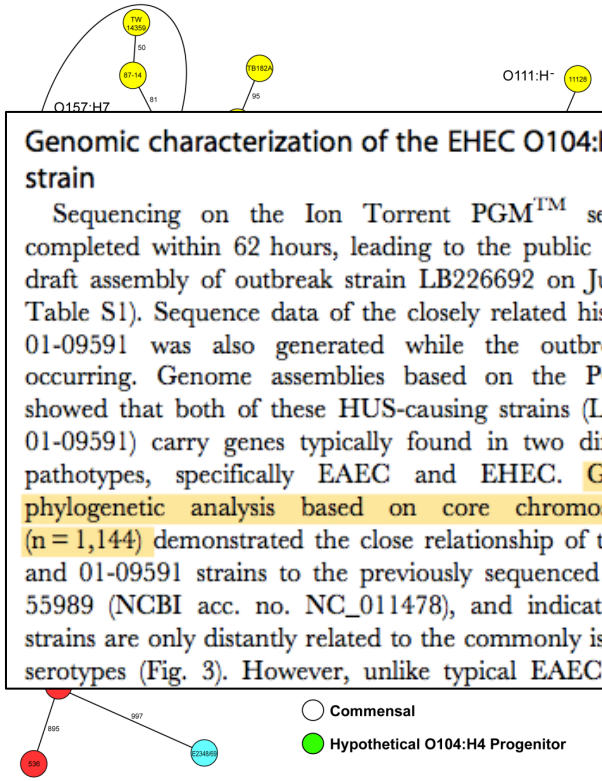
OPEN ACCESS Freely available online

PLoS one

Prospective Genomic Characterization of the German Enterohemorrhagic *Escherichia coli* O104:H4 Outbreak by Rapid Next Generation Sequencing Technology

Alexander Mellmann[1☯], Dag Harmsen[2*☯], Craig A. Cummings[3☯], Emily B. Zentz[4], Shana R. Leopold[1], Alain Rico[5], Karola Prior[2], Rafael Szczepanowski[2], Yongmei Ji[3], Wenlan Zhang[1], Stephen F. McLaughlin[3], John K. Henkhaus[4], Benjamin Leopold[1], Martina Bielaszewska[1], Rita Prager[6], Pius M. Brzoska[3], Richard L. Moore[4], Simone Guenther[5], Jonathan M. Rothberg[7], Helge Karch[1]

1 Institute of Hygiene, University Münster, Münster, Germany, **2** Department of Periodontology, University Münster, Münster, Germany, **3** Life Technologies, Foster City, California, United States of America, **4** OpGen, Gaithersburg, Maryland, United States of America, **5** Life Technologies, Darmstadt, Germany, **6** Robert Koch Institute, Wernigerode Branch, Wernigerode, Germany, **7** Ion Torrent by Life Technologies, Guilford, Connecticut, United States of America



Genomic characterization of the EHEC O104:H4 outbreak strain

Sequencing on the Ion Torrent PGM™ sequencer was completed within 62 hours, leading to the public release of the draft assembly of outbreak strain LB226692 on June 3 (Fig. 1, Table S1). Sequence data of the closely related historical isolate 01-09591 was also generated while the outbreak was still occurring. Genome assemblies based on the PGM™ reads showed that both of these HUS-causing strains (LB226692 and 01-09591) carry genes typically found in two different *E. coli* pathotypes, specifically EAEC and EHEC. Genome wide phylogenetic analysis based on core chromosomal ORFs (n = 1,144) demonstrated the close relationship of the LB226692 and 01-09591 strains to the previously sequenced EAEC strain 55989 (NCBI acc. no. NC_011478), and indicated that these strains are only distantly related to the commonly isolated EHEC serotypes (Fig. 3). However, unlike typical EAEC strains, both

Phylogenetic Analysis of EHEC 0104:H4

Method

- **first real-time** prospective outbreak **genomics outbreak analysis**. Hybrid assembly from reference mapping & ***de novo*** **assembly** with **Ion Torrent PGM desktop machine** WGS data and **BIGSdb genome-wide gene-by gene allele** calling against a **fixed set of loci/targets**
- n = 1,144 STEC **core genome gene scheme** defined **before** outbreak analysis and SeqSphere minimum-spanning tree (not yet termed so but **first cgMLST application**; internally called at that time 'super MLST' and/or 'MLST on steroids')

Mellmann *et al*. (July, 2011). *PLoS One*. **6:** e22751 [PubMed].

F1000 FACULTY OF 1000 POST-PUBLICATION PEER REVIEW

nature biotechnology

Home | Current issue | News & comment | Research | Archive ▾ | Authors & referees ▾ | About the

Updating benchtop sequencing performance comparison

Sebastian Jünemann, Fritz Joachim Sedlazeck, Karola Prior, Andreas Albersmeier, Uwe John, Jörn Kalinowski, Alexander Mellmann, Alexander Goesmann, Arndt von Haeseler, Jens Stoye & Dag Harmsen

valuation of contiguity and consensus accuracy
nchtop
uity for
the *de*
erated
udo- or
E. coli
. This
alysis
here+
ensus
eq 2 ×
sensus
Sakai
here+
ariants
ng are

generated NGS data and the reference were resolved by bidirectional Sanger sequencing. The finally 75 validated benchtop NGS consensus errors were visualized and reported (**Fig. 1b**) as Venn diagram (*variants_assembly.R*) using the VennDiagram R package[18].

Outlook. A genome-wide gene-by-gene analysis as used in this study can be regarded as an obvious higher discriminatory extension of the highly popular and successful Multi-Locus Sequence Typing (MLST) approach[19]. Such an 'extended' MLST or MLST+* was used for the first time with NGS data during the phylogenetic analysis of the German 2011 EHEC

Jünemann *et al.* (April, 2013). *Nature Biotechnology* **31:** 294 [PubMed].

***Avoidance of the term core genome as core genome genes are here determined from DNA with rather high similarity values! – term MLST+ abandoned by Ridom September 2015.**

OPINION

MLST revisited: the gene-by-gene approach to bacterial genomics

Martin C. J. Maiden, Melissa J. Jansen van Rensburg, James E. Bray, Sarah G. Earle, Suzanne A. Ford, Keith A. Jolley and Noel D. McCarthy

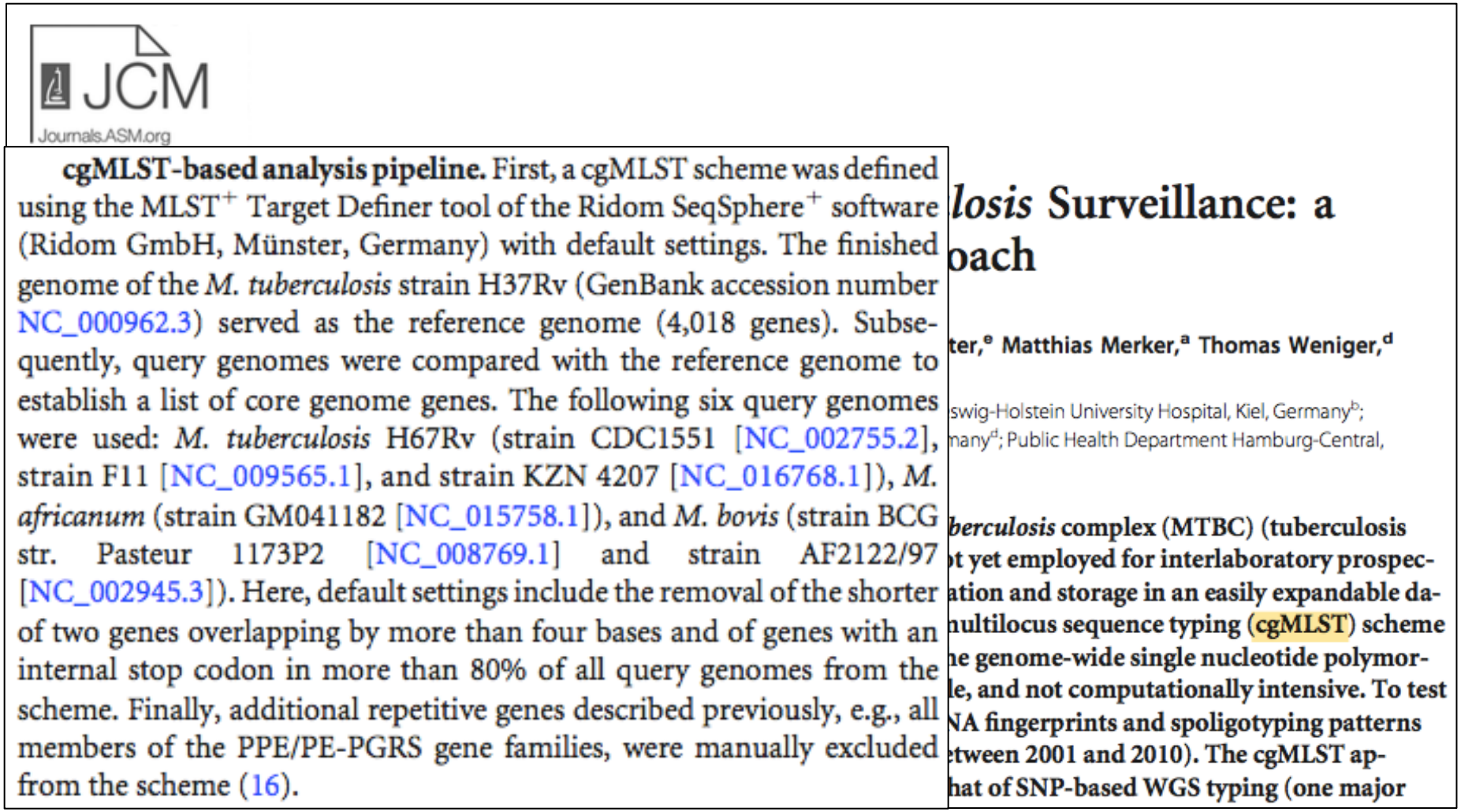
Abstract | Multilocus sequence typing (MLST) was proposed in 1998 as a portable sequence-based method for identifying clonal relationships among bacteria. Today, in the whole-genome era of microbiology, the need for systematic, standardized descriptions of bacterial genotypic variation remains a priority. Here, to meet this need, we draw on the successes of MLST and 16S rRNA gene sequencing to propose a hierarchical gene-by-gene approach that reflects functional and evolutionary relationships and catalogues bacteria 'from domain to strain'. Our gene-based typing approach using online platforms such as the Bacterial Isolate Genome Sequence Database (BIGSdb) allows the scalable organization and analysis of whole-genome sequence data.

Maiden *et al.* (October, 2013). *Nature Rev. Microbiol.* **11:** 728 [PubMed].

Different levels of sequence information can be associated with different taxonomic levels. Analysis of a single locus, for example, is often sufficient to distinguish many groups (phylum, class, order, family or genus), whereas determining speciation and subspeciation requires higher resolution, which can be attained by increasing the number of loci analysed. Researchers can apply a set of MLST approaches, each using different numbers of loci and each suitable for addressing different levels of isolate discrimination (FIG. 2). The highest level of resolution using a gene-by-gene approach can be termed whole-genome MLST (wgMLST), in which all the loci of a given isolate are compared to equivalent loci in other isolates. The wgMLST approach is applicable to single-clone pathogens with closed genomes or to very closely related variants of more diverse organisms. This analysis can involve the entire genome if loci corresponding to intergenic regions are also defined. Few bacteria share all loci, so comparisons of the core genome of a given group (core-genome MLST (cgMLST)) provides high-resolution data across a group of related but

cgMLST at that time for the authors **NOT** a fixed set of loci but 'shared' loci of selected isolates under study.

PATHO NGenTrace



JCM
Journals.ASM.org

cgMLST-based analysis pipeline. First, a cgMLST scheme was defined using the MLST$^+$ Target Definer tool of the Ridom SeqSphere$^+$ software (Ridom GmbH, Münster, Germany) with default settings. The finished genome of the *M. tuberculosis* strain H37Rv (GenBank accession number NC_000962.3) served as the reference genome (4,018 genes). Subsequently, query genomes were compared with the reference genome to establish a list of core genome genes. The following six query genomes were used: *M. tuberculosis* H67Rv (strain CDC1551 [NC_002755.2], strain F11 [NC_009565.1], and strain KZN 4207 [NC_016768.1]), *M. africanum* (strain GM041182 [NC_015758.1]), and *M. bovis* (strain BCG str. Pasteur 1173P2 [NC_008769.1] and strain AF2122/97 [NC_002945.3]). Here, default settings include the removal of the shorter of two genes overlapping by more than four bases and of genes with an internal stop codon in more than 80% of all query genomes from the scheme. Finally, additional repetitive genes described previously, e.g., all members of the PPE/PE-PGRS gene families, were manually excluded from the scheme (16).

losis Surveillance: a
oach

ter,[e] Matthias Merker,[a] Thomas Weniger,[d]

swig-Holstein University Hospital, Kiel, Germany[b];
nany[d]; Public Health Department Hamburg-Central,

berculosis complex (MTBC) (tuberculosis
t yet employed for interlaboratory prospec-
ation and storage in an easily expandable da-
ultilocus sequence typing (cgMLST) scheme
e genome-wide single nucleotide polymor-
le, and not computationally intensive. To test
NA fingerprints and spoligotyping patterns
tween 2001 and 2010). The cgMLST ap-
hat of SNP-based WGS typing (one major

Kohl *et al.* (April, 2014). *JCM* **52:** 2479 [PubMed].

First original publication using the term cgMLST and using a fixed genome-wide set of genes.

cgMLST Nomenclature Harmonization

Vaz *et al. Journal of Biomedical Semantics* 2014, **5**:43
http://www.jbiomedsem.com/content/5/1/43

JOURNAL OF BIOMEDICAL SEMANTICS

RESEARCH Open Access

TypOn: the microbial typing ontology

Cátia Vaz[1,2*], Alexandre P Francisco[1,3], Mickael Silva[4], Keith A Jolley[5], James E Bray[5], Hannes Pouseele[6], Joerg Rothganger[7], Mário Ramirez[4] and João A Carriço[4]

Vaz *et al.* (October, 2014). *J Biomed Semantics* **5:** 43 [PubMed].

The TypON microbial typing ontology foresees immediately a **REST** application programming interface (**API**) for cgMLST allele nomenclature services that allows software tools to bi-directional communicate with each other.

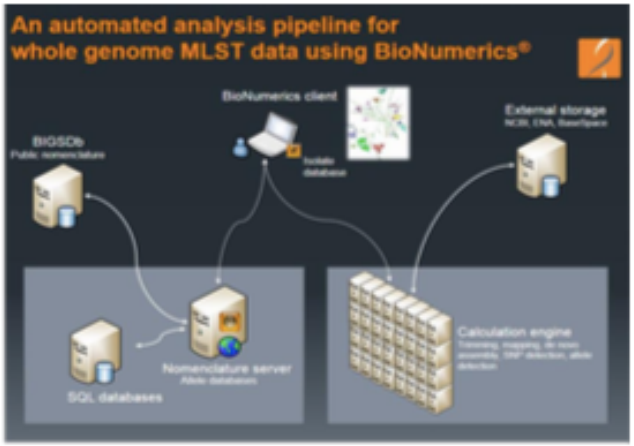
APPLIED MATHS
Manage. Analyze. Discover.

BioNumerics SEVEN

PATH NGenTrace

WHOLE GENOME MLST PLUGIN

As next-generation sequencing is increasingly replacing Sanger sequencing, conventional MLST is gradually extending to whole genome MLST (wgMLST), providing higher resolution. Given the amount of data and the demanding calculations, we developed an automated pipeline with an integrated calculation engine and external storage, so it remains workable from a good average client computer.



Demanding calculations such as de novo assemblies can be performed on an external calculation engine. The choice here is offered between virtually setup-free pay-per-use cloud solutions (e.g. via Amazon) or a local deployment e.g. on a computer cluster (requires custom services). Only the wgMLST allelic profiles are stored in the BioNumerics database, resulting in a lightweight and responsive strain database.

- Import of sequence read sets from various sequencers (Roche 454*, Illumina*, PacBio*, IonTorrent*) and different sources (NCBI, EMBL-EBI, Illumina* Basespace* or Amazon S3). Sequence read sets can be imported as links to these online databases, bypassing the tedious step of up- and downloading the files.
- Batch job processing on the calculation engine to calculate read statistics, perform de novo assemblies, detect loci presence and perform allele identifications using both assembly-based and assembly-free methods.
- Overview and management of the submitted jobs with details on users, submission time, job status and progress is accessible through the Overview window. Double-click on a job displays the job log information.
- The import of job results in the database once processing is finished and automated linkage to the corresponding entry information
- Quality assessment of the wgMLST results, including imperfect and new allele matches, multiple allele matches or non-consensus allele calls.
- Automated submission of new alleles to the allele nomenclature server.
- Synchronization between the wgMLST sample database and the wgMLST typing schemes defined at server side.
- Automated assignments of sequence type and clonal complex information.

Matanock *et al.* (March, 2016). *JCM* **54:** 768 [PubMed].

cgMLST Workshop

cgMLST WORKSHOP
Department of Zoology, Oxford, meeting room D37
2nd March 2015, 12:00 – 17:00
3rd March 2015, 9:00 – 16:00

UNIVERSITY OF OXFORD

Programme:

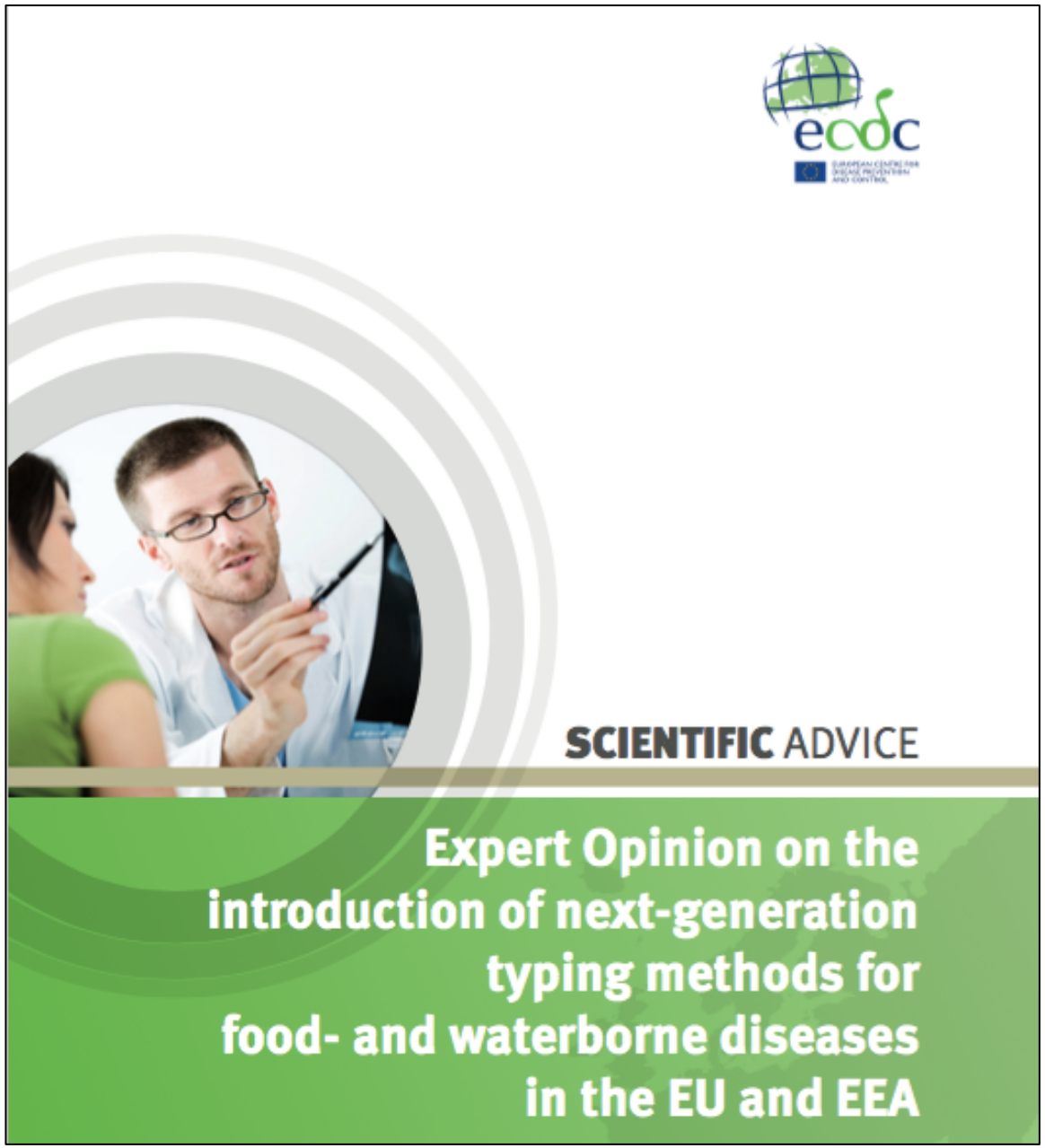
Monday 2 March

- 12:00 Lunch
- 13:00 **Session I cgMLST Strategy**
  - Welcome and Introduction –
    - **Martin Maiden**, Department of Zoology, Oxford
    - **Dag Harmsen**, University Hospital Münster, Germany
- Perspectives on cgMLST
  - **Noel McCarthy**, Medical School, University of Warwick
  - **Collette Fitzgerald**, CDCP Atlanta, USA
  - **Kathie Grant**, Gastrointestinal Bacteria Reference Unit, Public Health England
  - **Sylvain Brisse**, Institut Pasteur, Paris
- 18:30 Pre-dinner drinks and **Dinner at Hertford College**, Oxford

Tuesday 3 March

- 9:00 **Session II Implementation of Interfaces**
  - **João Carriço**, Faculty of Medicine Lisbon/Alexandre Francisco, IS Tecnico Lisbon
  - **Keith Jolley**, Department of Zoology, Oxford
  - **Hannes Pouseele**/Bruno Pot, Applied Maths Sint-Martens-Latem, Belgium
  - **Jörg Rothgänger**, Ridom GmbH, Münster, Germany
  - **Tim Dallman**, Gastrointestinal Bacteria Reference Unit, Public Health England
- 12:00 Lunch
- 13:00 **Session III Defining cgMLST Schemes**
  - **Alison Cody**, Department of Zoology, Oxford
  - **Angela Brueggemann**, Medawar Building for Pathogen Research, Oxford
  - **Mark Achtman**, Medical School, University of Warwick
  - **Stefan Niemann**, Research Centre Borstel, Germany

Informal agreement that a **cgMLST** scheme is a **fixed and agreed upon number of genes for each species or group of closely related species** that is going to be at least the minimum denominator for analyzing whole genome shotgun (WGS) sequence data for surveillance purposes!



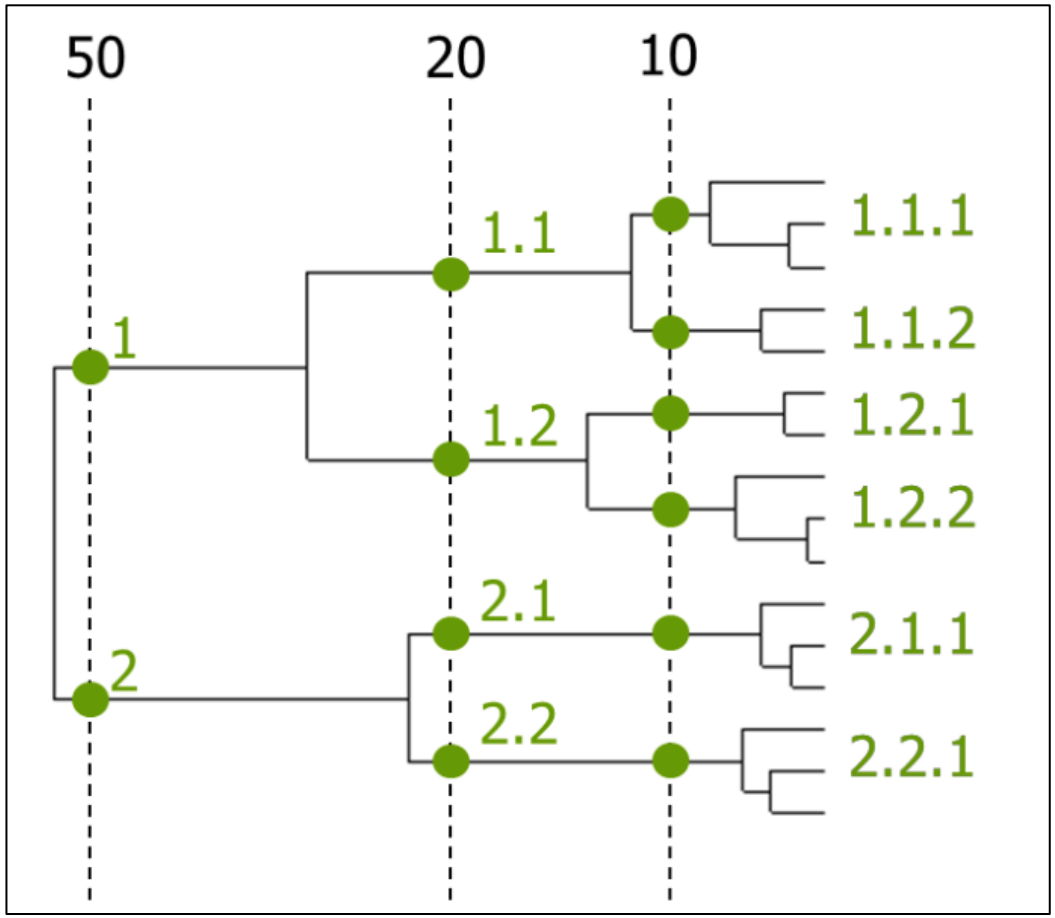
Describes a **top-down approach** that includes also several tears of reporting (e.g., national and international).

However, in the past the most successful bacterial genotyping initiatives (e.g., **MLST**, **spa-typing**, or **MIRU-VNTR**) followed a **bottom-up** - grass-root basic democratic or even anarchic - approach.

Only the **PulseNet** initiative followed a **top-down approach** but never resulted in a public nomenclature and involved 'heavy' investment by CDC.

ECDC (October, 2015).
http://ecdc.europa.eu/en/publications/Publications/food-and-waterborne-diseases-next-generation-typing-methods.pdf.

Taxonomical ('phylogenetic') Nomenclature

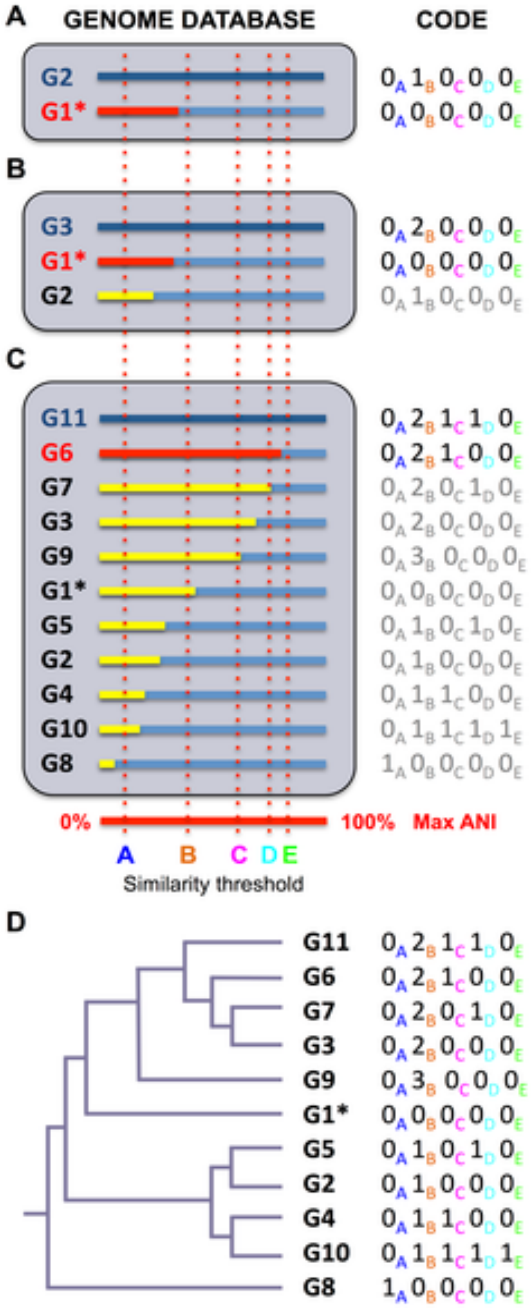


Taxonomical nomenclature principle based on SNP or wgMLST dendrogram.*

*Desirable BUT **impossible for an additive expandable nomenclature system** as there will be always changes in the tree (was not possible in the past with MLST or canonical SNPs of monomorphic bacteria; would violate **stability of nomenclature**). Furthermore, if done with 'SNP addresses' and not with alleles very compute-intensive to calculate.

ECDC (October, 2015). Expert Opinion on the introduction of next-generation typing methods for food- and waterborne diseases in the EU and EEA. http://ecdc.europa.eu/en/publications/Publications/food-and-waterborne-diseases-next-generation-typing-methods.pdf.

Genome Similarity-based Nomenclature ('Genome Codes')



Genome codes are **hierarchical** whereby every position in the code reflects a different level of similarity between organisms. The bin size – which is predictive for diversity at this level - decreases moving from left to right

Like MLST STs numeric value at a certain position does not express relationship!

The information content of genome codes consists exclusively in the extend of **shared code positions**.

However, depending on the **submission order** the **last common position** of two organisms might be **occasionally wrong** ('triangulation issue; similar to CT).

In contrast to **phylogeny-based codes**, **genome codes** (**ANIb** based) could be assigned to an organism automatically as soon as its genome become available and the codes would be **stable**.

… for applications in molecular disease epidemiology … to assign codes based only on vertically inherited **core genomes** …

Marakeby *et al.* (February, 2014). *PLoS ONE* **9:** e89142 [PubMed].

Future Microbial NGS Developments and Genomic Nomenclature

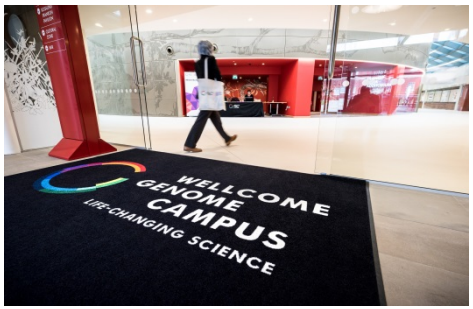
Dag Harmsen
University Münster, Münster, Germany
dharmsen@uni-muenster.de

Applied Bioinformatics & Public Health Microbiology

17-19 May 2017
Wellcome Genome Campus, Hinxton, UK

Registration will open soon!

For further information, contact: scientificconferences@wellcomegenomecampus.org



WELLCOME GENOME CAMPUS | ADVANCED COURSES AND SCIENTIFIC CONFERENCES

2nd ASM Conference on Rapid Applied Microbial Next-Generation Sequencing and Bioinformatic Pipelines

Date: Friday, September 15 – Monday, September 18, 2017
Place: Crystal Gateway Marriott, Crystal City, Virginia, US

Program Chairs: Dag Harmsen (Univ. Münster, DE), Jennifer Gardy (Univ. BC, CA)

Program Advisory Committee: Marc Allard and Eric Brown (both FDA, US), Elodie Ghedin (New York Univ., US), Paul Keim (Northern Arizona Univ., US), Duncan MacCannell (CDC, US), Adam Phillippy (NIH, US)

ASM Conference Committee Liaison: Gary Procop (Cleveland Clinic Foundation, US)

ASM Conferences Program Manager: Lisa Nalker (Washington DC, US)

Registration and abstract submission will open **April, 2017**.

Twitter: #ASMNGS

PATH NGenTrace

3nd Conference
Rapid Microbial NGS and Bioinformatics: Translation Into Practice

The event will gather experts from all over the world active in applying Next Generation Sequencing (NGS) techniques to discover the epidemiology, anti-microbial resistance, ecology and evolution of microorganisms. The program will be designed to build a bridge between software developers and end-users.

Base pairs 2 Bedside
Bacterial Genome Detectives

Date: May, 2018 (?)
Place: Hamburg, Germany (?)