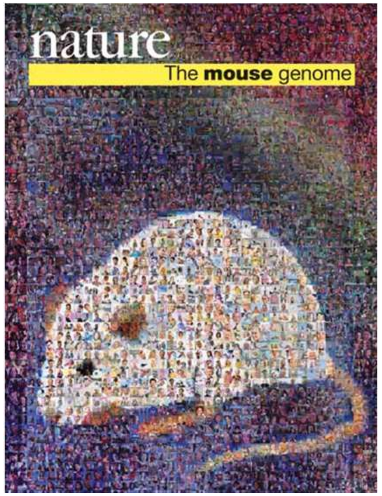
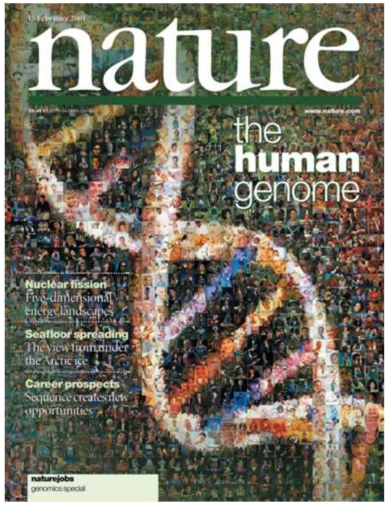
Next Generation Sequencing for Metagenomics

Genève, 13.10.2016

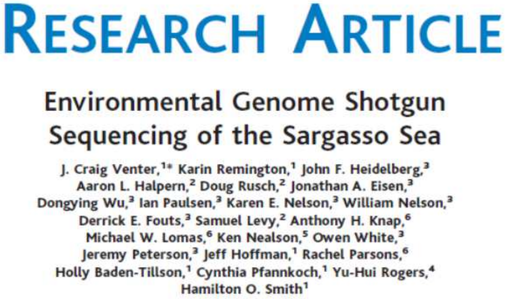
Patrick Wincker, Genoscope-CEA

Human and model organisms sequencing were initially based on the Sanger method





Sanger shotgun sequencing was used subsequently for environmental metagenomics



RESEARCH ARTICLE

Environmental Genome Shotgun Sequencing of the Sargasso Sea

J. Craig Venter,[1*] Karin Remington,[1] John F. Heidelberg,[3] Aaron L. Halpern,[2] Doug Rusch,[2] Jonathan A. Eisen,[3] Dongying Wu,[3] Ian Paulsen,[3] Karen E. Nelson,[3] William Nelson,[3] Derrick E. Fouts,[3] Samuel Levy,[2] Anthony H. Knap,[6] Michael W. Lomas,[6] Ken Nealson,[5] Owen White,[3] Jeremy Peterson,[3] Jeff Hoffman,[1] Rachel Parsons,[6] Holly Baden-Tillson,[1] Cynthia Pfannkoch,[1] Yu-Hui Rogers,[4] Hamilton O. Smith[1]

articles

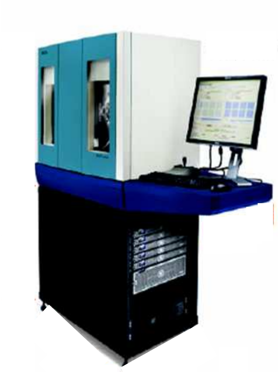
Community structure and metabolism through reconstruction of microbial genomes from the environment

Gene W. Tyson[1], Jarrod Chapman[3,4], Philip Hugenholtz[1], Eric E. Allen[1], Rachna J. Ram[1], Paul M. Richardson[4], Victor V. Solovyev[4], Edward M. Rubin[4], Daniel S. Rokhsar[3,4] & Jillian F. Banfield[1,2]

[1]*Department of Environmental Science, Policy and Management,* [2]*Department of Earth and Planetary Sciences, and* [3]*Department of Physics, University of California, Berkeley, California 94720, USA*
[4]*Joint Genome Institute, Walnut Creek, California 94598, USA*

2006-2011 : Sequencing by flux changes the scale and cost of genomics projects

All these methods use DNA amplification and synthesis but with different approaches
→ All produced short reads (100s bp)



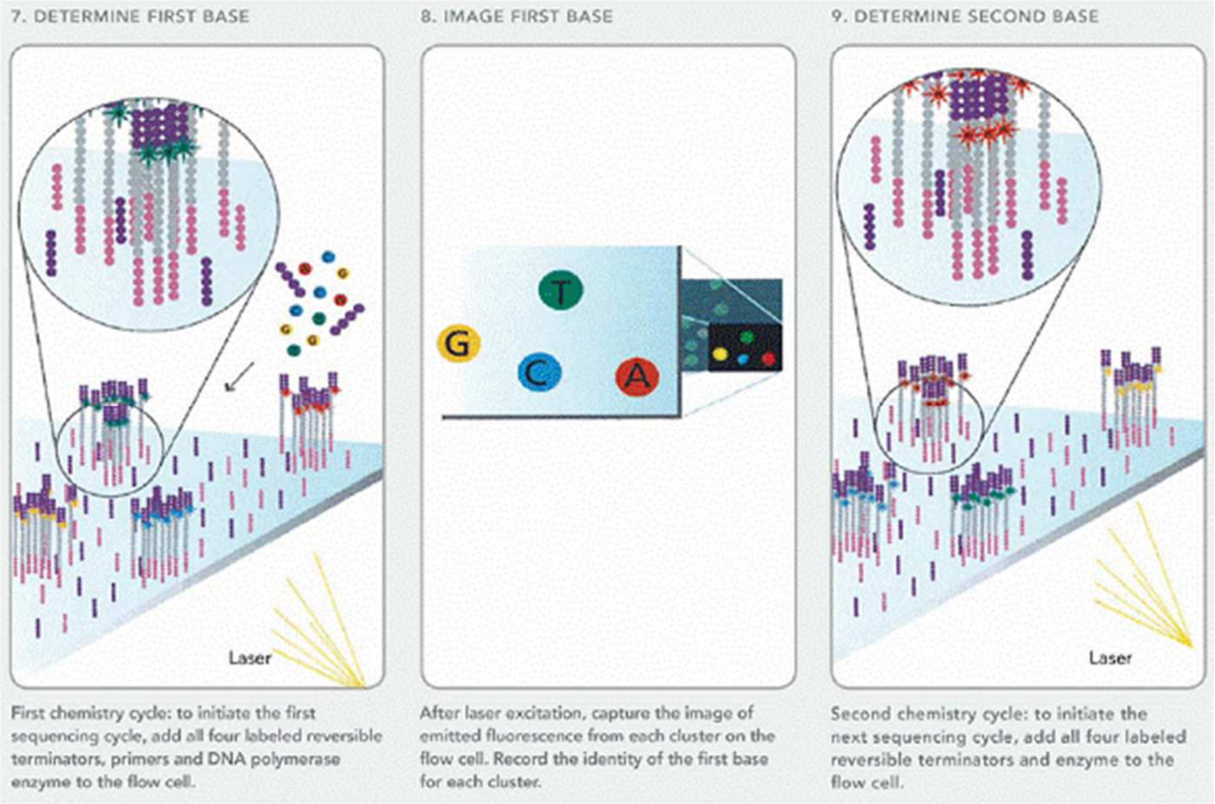
454
SEQUENCING

life
technologies™

illumina®

Complete Genomics
Powering large-scale
human genome studies

Illumina *Sequencing by Synthesis*



7. DETERMINE FIRST BASE

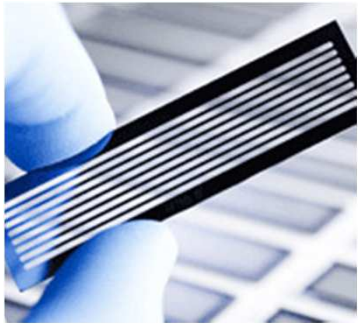
8. IMAGE FIRST BASE

9. DETERMINE SECOND BASE

First chemistry cycle: to initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.

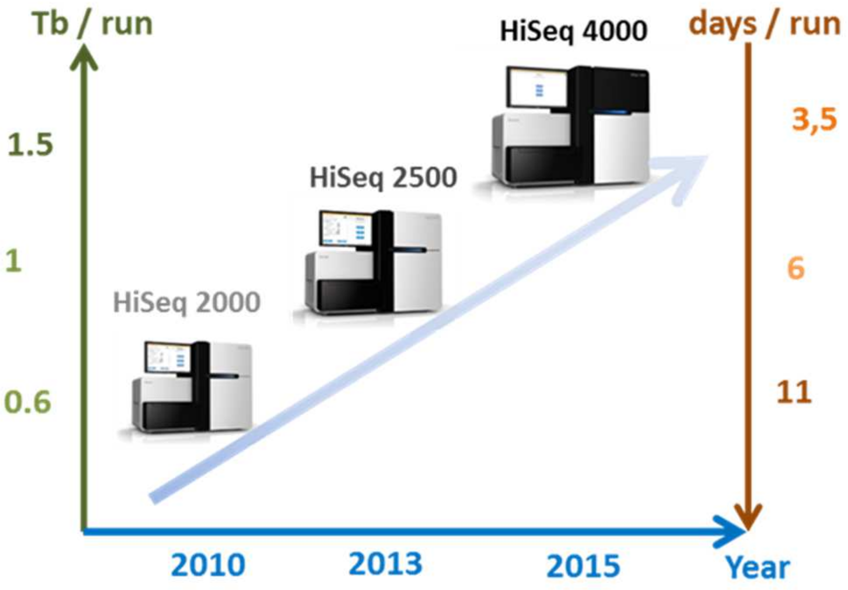
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

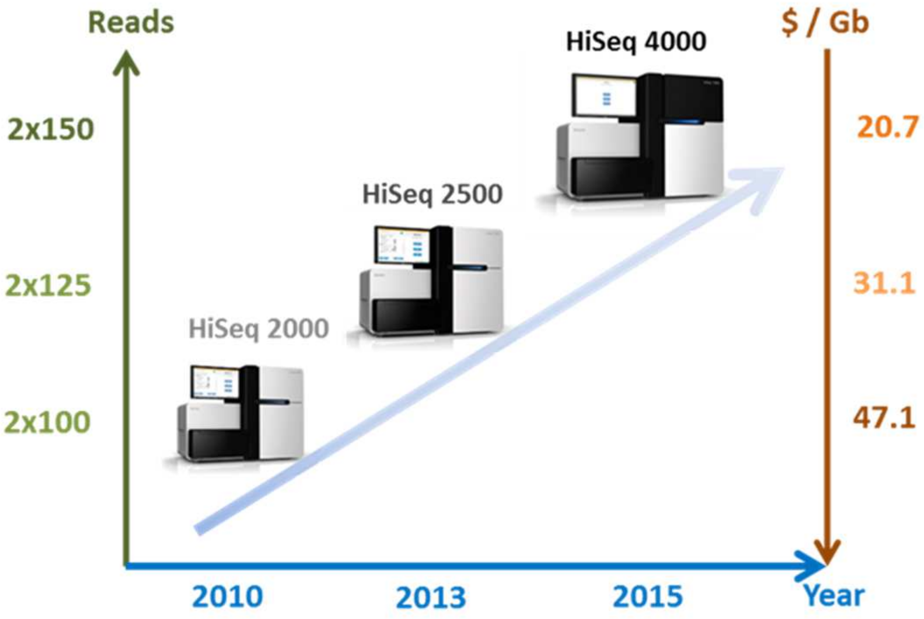
Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.



Very high throughput
Low error rate (<1%)
Short read length
High turnover time

Technological improvements at all scales are routine every year



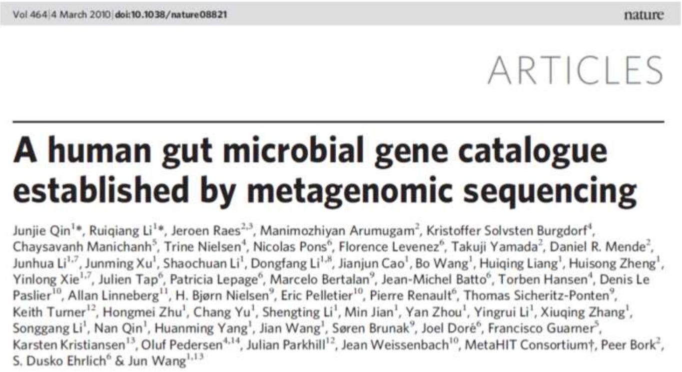


Sequencing by flux for Metagenomics

Deep coverage of complex microbial communities

High fragmentation of the individual species genomes

Method is efficient at building gene catalogues

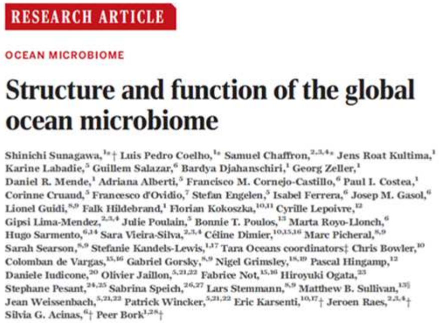


Vol 464 | 4 March 2010 | doi:10.1038/nature08821 nature

ARTICLES

A human gut microbial gene catalogue established by metagenomic sequencing

Junjie Qin[1]*, Ruiqiang Li[1]*, Jeroen Raes[2,3], Manimozhiyan Arumugam[2], Kristoffer Solvsten Burgdorf[4], Chaysavanh Manichanh[5], Trine Nielsen[4], Nicolas Pons[6], Florence Levenez[6], Takuji Yamada[2], Daniel R. Mende[2], Junhua Li[1,7], Junming Xu[1], Shaochuan Li[1], Dongfang Li[1,8], Jianjun Cao[1], Bo Wang[1], Huiqing Liang[1], Huisong Zheng[1], Yinlong Xie[1,7], Julien Tap[6], Patricia Lepage[6], Marcelo Bertalan[9], Jean-Michel Batto[6], Torben Hansen[4], Denis Le Paslier[10], Allan Linneberg[11], H. Bjørn Nielsen[9], Eric Pelletier[10], Pierre Renault[6], Thomas Sicheritz-Ponten[9], Keith Turner[12], Hongmei Zhu[1], Chang Yu[1], Shengting Li[1], Min Jian[1], Yan Zhou[1], Yingrui Li[1], Xiuqing Zhang[1], Songgang Li[1], Nan Qin[1], Huanming Yang[1], Jian Wang[1], Søren Brunak[9], Joel Doré[6], Francisco Guarner[5], Karsten Kristiansen[13], Oluf Pedersen[4,14], Julian Parkhill[12], Jean Weissenbach[10], MetaHIT Consortium†, Peer Bork[2], S. Dusko Ehrlich[6] & Jun Wang[1,13]



RESEARCH ARTICLE

OCEAN MICROBIOME

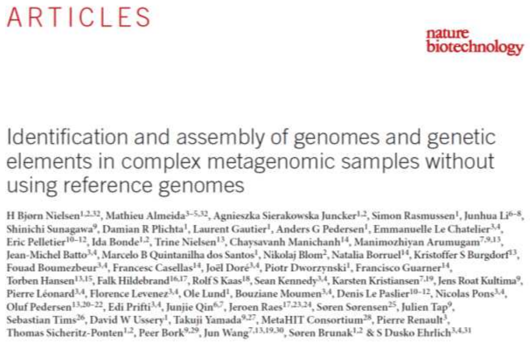
Structure and function of the global ocean microbiome

Shinichi Sunagawa,[1]*† Luis Pedro Coelho,[1]* Samuel Chaffron,[2,3,4]* Jens Roat Kultima,[1] Karine Labadie,[5] Guillem Salazar,[6] Bardya Djahanschiri,[1] Georg Zeller,[1] Daniel R. Mende,[1] Adriana Alberti,[5] Francisco M. Cornejo-Castillo,[6] Paul I. Costea,[1] Corinne Cruaud,[5] Francesco d'Ovidio,[7] Stefan Engelen,[5] Isabel Ferrera,[6] Josep M. Gasol,[6] Lionel Guidi,[8,9] Falk Hildebrand,[1] Florian Kokoszka,[10,11] Cyrille Lepoivre,[12] Gipsi Lima-Mendez,[2,3,4] Julie Poulain,[5] Bonnie T. Poulos,[13] Marta Royo-Llonch,[6] Hugo Sarmento,[6,14] Sara Vieira-Silva,[2,3,4] Céline Dimier,[10,15,16] Marc Picheral,[8,9] Sarah Searson,[8,9] Stefanie Kandels-Lewis,[1,17] Tara Oceans coordinators‡ Chris Bowler,[10] Colomban de Vargas,[15,16] Gabriel Gorsky,[8,9] Nigel Grimsley,[18,19] Pascal Hingamp,[12] Daniele Iudicone,[20] Olivier Jaillon,[5,21,22] Fabrice Not,[15,16] Hiroyuki Ogata,[23] Stephane Pesant,[24,25] Sabrina Speich,[26,27] Lars Stemmann,[8,9] Matthew B. Sullivan,[13]§ Jean Weissenbach,[5,21,22] Patrick Wincker,[5,21,22] Eric Karsenti,[10,17]† Jeroen Raes,[2,3,4]† Silvia G. Acinas,[6]† Peer Bork[1,28]†

Sequencing by flux for Metagenomics

Short-read sequencing brings opportunity to compare microbiome samples for gene content, but genomic context would be valuable to better interpret the data

Bioinformatics methods have been established to reconstruct genome fragments from metagenomes



ARTICLES

nature biotechnology

Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes

H Bjørn Nielsen[1,2,32], Mathieu Almeida[3–5,32], Agnieszka Sierakowska Juncker[1,2], Simon Rasmussen[1], Junhua Li[6–8], Shinichi Sunagawa[9], Damian R Plichta[1], Laurent Gautier[1], Anders G Pedersen[1], Emmanuelle Le Chatelier[3,4], Eric Pelletier[10–12], Ida Bonde[1,2], Trine Nielsen[13], Chaysavanh Manichanh[14], Manimozhiyan Arumugam[7,9,13], Jean-Michel Batto[3,4], Marcelo B Quintanilha dos Santos[1], Nikolaj Blom[2], Natalia Borruel[14], Kristoffer S Burgdorf[13], Fouad Boumezbeur[3,4], Francesc Casellas[14], Joël Doré[3,4], Piotr Dworzynski[1], Francisco Guarner[14], Torben Hansen[13,15], Falk Hildebrand[16,17], Rolf S Kaas[18], Sean Kennedy[3,4], Karsten Kristiansen[7,19], Jens Roat Kultima[9], Pierre Léonard[3,4], Florence Levenez[3,4], Ole Lund[1], Bouziane Moumen[3,4], Denis Le Paslier[10–12], Nicolas Pons[3,4], Oluf Pedersen[13,20–22], Edi Prifti[3,4], Junjie Qin[6,7], Jeroen Raes[17,23,24], Søren Sørensen[25], Julien Tap[9], Sebastian Tims[26], David W Ussery[1], Takuji Yamada[9,27], MetaHIT Consortium[28], Pierre Renault[3], Thomas Sicheritz-Ponten[1,2], Peer Bork[9,29], Jun Wang[7,13,19,30], Søren Brunak[1,2] & S Dusko Ehrlich[3,4,31]

In the absence of exhaustive reference genomes for uncultured microbes, high-continuity assemblies are requested from environmental shotgun data

Recent trends : back to high continuity

PACIFIC BIOSCIENCES

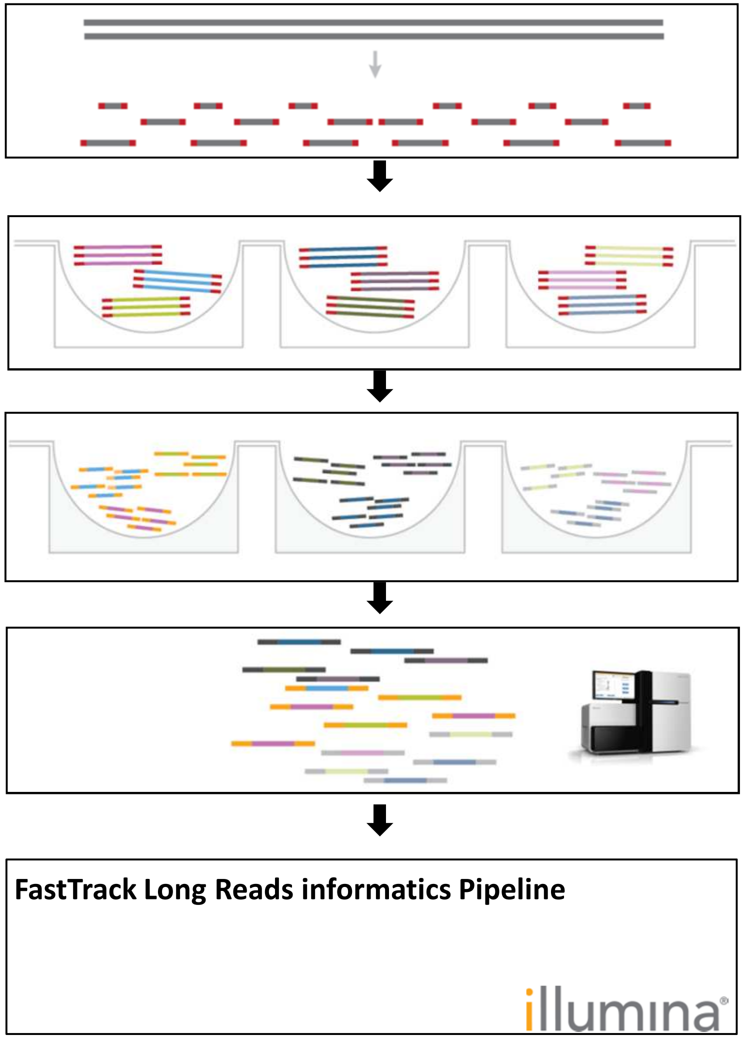
illumina
moleculo

Oxford NANOPORE Technologies

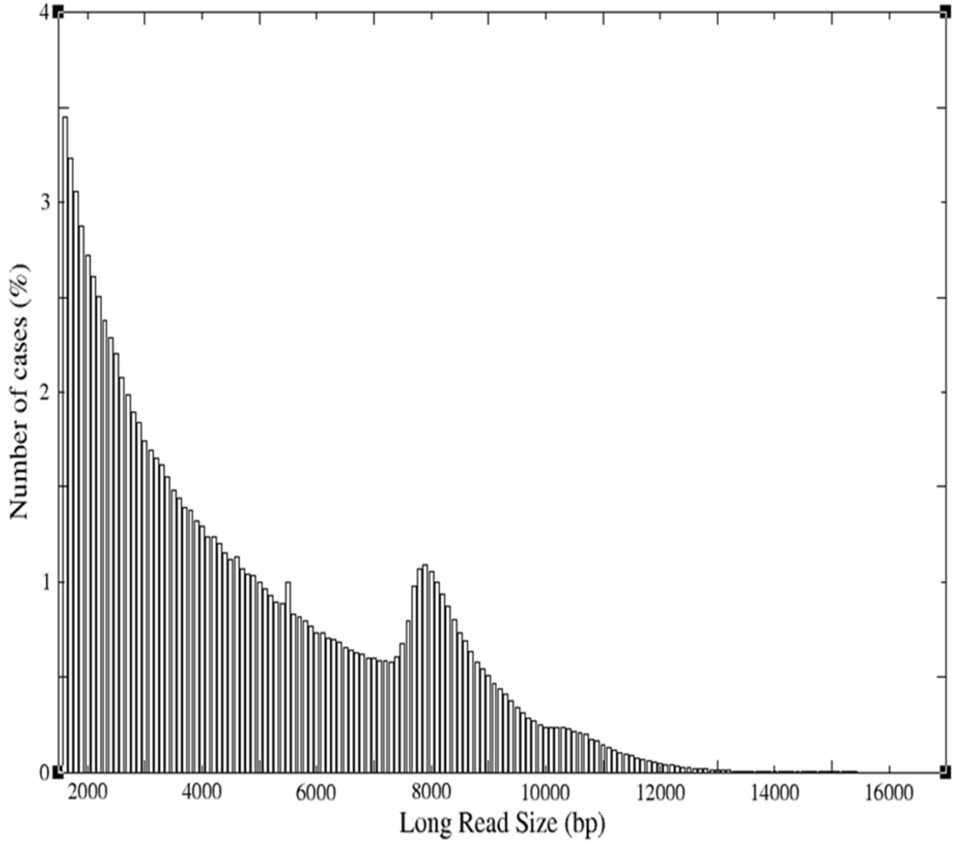
10x GENOMICS

Long reads or re-phased short reads

Synthetic long reads (« moleculo » approach)



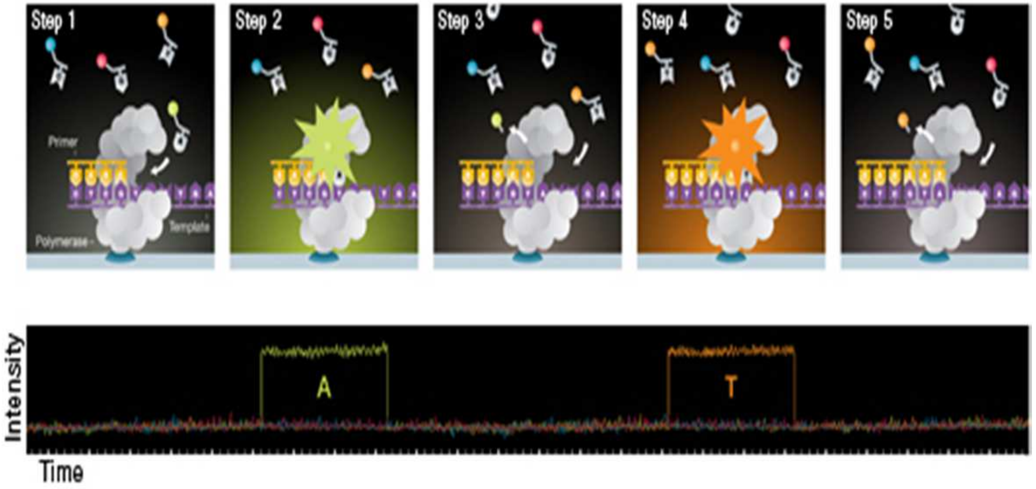
FastTrack Long Reads informatics Pipeline

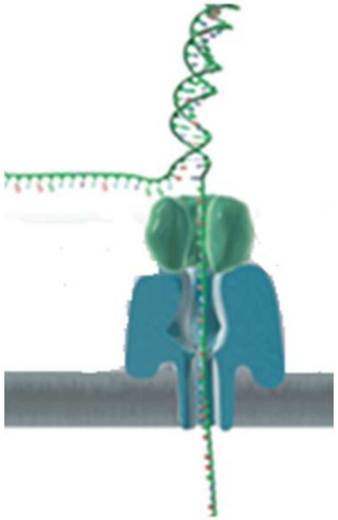


From 2011 : Real-time single molecule sequencing

PACIFIC BIOSCIENCES

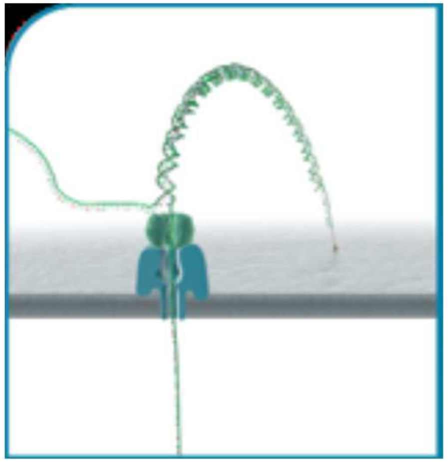
Oxford NANOPORE Technologies



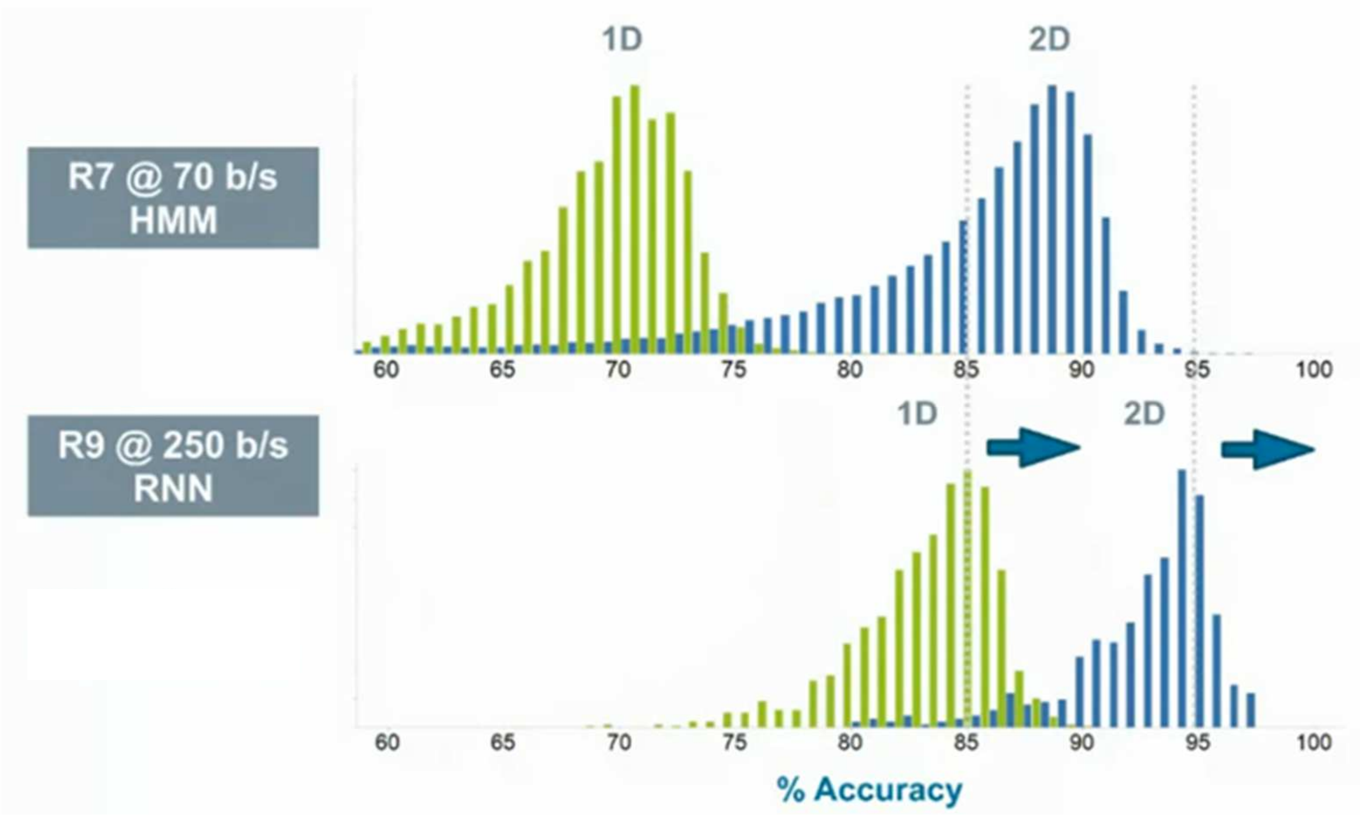


OXFORD NANOPORE TECHNOLOGIES: MINION

☑ **2014: start of the MinION platform**



MinIon : Improving accuracy

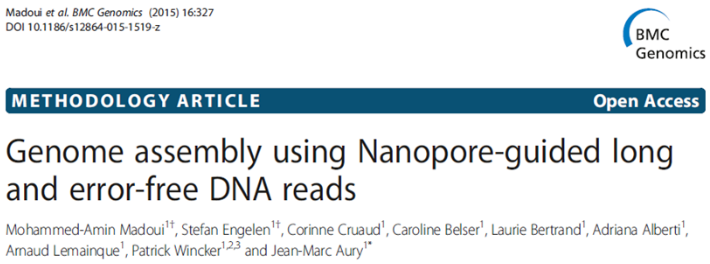


MIXING SHORT AND LONG READS FOR OPTIMIZED GENOME ASSEMBLIES

Long but error-prone sequences can be complemented by short accurate reads

Long reads provide continuity

Short reads may be used to correct or to re-assemble sequences based on the long reads



Madoui *et al. BMC Genomics* (2015) 16:327
DOI 10.1186/s12864-015-1519-z

BMC Genomics

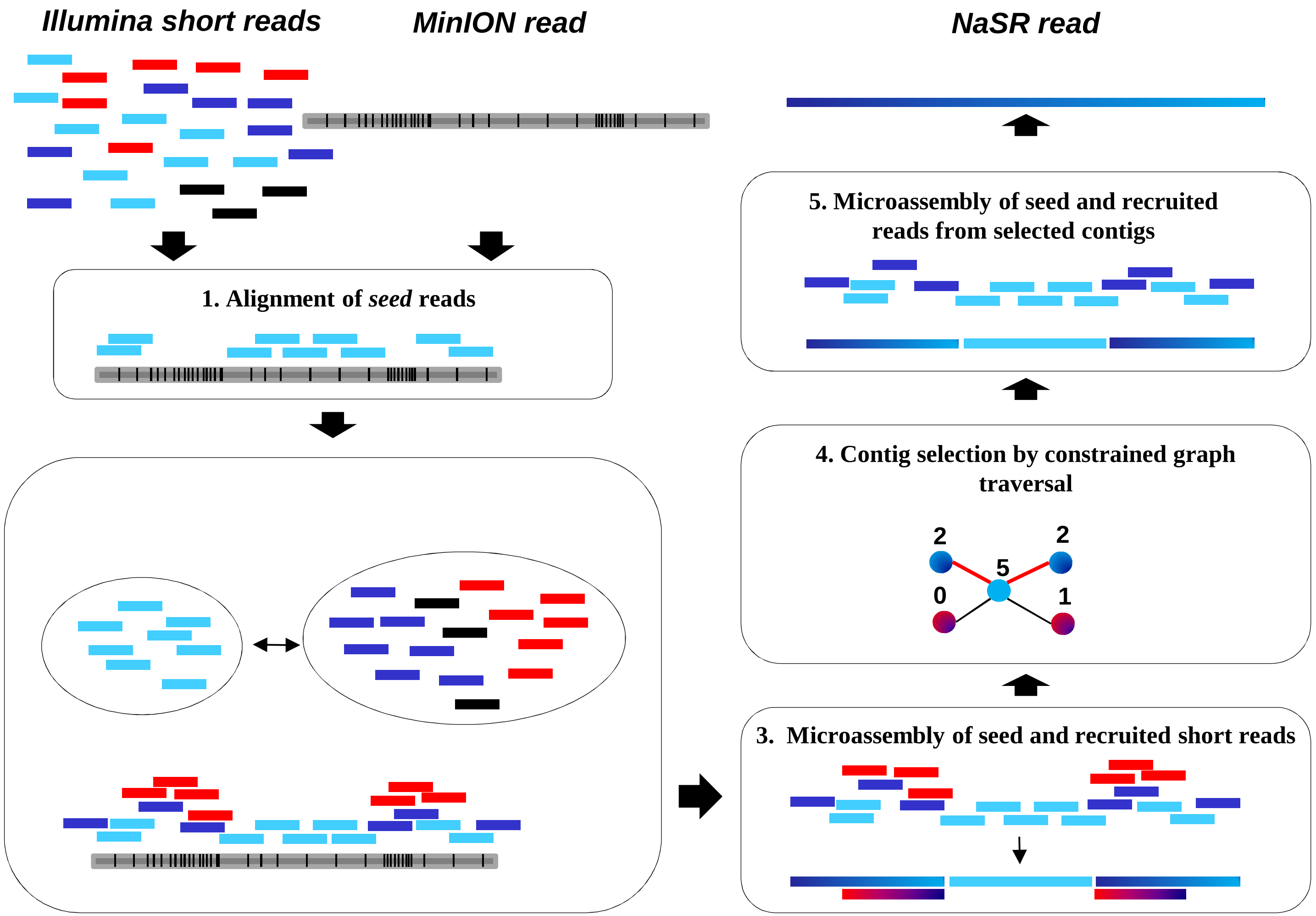
METHODOLOGY ARTICLE **Open Access**

Genome assembly using Nanopore-guided long and error-free DNA reads

Mohammed-Amin Madoui[1†], Stefan Engelen[1†], Corinne Cruaud[1], Caroline Belser[1], Laurie Bertrand[1], Adriana Alberti[1], Arnaud Lemainque[1], Patrick Wincker[1,2,3] and Jean-Marc Aury[1*]

97% of NaS reads have no error

A. baylyi genome assembled as a single contig



1. Alignment of *seed* reads

3. Microassembly of seed and recruited short reads

4. Contig selection by constrained graph traversal

5. Microassembly of seed and recruited reads from selected contigs

LONG READ TECHNOLOGIES AND METAGENOMICS

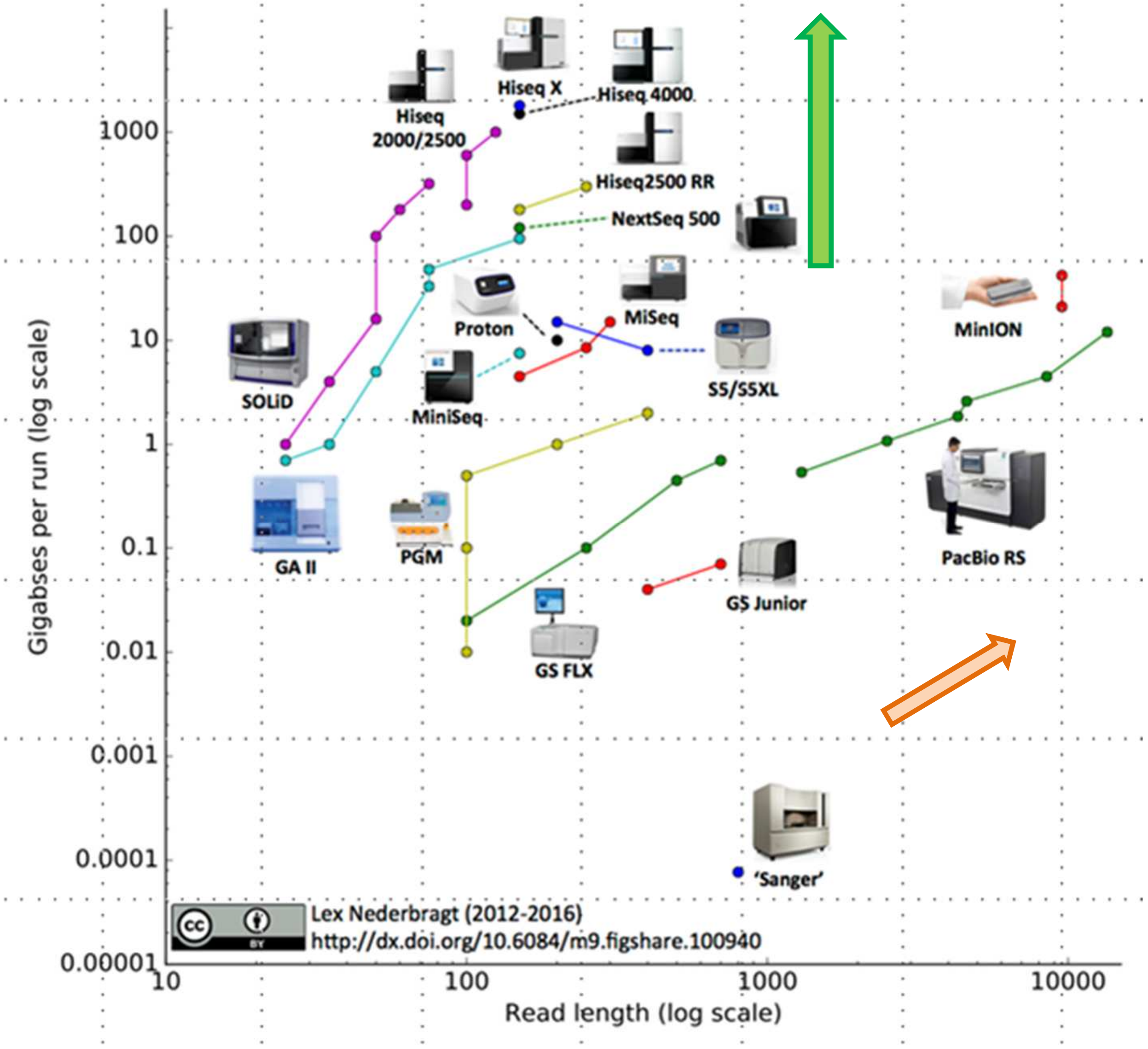
Still little use of long reads for complex metagenomes

Proof of concept for usefulness in reconstituting large genome regions from metagenome data

Main constraint is cost, particularly for studies with high sample numbers

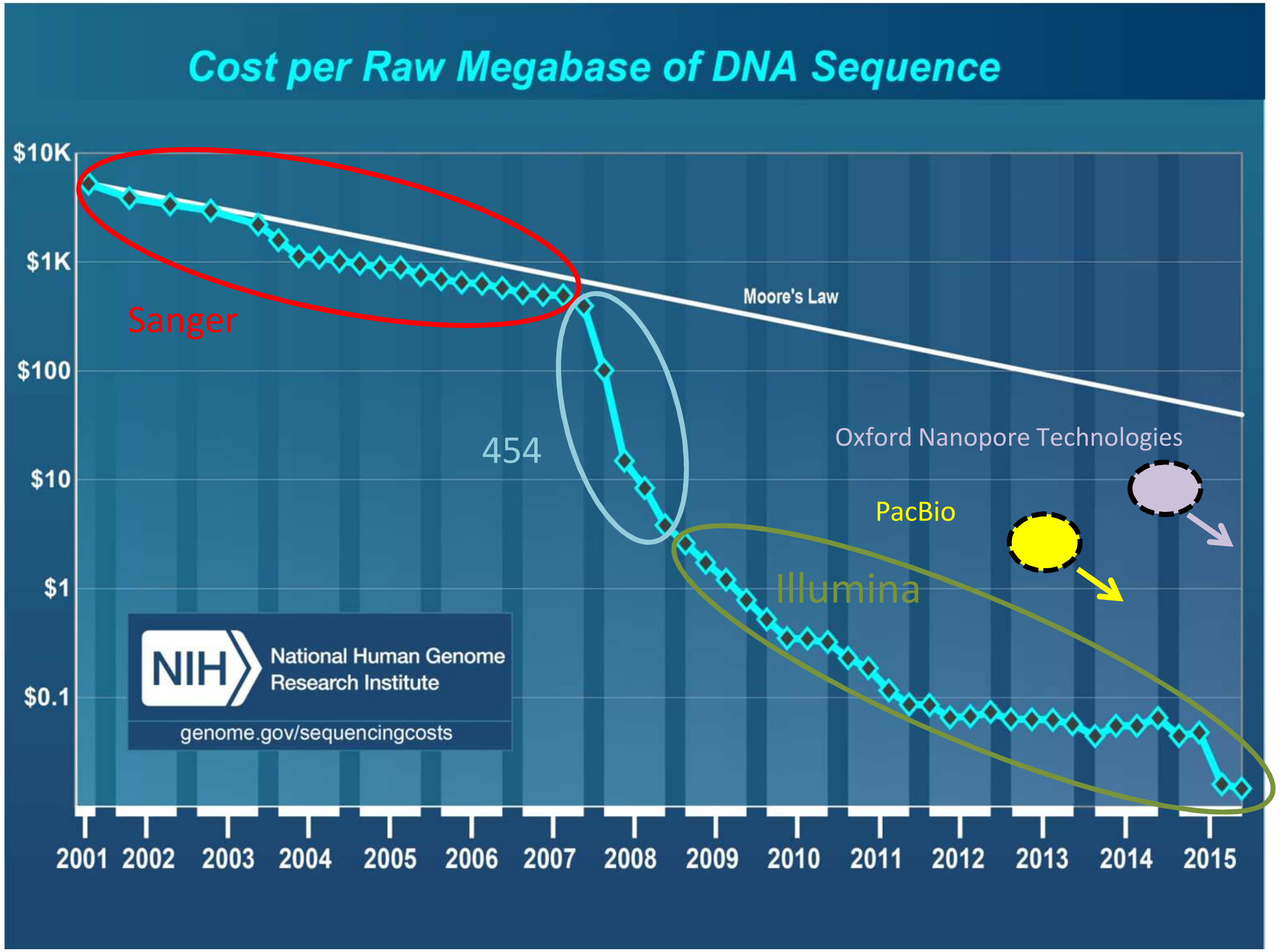
Bioinformatic tools for metagenomic assemblies are still underdeveloped for long error-prone sequences

Sequencing technologies : read length and throughput



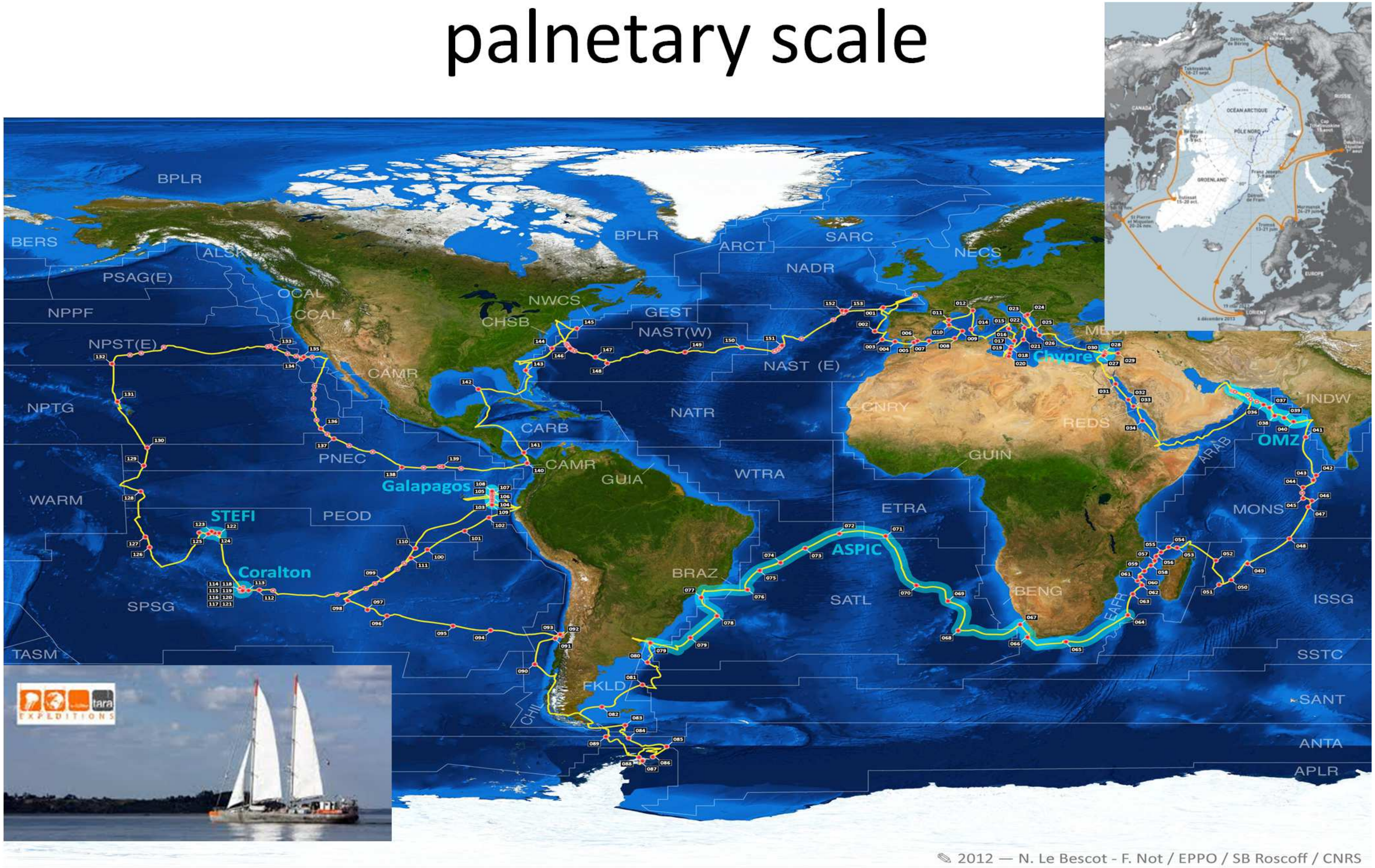
Lex Nederbragt (2012-2016)
http://dx.doi.org/10.6084/m9.figshare.100940

Evolution of sequencing costs



Sequencing complex microbiomes to near-completion ?

Deciphering the microbial gene content of the surface oceans at palnetary scale

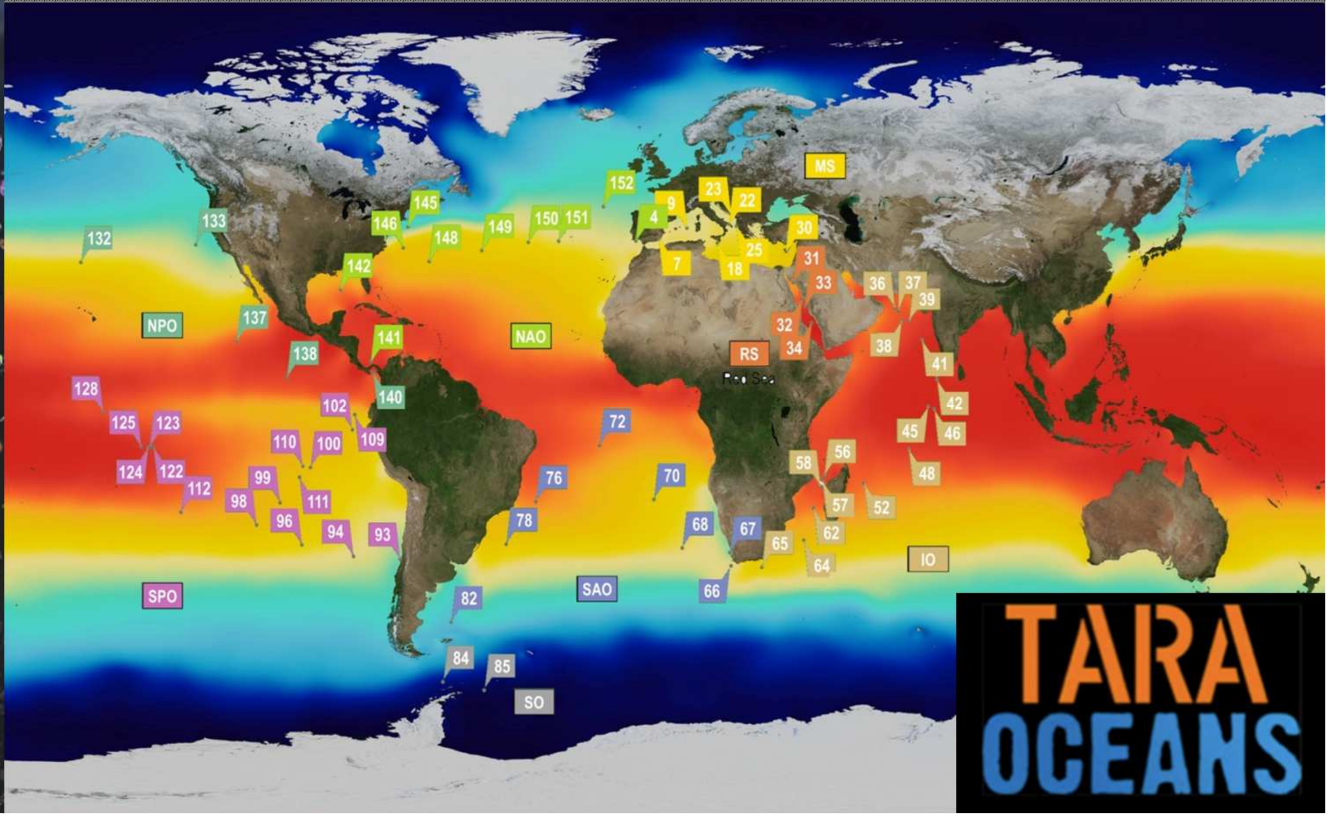


✎ 2012 — N. Le Bescot - F. Not / EPPO / SB Roscoff / CNRS

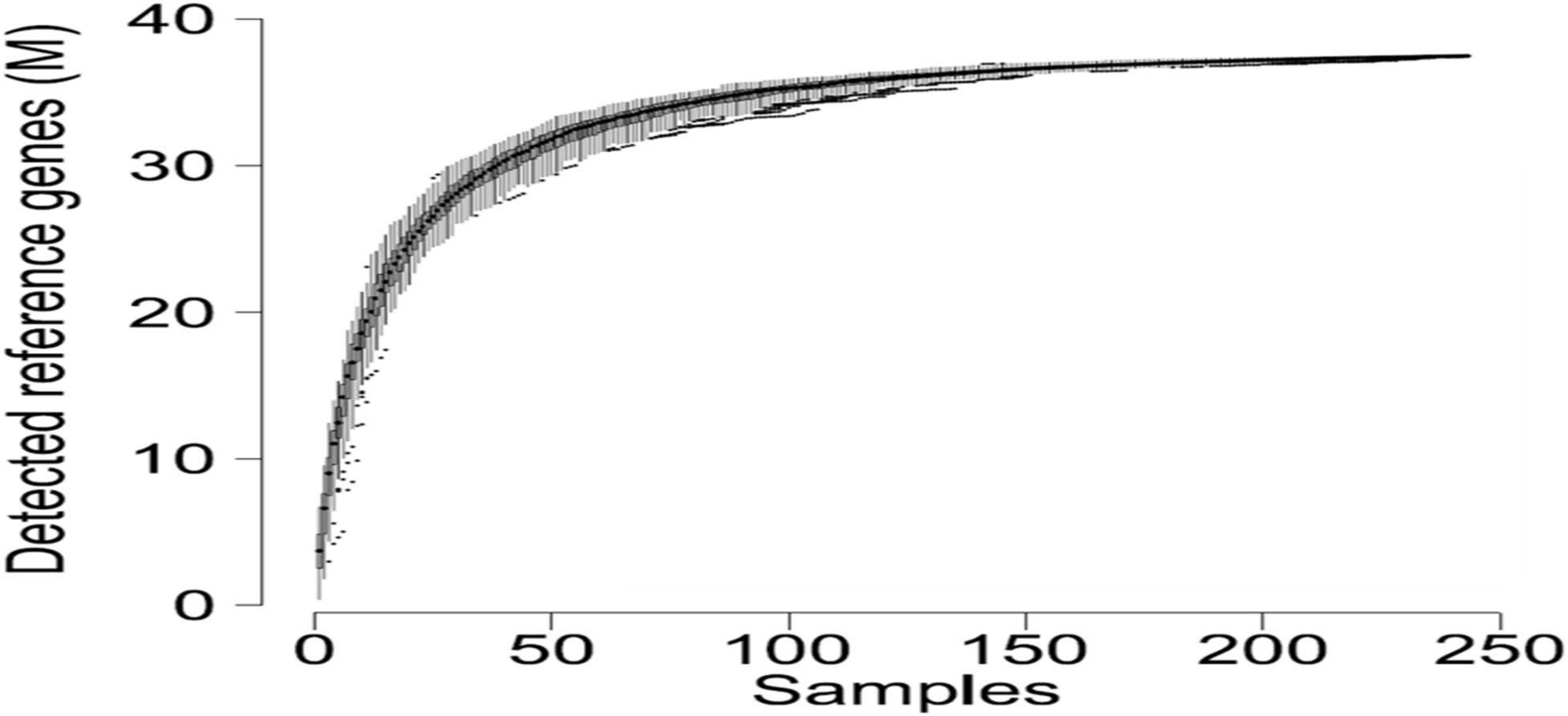
Ocean eco-systems biology

- 68 stations
- 3 depths
- 243 samples

7.2 Tbp DNA data
in the context of
the environment



The microbial gene content of the open ocean



Estimated diversity : 37,000 OTUs (97%)

S. Sunagawa et al.

CONCLUSION : SPECIES AND GENE CONTENT OF A COMPLEX ECOSYSTEM

… can be described using short-read massive sequencing for prokaryotes and viruses

The eukaryote gene content is the next target

Reconstructing genome context may be helped by new long read technologies

Acknowledgements

Sequencing team
Karine Labadie
Arnaud Lemainque
Corinne Cruaud

Development team
Valérie Barbe
Adriana Alberti
Julie Poulain

Informatics
Jean-Marc Aury
Eric Pelletier
Betina Porcel
Olivier Jaillon
Benjamin Noel
Yoann Seeleuthner
Stephane Engenlen
M. Amin Madoui
Caroline Belser
Arnaud Couloux

TARA OCEANS Consortium