

2nd CeBiTec
Symposium

July, 4-6
2007



**The Future of Genome
Research in the Light of
Ultrafast Sequencing
Technologies**

Center for Interdisciplinary
Research (ZiF) of Bielefeld University

Book of Abstracts



Welcome to the 2nd CeBiTec Symposium at the Center for Interdisciplinary Research (ZiF) of Bielefeld University

The Center for Biotechnology (CeBiTec) of Bielefeld University together with the German genome research network "Functional Genome Research on Bacteria Relevant for Agriculture, Environment and Biotechnology" funded by the Federal Ministry of Education and Research (BMBF) under the GenoMik-Plus guideline welcome you to the 2nd CeBiTec symposium. The international symposium which is entitled "**The Future of Genome Research in the Light of Ultrafast Sequencing Technologies**" belongs to a series of periodic conferences. The symposium is the follow-up of the 1st CeBiTec symposium which was dedicated to "**Molecular Systems Biology**" and was held at the ZiF from June 6th to 9th, 2006.

Just 30 years ago, two different methods for establishing the sequence of DNA molecules, the Sanger and the Maxam-Gilbert method were published. In the subsequent years, in particular the Sanger method was developed to a high-throughput method. Speed, data yield and cost reduction were constantly increased, resulting in the unexpectedly fast establishment of the human genome sequence in the year 2000.

For a short time, several high-throughput DNA sequence technologies are in operation which revolutionize our capabilities to decipher the blueprint of living organisms in an unprecedented speed, scale and cost-efficiency. These novel technologies will exert highest impact on research fields such as medicine. Moreover, technical developments are on the horizon which will lead to a paradigm shift in genome research once they are in place.

The major objectives of the 2nd CeBiTec symposium are to present a survey of current and future next generation sequencing technologies as well as the presentation of results arising from the use of these novel technologies in the various fields of genome research.

The symposium will feature exhibition booths of the distributors of these novel technologies, i.e. Roche Diagnostics, Applied Biosystems and Illumina Inc. as well as GATC Biotech as a service provider. In this context, the organizers gratefully acknowledge the sponsorship of Roche Diagnostics, Applied Biosystems and Illumina Inc., respectively. Looking forward to welcome you in Bielefeld to an exciting meeting!

Alfred Pühler (Chair of the organizing committee)

Program

The Future of Genome Research in the Light of Ultrafast Sequencing Technologies

Organizing committee:

- R. Giegerich (Bielefeld University, D)
- A. Pühler (Bielefeld University, D/Chair)
- S. Schuster (PSU, USA)
- Ch. Sensen (University of Calgary, CA)
- J. Stoye (Bielefeld University, D)
- B. Weisshaar (Bielefeld University, D)

Wednesday, July 4th, 2007

12.00 Registration + Snacks

Opening Session

13.30 Welcome address - Introduction

Alfred Pühler, CeBiTec, Bielefeld University, Germany

Ipke Wachsmuth, ZiF, Bielefeld University, Germany

14.00 Mammoth Genomics

Stephan C. Schuster, Pennsylvania State University, U.S.A.

14.45 Genome Sequencer FLX: next generation sequencing technology that enables the sequencing of personal human genomes

Maithreyan Srinivasan, 454 Life Sciences/Roche Diagnostics, Branford, U.S.A.

15.15 How projects benefit from the integration of (ultra-)short to long reads

Richard Reinhardt, Max-Planck-Institute of Molecular Genetics, Berlin, Germany

15.45 Coffee break

Session Technologies

16.15 The Genome Sequencer FLX System from 454/Roche – The 2nd generation sequencing system offering the broadest range of applications

Marcus Dröge, Roche Diagnostics GmbH, Penzberg, Germany

16.45 Enabling genome research with Solexa sequencing

Kevin Hall, Illumina Inc., Essex, UK

17.15 The SOLiD™ system: a technology enabling new applications

Michael Rhodes, Applied Biosystems, Foster City, U.S.A.

17.45 Real-time DNA sequencing

Susan Hardin, VisiGen Inc., Huston, Texas, U.S.A.

18:20 Greetings of the Rector of Bielefeld University

Dieter Timmermann, Bielefeld University, Germany

18:30 Mixer and Exhibition

Thursday, July 5th, 2007

Session Bioinformatics

- 09.00 Two base encoding in the SOLiD™ sequencing system**
Michael Rhodes, Applied Biosystems, Foster City, U.S.A.
- 09.30 Parallel scalable assembly of mixed sequencing types using Forge**
Darren Platt, Joint Genome Institute, Walnut Creek, U.S.A.
- 10.00 Assembly and annotation of short read data sets**
Heinz Himmelbauer, Max-Planck-Institute of Molecular Genetics, Berlin, Germany
- 10.30 Coffee break**

Session Bioinformatics (continued)

- 11:00 Forging Ahead: testing a hybrid genome assembler**
Dirk Evers, CeBiTec, Bielefeld University, Germany
- 11:30 MEGAN analysis of metagenomic data**
Daniel Huson, Tübingen University, Germany
- 12.00 Computational characterization of short environmental DNA fragments**
Jens Stoye, CeBiTec, Bielefeld University, Germany
- 12.30 Lunch**

Session Microbial Genomics + Metagenomics

- 14.00 New sequencing strategies for microbes**
Paul Richardson, Joint Genome Institute, Walnut Creek, U.S.A.
- 14.30 Ultrafast *de novo* sequencing of the human pathogen *Corynebacterium urealyticum* with the Genome Sequencer 20 System**
Andreas Tauch, Bielefeld University, Germany
- 15.00 Access to the plasmid mobilome of wastewater treatment plant bacteria by applying the ultra-fast 454-sequencing technology**
Andreas Schlüter, Bielefeld University, Germany
- 15.30 Next Generation Sequencing: comparison of the technologies for bacterial genome sequencing**
Kerstin A. Stangier, GATC, Konstanz, Germany
- 15.45 Coffee break**

Session Plant Genomics

- 16.15 *Vitis vinifera* cv. Pinot Noir: dealing with heterozygous genomes by merging Sanger and pyrosequencing methods**
Michela Troglio, San Michelle all'Adige, Italy
- 16.45 Profiling plant transcriptomes by massively-parallel pyrosequencing**
Andreas Weber, Düsseldorf University, Germany
- 18.00 Reception and Conference Dinner**

Friday, July 6th, 2007

Session Human and Animal Genomics

- 09.00 Charting and sequencing structural variation using high resolution paired-end mapping (HR-PEM)**
Jan Korbel, Yale University, New Haven, U.S.A.
- 09.30 Mutation detection in cancer: parallel sequencing and mass-spectrometric genotyping for sensitive and high-throughput analysis of human cancer mutation space**
Roman Thomas, Max-Planck Institute for Neurological Research, Cologne, Germany
- 10.00 Chronic inflammatory diseases of barrier organs: from genetic etiology to genomic pathophysiology**
Philip Rosenstiel, Institut für Klinische Molekularbiologie, Kiel, Germany
- 10.30 Coffee break**

Session Human and Animal Genomics (continued)

- 11:00 Mutation discovery using SOLiD sequencing for the identification of gene knockouts in model organisms**
Edwin Cuppen, Hubrecht Institute, Utrecht, The Netherlands
- 11:30 Pair-end-ditag (PET) based sequencing approach for transcriptome analysis and genome annotation**
Chia-Lin Wei, Genome Institute of Singapore, Singapore
- 12:00 Neanderthal genomics**
Johannes Krause, Max-Planck Institute for Evolutionary Anthropology, Leipzig, Germany
- 12.30 Concluding remarks**

ORAL PRESENTATIONS

Contribution

Page

The Future of Genome Research in the Light of Ultrafast Sequencing Technologies at the Center for Biotechnology at Bielefeld University

Alf Pühler

1

Mammoth Genomics

Stephan C. Schuster

2

Genome Sequencer FLX: Next Generation Sequencing Technology that enables the sequencing of personal human genomes

Maithreyan Srinivasan

4

How projects benefit from the integration of (ultra-)short to long reads

Richard Reinhardt

5

The Genome Sequencer FLX System from 454/Roche –The 2nd generation sequencing system offering the broadest range of applications

Marcus Dröge

6

Enabling genome research with Solexa sequencing

Kevin P Hall

7

The SOLiD™ System : a technology enabling new applications

Michael Rhodes

8

Real-time DNA Sequencing

Susan Hardin

9

Two base encoding in the SOLiD™ Sequencing System Michael Rhodes	10
Parallel scalable assembly of mixed sequencing types using Forge Darren Platt	11
Assembly and annotation of short read data sets Heinz Himmelbauer	12
Forging Ahead: Testing a Hybrid Genome Assembler Dirk J. Evers	13
MEGAN analysis of metagenomic data Daniel Huson	14
Computational characterization of short environmental DNA fragments Jens Stoye	15
New sequencing strategies for microbes Paul Richardson	16
Ultrafast <i>de novo</i> sequencing of the human pathogen <i>Corynebacterium urealyticum</i> with the Genome Sequencer 20 System Andreas Tauch	18
Access to the plasmid mobilome of wastewater treatment plant bacteria by applying the ultra-fast 454-sequencing technology Andreas Schlüter	20
Next Generation Sequencing: Comparison of the technologies for bacterial genome sequencing Kerstin A. Stangier	22
<i>Vitis vinifera</i> cv. Pinot Noir: dealing with heterozygous genomes by merging Sanger and pyrosequencing methods Michela Troggio	24

Profiling plant transcriptomes by massively-parallel pyrosequencing Andreas P.M. Weber	26
Charting and Sequencing Structural Variation using High-Resolution Paired-End Mapping (HR-PEM) Jan Korbel	28
Mutation detection in cancer: parallel sequencing and mass-spectrometric genotyping for sensitive and high-throughput analysis of human cancer mutation space Roman Thomas	30
Chronic inflammatory diseases of barrier organs: From genetic etiology to genomic pathophysiology Philip Rosenstiel	31
Mutation discovery using SOLiD sequencing for the identification of gene knockouts in model organisms Edwin Cuppen	32
Pair-end-ditag (PET) based Sequencing Approach for Transcriptome Analysis Chia-Lin Wei	33
Neanderthal Genomics Johannes Krause	34

The Future of Genome Research in the Light of Ultrafast Sequencing Technologies at the Center for Biotechnology at Bielefeld University

Alfred Pühler

Department of Genetics and Center for Biotechnology, Bielefeld University

The Center for Biotechnology (CeBiTec) at Bielefeld University hosts the Institute for Bioinformatics and the Institute for Genome Research and Systems Biology. One of the larger research projects carried out by these institutes concerns the “Functional Genome Research on Bacteria Relevant for Agriculture, Environment and Biotechnology”. In the frame of this research project nearly ten bacterial genomes have been sequenced by the conventional Sanger technology. Last year, several sequencing projects have been carried out by using the 454 sequencing technology. Two of these projects, namely the sequencing of the genome of *Corynebacterium urealyticum* and the sequencing of a metagenome representing plasmid DNA isolated from a bacterial community residing in a wastewater treatment plant will be presented during this symposium. A third project dealing with *de novo* 454 sequencing of the genome of the agriculturally important *Sinorhizobium meliloti* strain SM11 has been started. In the meantime, the sequence of the megaplasmid pSymA has been completed. Gap closure was achieved by using a fosmid library and the bioinformatics tool BACCardI. The strategy for *de novo* sequencing of bacterial genomes employing 454 pyrosequencing will be discussed.

Mammoth Genomics

Stephan C. Schuster

Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, USA

Next-generation sequencing in its initial phase is limited by shorter read-length than the more mature Sanger capillary-sequencing. While this is considered a short-coming by some, its bias-free sequencing without any cloning or amplification biases made it the method of choice for environmental or ancient DNA projects. This is in particular true as ancient DNA is restricted in a post-mortem degenerative process to short fragments that matches the read-length of the currently available technology. We will describe the use of the sequencing-by-synthesis approach on DNA samples that have been extracted from mammoth fossils.

References:

H. N. Poinar, C. Schwarz, Ji Qi, B. Shapiro, R. D. E. MacPhee, B. Buigues, A. Tikhonov, D. H. Huson, L. P. Tomsho, A. Auch, M. Rampp, W. Miller, S. C. Schuster, Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA, *Science* 311:392-394, 2006.

MT Gilbert, Binladen J, Miller W, Wiuf C, Willerslev E, Poinar H, Carlson JE, Leebens-Mack JH and S.C. Schuster, Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis, *Nucleic Acids Res.* 35(1), 1-10. Epub 2006 Aug 18.

D.H. Huson, A. Auch, Ji Qi and S.C. Schuster, MEGAN Analysis of Metagenomic Data, *Genome Research*,17:377-386, 2007.

Genome Sequencer FLX: Next Generation Sequencing Technology that enables the sequencing of personal human genomes

Maithreyan Srinivasan

454 Life Sciences/Roche Diagnostics, Branford, CT 06405

454 Sequencing™ (pyrophosphate-based sequencing optimized for beads in picoliter wells) involves subjecting template DNA to solid-phase amplification on to beads in emulsions. Beads that contain amplified template DNA are deposited into the wells of a picotiter plate (fused fiber-optic reaction vessels arranged in tandem) and 454 Sequencing™ is performed to obtain reads with average read lengths of 250-bases. The 2nd generation Roche/454 sequencer, GSFLX, is capable of generating 400000 reads in an 8 hour run. We have used the GSFLX to sequence the first individual human genome – the genome of DNA pioneer and Nobel Laureate, James D. Watson. The presentation will review what we have learned from sequencing an individual genome including summary statistics such as coverage of the genome and accuracy of the reads. With the first unbiased human genome sequence in hand, we will be able to describe the degree of completion of the human genome assembly as well as ascertain the degree of individual human variation.

How projects benefit from the integration of (ultra-)short to long reads

Richard Reinhardt, Wei Chen, Peter Marquardt, Bernd Timmermann,
Aleksey Soldatov

Max-Planck-Institut for Molecular Genetics, D-14195 Berlin-Dahlem

An overview of the capability of the Illumina/Solexa and Roche /454 systems will be presented and specifically describing the integration into the overall data network of the institute with special demands for the Solexa instrument.

Data from some projects, so far evaluated on the systems will be shown and discussed together with the integration of standard capillary sequence reads.

The overall capacity of the systems versus the cost discussion will at least be presented.

The Genome Sequencer FLX System from 454/Roche – The 2nd generation sequencing system offering the broadest range of applications

Marcus Dröge

Roche Diagnostics GmbH, Prenzberg, Germany

In December 2006, Roche has launched its next generation system, the Genome Sequencer FLX system. FLX stands for Flexibility and reflects the outstanding versatility of the system. Due to its unique combination of read length and reads per run and its significantly improved sequence accuracy, the broadest variety of break-through applications can be addressed. It offers applications for research fields such as cancer research, genetic diseases, infectious diseases, plant genomics, metagenomics and many more.

The presentation will give an overview about applications possible with the Genome Sequencer FLX, with special emphasis on transcriptome and gene regulation studies.

Enabling genome research with Solexa sequencing

Kevin P Hall

Illumina Inc. Chesterford Research Park, Cambridge, UK, Industrial Blvd.
Hayward, CA & 9885 Towne Centre Drive, San Diego, CA

DNA sequence information underpins human genetic research, enabling discoveries of important biological or medical benefit. However, the routine application of sequencing in everyday research has been limited by the cost and capacity of existing methods. New technologies have the potential to meet the challenge of generating more than a billion bases of highly accurate raw data per experiment at very low cost. For example, in targeted or whole genome sequencing approximately 30 million 30-40 base reads are generated per flowcell using Solexa technology and aligned to a suitable reference sequence for detection of insertions and deletions, and copy number variants, as well as providing information to assist de novo sequence assembly in the absence of a reference sequence.

High-throughput sequencing on this scale enables the routine acquisition of novel information in genomic sequencing studies, for example to search for causative variants in regions associated with disease, or to establish complete catalogues of sequence variation for population studies. Short read sequencing is equally applicable to tag-based analyses of DNA or RNA that leads to discovery and annotation of new genomic features including genes, regulatory sites in chromatin, and digital gene expression information. The precision and depth of our knowledge of human and other genomes will increase rapidly as we enter a new era of medical and biological research with comprehensive datasets of accurate, low-cost sequence information.

The SOLiD™ System : a technology enabling new applications

Michael Rhodes

Applied Biosystems, 850 Lincoln Centre Dr., Foster City, CA 94404

The decreasing cost of sequencing is driving the development of genome-wide applications based on sequence data. Instead of needing different platforms for applications such as gene expression, genotyping and sequencing, it is becoming possible to use sequencing as the universal currency for most, if not all, genetic analysis. The SOLiD™ System is a ligation-based, massively parallel sequencing system that can currently generate in excess of 1 Gb of DNA sequence per run. Investigators can elect to use either fragment or mate-paired libraries and can run one or multiple samples at a time. We are taking advantage of the high throughput and system flexibility of the SOLiD™ system to collaborate with a number of scientific groups to develop a range of applications. An update on application development and examples of the data generated will be presented.

Real-time DNA Sequencing

Susan Hardin and The VisiGen Development Team

VisiGen Biotechnologies, Inc., 2575 West Bellfort, Suite 250, Houston, Texas
77054 USA

VisiGen Biotechnologies, Inc, is developing a sequencing platform that will enable comprehensive genome analysis. Cutting-edge technologies, including single-molecule detection, fluorescent molecule chemistry, computational biochemistry, and biomolecule engineering and purification, are combined to create this revolutionary platform. VisiGen scientists have engineered polymerase and nucleotides to act together as direct molecular sensors of DNA sequence information during DNA synthesis. As a nucleotide is incorporated into the nascent DNA strand, energy transfers from the excited donor fluorophore attached to the polymerase to the acceptor fluorophore bonded to the nucleotide's gamma-phosphate, thereby stimulating the emission and detection of a base-type-specific signature. Because the acceptor fluorophore is naturally removed during nucleotide incorporation, VisiGen's strategy enables a real-time approach to sequence determination. The technology is scalable: these nanosequencing machines are monitored in massively parallel arrays to produce a sequencing platform that will be capable of collecting sequence data at rates approaching 1 million bases per second.

Two Base Encoding in the SOLiD™ Sequencing System

Michael Rhodes, Alan Blanchard and Heather Peckham

Applied Biosystems, 850 Lincoln Centre drive, Foster City CA

The SOLiD (Sequencing by Oligonucleotide Ligation and Detection) uses a unique ligation chemistry, utilizing 8 base ligation probes. Instead of measuring incorporation of a single base, ligation of a specific di-nucleotide (specifically, bases 4, 5 in our system) is measured using four fluorescent dyes. This results in each fluorescent signal representing four potential di-nucleotides. This methodology has the advantage that every base is interrogated twice, thus improving basecalling accuracy over single interrogation. As a consequence it is possible to determine measurement errors from a real polymorphism. By identifying error it is possible to carry out analysis excluding miscalled signals. This has been demonstrated in practice to lead to greater than a twenty fold decrease in error. Data will be presented on the 2 base encoding schema and the consequences of this schema. The ability to use color space to reduce errors will be discussed as well as some of the requirements to go from color space to base space. Data will be presented showing how two base encoding improved accuracy over single base encoding.

Parallel Scalable Assembly of Mixed Sequencing Types using Forge

Darren Platt⁽¹⁾, and Dirk Evers⁽²⁾

⁽¹⁾ (Joint Genome Institute Walnut Creek, USA),

⁽²⁾ (University of Bielefeld)

Changing sequencing technologies are simultaneously putting many different pressures on genome assembly algorithms. Read lengths have been getting considerably shorter, there are new error models and the overall volume of data to be handled is rapidly growing. We describe a set of modifications to the genome assembly algorithm Forge that enable it to assemble mixtures of Sanger, 454, Illumina and SOLiD sequencing reads. The implementation is based on MPI, the parallel programming standard and can utilize the memory across many machines to handle large genome assemblies. We will describe statistical considerations for dealing with read mixtures, error handling issues specific to particular sequence types and challenges detecting overlaps in very short reads. Forge has been a run on a variety of modern and ancient DNA sets and different assembly results and issues will be demonstrated with these sets.

Assembly and annotation of short read data sets

Heinz Himmelbauer¹, Juliane Dohm¹, Claudio Lottaz^{1,2}

1 Max-Planck-Institute for molecular Genetics, Berlin, Germany

2 University of Regensburg, Germany

Short-read sequencing technologies allow the rapid production of large sequence data sets at low costs. Their availability is just about to show its impact on everyday experimental work in the lab. Applications will encompass transcriptome characterization, mutation discovery, high-resolution genetic mapping and haplotyping, as well as genomic sequencing. We illustrate some of the emerging applications of short-read technologies for transcriptomics and genomics. A particular challenge is assembling genomic sequences *de novo* using short-read data. We have therefore implemented SHARCGS (SHort-read Assembler based on Robust Contig-extension for Genome Sequencing), an algorithm to permit *de novo* assemblies from 25-50mer read data sets. We tested the efficiency of SHARCGS in simulations on BAC inserts, yeast chromosomes, and bacterial genomes. In addition, the performance of SHARCGS was evaluated using real Solexa/Illumina data. Concepts and present status of the work will be outlined.

Forging Ahead: Testing a Hybrid Genome Assembler

Dirk J. Evers

CeBiTec – Center for Biotechnology, Bielefeld University

Currently, new DNA sequencing technology is reaching the market with the promise of significant improvements in overall cost, speed, and quantity (Roche/454, Illumina/Solexa, Applied Biosystems, and others). On the downside are considerably shorter average read lengths and mostly unpaired reads. The forge genome assembler is capable of combining conventional paired Sanger reads with single short reads in one hybrid assembly.

All phases of forge are implemented as parallel algorithms targeted toward optimal performance on a typical cluster of industry standard personal computers. In the hybrid case, overlap statistics, alignment, and consensus calculations need to take into account mixed populations of different read lengths and error characteristics.

A variety of use cases, from conventional sequencing, pure short read assembly, to hybrid assemblies ranging from bacteria to mammals will be presented together with run time analyses and quality assessments.

MEGAN analysis of metagenomic data

Daniel Huson

Center for Bioinformatics, Tuebingen University

A main bioinformatics challenge is to develop methods for identifying the taxonomical content of a sample of organisms from sequencing data, often obtained by ultrafast sequencing technologies. This talk will give an overview of metagenomics, in particular focusing on new computational challenges, and will introduce MEGAN, a new program for metagenome analysis.

References:

D.H. Huson, A. Auch, Ji Qi and S.C. Schuster, MEGAN Analysis of Metagenomic Data, *Genome Research*, 17:377-386, 2007.

H. N. Poinar, C. Schwarz, Ji Qi, B. Shapiro, R. D. E. MacPhee, B. Buigues, A. Tikhonov, D. H. Huson, L. P. Tomsho, A. Auch, M. Rampp, W. Miller, S. C. Schuster, Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA, *Science* 311:392-394, 2006.

Computational characterization of short environmental DNA fragments

Jens Stoye(1), Lutz Krause(1), Robert A. Edwards(2), Forest Rohwer(2), Naryttza N. Diaz(1), Alexander Goesmann(1), Scott Kelley(2), Alfred Pühler(1)

(1) Bielefeld University

(2) San Diego State University

Advances in sequencing methods have dramatically changed the way microbial ecosystems are studied, ushering into the emerging field of "metagenomics" - the sequence-based study of the ecology and evolution of free-living microbes. The recently developed 454 pyrosequencing has significantly dropped the time and cost constraints of environmental DNA sequencing while at the same time avoiding a potential bias that the cloning procedure may introduce. However, the main weakness of this technique is that at present, only short sequence reads of length 100-300 bp are produced.

Here, we show that despite the short length of reads, metagenomes obtained by 454 pyrosequencing can be used to readily characterize the genetic diversity and taxonomic composition of natural microbial communities. A strategy for finding short fragments of genes in the un-assembled reads of a sample using Pfam profile hidden Markov models is presented. These snippets of genes are subsequently assigned to a functional role and a taxonomic origin, yielding a community-specific genetic and taxonomic profile. In a comparative analysis of microbial samples from different environments, the applicability of genetic and taxonomic profiling will be illustrated. As a result, important trends in the metabolic activities and taxonomic composition of each community are unveiled.

New Sequencing Strategies for Microbes

Feng Chen, Alla Lapidus, Ed Kirton, Stephan Trong, Eugen Goltsman, Kurt Labutti, and **Paul Richardson**

US Department of Energy Joint, Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94549.

The advent of new sequencing technologies requires a new analysis of how to sequence microbial genomes in the most efficient manner. At the Joint Genome Institute, we have sequenced and finished over 150 microbial genomes and we are continually trying to improve the process. Our previous standard method utilized 3 whole genome shotgun libraries of various insert sizes sequenced to approximately 8.5X depth on capillary instruments. In the last year we have added 454 pyrosequencing coverage to approximately 15X while eliminating one shotgun library and reducing Sanger coverage to about 4-5X. Our goal is to ultimately eliminate Sanger coverage altogether, and we envision a step-wise process in which we will continue to reduce Sanger coverage, optimize pyrosequencing, and also layer in newer short read technologies to supplement coverage and eliminate errors.

Our hybrid Sanger/pyrosequencing approach involves first assembling the pyrosequencing reads into contigs using available software. We then shear those contigs (in-silico) into 1000bp overlapping fragments and combine them with Sanger reads using standard assemblers such as Phrap, PGA and Arachne. The resulting draft genome is then annotated and analyzed before proceeding to

the finishing process to close gaps and eliminate errors. We have recently begun to improve this process by analyzing error rates in pyrosequencing contigs to try to pinpoint areas for improvement by Sanger sequencing. We have also begun to analyze newer short reads may lead to genome sequence improvements.

Ultrafast *de novo* sequencing of the human pathogen *Corynebacterium urealyticum* with the Genome Sequencer 20 System

Andreas Tauch

Center for Biotechnology, Bielefeld University, Germany

Corynebacterium urealyticum is a lipid-requiring microorganism that is considered as part of the normal human skin flora. On the other hand, *C. urealyticum* is frequently isolated from urine samples of catheterized intensive care patients and causes urinary tract infections that are significantly associated with struvite (NH_4MgPO_4) stone formation. Many clinical isolates of *C. urealyticum* are multi-resistant against antibiotics and are often only susceptible to glycopeptides, such as vancomycin and teicoplanin. The Genome Sequencer System 20 was applied for *de novo* sequencing of the type strain *C. urealyticum* DSM7109 originally isolated from urine samples of an inpatient suffering from bladder stones [1]. Sequencing of the *C. urealyticum* DSM7109 genome with the Genome Sequencer System 20 yielded 657,410 sequence reads that were finally used for *de novo* genome assembly. By applying a contig length cut-off of 500 bp, a total number of 2,294,755 bases were assembled into 69 contigs, covering approx. 97% of the DSM7109 genome. Finishing of the *C. urealyticum* genome sequence was performed with template clones selected from a fosmid library. The genome of *C. urealyticum* DSM7109 consists of a circular chromosome of 2,366,079 bp and contains 2047 predicted protein-coding sequences. Synteny analysis at the protein level with other corynebacterial genome sequences provided evidence for the occurrence of genomic rearrangements in a distinct sub-line of the genus *Corynebacterium*. Metabolic analysis of the genome

sequence indicated that the lipid-requiring phenotype of *C. urealyticum* most likely originates from the absence of a fatty acid synthase gene and thus represents a fatty acid auxotrophy. Accordingly, both the gene repertoire for fatty acid metabolism and the deduced lifestyle of *C. urealyticum* largely reflect the strict dependence of growth on the presence of exogenous fatty acids. The annotated genome sequence of *C. urealyticum* DSM7109 also revealed valuable insights into the gene inventory related to antibiotic resistance, indicating that horizontal gene transfer played an important role in the acquisition of genes involved in multidrug resistance. Furthermore, a draft genome sequence with an approx. 30-fold coverage of the antibiotic susceptible clinical isolate *C. urealyticum* DSM7111 was established with the Genome Sequencer 20 System and used for comparative genome analysis.

Reference:

- [1] Tauch et al. 2006. Ultrafast *de novo* sequencing of *Corynebacterium urealyticum* using the Genome Sequencer 20 System. *Biochemica* 4: 4-6.

Access to the plasmid mobilome of wastewater treatment plant bacteria by applying the ultra-fast 454-sequencing technology

Andreas Schlüter⁽¹⁾, Rafael Szczepanowski⁽¹⁾, Thomas Bekel⁽²⁾,
Alexander Goesmann⁽²⁾, Lutz Krause⁽²⁾, Holger Krömeke⁽²⁾, and Alfred Pühler⁽¹⁾

⁽¹⁾ Department of Genetics, Bielefeld University, Universitätsstr. 25,
33615 Bielefeld

⁽²⁾ BRF, CeBiTec, Bielefeld University, Universitätsstr. 25, 33615 Bielefeld

Wastewater treatment plants (WWTP) are a reservoir for bacteria harbouring antibiotic resistance plasmids. Genomics of resistance plasmids isolated from WWTP bacteria revealed a high diversity of plasmid replicons and accessory genetic modules carrying resistance genes. Likewise, PCR analyses and microarray hybridisations confirmed that resistance genes covering all major classes are detectable in plasmid-DNA preparations from WWTP bacteria. To get a more complete picture of the plasmid mobilome of WWTP bacteria an ultra-fast sequencing approach applying the 454-technology was carried out. One 454-run yielded 346,427 reads with an average reading length of 104 bases resulting in a total of 36,071,493 bases sequence data. Categorisation of reads according to Gene Ontology (GO terms) and Pfam (protein family database) revealed that many reads represent genes involved in plasmid replication, mobilisation and plasmid stability. The obtained data confirm a high diversity of plasmids residing in WWTP bacteria. Accessory modules of plasmids encode different transposons, insertion sequences, resistance and virulence determinants and

other catabolic genes. Approximately 49,000 454-sequencing reads could be mapped to known plasmid genes deposited in databases. Assembly of mobilome sequencing reads resulted in 605 contigs with a minimum length of 500 bp. Contig annotation by means of SAMS (Sequence Analysis and Management System) and the Annotation Tool GenDB will be presented.

Next Generation Sequencing: Comparison of the technologies for bacterial genome sequencing

Kerstin A. Stangier, Christopher Bauser

GATC, Konstanz, Germany

The next generation sequencing technologies, embodied by the Roche Diagnostics/454 and Illumina/Solexa sequencing machines, provide the opportunity to generate huge amounts of sequence information for *de novo* and resequencing of bacterial genomes. The Paired-End method for both technologies, when available, will allow repetitive regions to be sequenced and assembled with accuracy approaching the Sanger method. GATC will present data of a bacterial genome that has been sequenced using all three available sequencing technologies: Sanger, 454 & Solexa. The availability of a reference genome allows a mapping of the 454 & Solexa reads. The longer read length of the 454 technology allows a *de novo* assembly of bacterial genomes while the high coverage of the Solexa technology is ideal for resequencing to find SNPs and other differences between bacterial strains. The Newbler assembler (454) provides a *de novo* assembly that can be imported into other assembly programs (e.g. SeqMan™ of the Lasergene™ suite from DNASTAR Inc., USA) and assembled with ABI 3730 data. The GS FLX from Roche Diagnostics will also provide the possibility of directly assembling data from different technologies. The mapping functions soon to be released for the Solexa technology will allow a visualization and analysis of differences (SNPs) between the genome of interest and the reference genome using a standard genome browser web interface.

With nearly two decades of experience, GATC Biotech is a leading provider of DNA sequencing services and bioinformatics software for industry and academia

worldwide. The company offers complete sequencing solutions, from sample prep to high throughput genome sequencing and bioinformatics. It is the only company in Europe that uses all three leading high throughput sequencing machines in house: the ABI 3730XL, the Illumina Genome Analyzer and the GS FLX from Roche Diagnostics, providing the flexibility to tailor programmes to meet customer requirements.

***Vitis vinifera* cv. Pinot Noir: dealing with heterozygous genomes by merging Sanger and pyrosequencing methods**

R. Velasco(1), A. Zharkikh(2), **M. Troggio**(1), S. Bhatnagar(2), M. Pindo(1), D. Cartwright(2), S. Vezzulli(1), G. Eldredge(2), G. Coppola(1), J. Mitchell(2), G. Malacarne(1), C. Segala(1), M. Stefanini(1), N. Gutin(2), M.S. Grando(1), D. Pruss(2), L. Dematté(1), A. Cestaro(1), S. Toppo(3), P. Fontana(1), M. Skolnick(2), A. Gutin(2), F. Salamini(4), R. Viola(1)

(1) IASMA Research Centre, Via E. Mach, 1 - I 38010 - San Michele all'Adige (TN); (2) Myriad Genetics Inc., 320 Wakara Way, Salt Lake City, Utah; (3) Dept. of Biological Chemistry, University of Padova, via U. Bassi, 58/b – I 35121 - Padova; (4) Parco Tecnologico Padano, Via Einstein – I 26900 loc. Cascina Codazza, Lodi

Given its cultural and economic importance, wine grape is an obvious candidate for the first woody crop to have its genome deciphered. Our research focused on the elite cultivar Pinot Noir with the multiple goals of genome assembly, gene identification and annotation, and identification of a maximum number of polymorphisms. Of special interest to biologists and breeders are polymorphisms in and around the coding regions. Pinot Noir is highly polymorphic with two clearly distinguishable haplotypes revealing several million SNPs and small indels. Based on a conservative estimate, 451,190 new SNPs and 67,580 new In/dels were revealed during the genome construction. This represents a substantial resource for molecular breeding programs, as well as trait and QTL marker association. Existing software and strategies were not adequate for the assembly of this highly heterozygous genome. We therefore largely focused on developing novel algorithms to address this challenge. A total coverage of seven genome equivalents of libraries of ascending size sequenced by the Sanger method, coupled with systematic highly parallel automated primer walking and

4.2 genome equivalents of 454 Life Science™ sequences, allowed us to create an effective genome sequence. Assembly was then reached by adding sequences of two BAC libraries and a fosmid library which were end-sequenced to assemble large meta-contigs. Contigs were oriented and ordered on appropriate chromosomes by high throughput marker development and genotyping in an F₁ cross of Syrah x Pinot Noir. Currently, 59,883 contigs merged into 705 meta-contigs covering 534,5 Mb have been submitted to the GenBank and EBI databases, and are available on the IASMA web site (<http://genomics.research.iasma.it>) organised in 19 chromosomes.

Profiling plant transcriptomes by massively-parallel pyrosequencing

Andreas P.M. Weber^{1,2}, Katrin L. Weber^{1,2}, Roshan P. Shrestha¹,
John B. Ohlrogge¹

¹Michigan State University, Department of Plant Biology, East Lansing, MI 48824, USA

²Current address: Heinrich-Heine-Universität, Institut für Biochemie der Pflanzen, Universitätsstraße 1, D-40225 Düsseldorf

The recent advances in sequencing technology, such as massively-parallel pyrosequencing (MPPS) of DNA molecules have made it possible to conduct economic high-throughput sequencing of transcript populations in short time and with very high coverage. Massively-parallel sequencing of DNA by pyrosequencing technology offers much higher throughput and lower cost than conventional Sanger sequencing. Although extensively used already for sequencing of genomes, relatively few applications of pyrosequencing to transcriptome analysis have been reported. In a recent proof-of-concept study, we found that two MPPS runs conducted on a non-normalized cDNA population from eight-day-old Arabidopsis seedlings provided 541,852 ESTs after quality control. Mapping of the ESTs to the Arabidopsis genome and to TAIR7 cDNA models indicated: 1) pyrosequencing detected transcription of 17,449 gene loci, providing very deep coverage of the transcriptome. Performing a second sequencing run only slightly increased the number of genes identified but significantly increased the overall sequence coverage by 50%. 2) Mapping of the ESTs to their corresponding full length transcripts indicated that all regions of the transcripts were represented regardless of transcript length or expression level. By comparison to microarrays, we found that short, medium and long transcripts

were equally represented. Both runs generated over 10 Mbp of non-redundant transcriptome sequence. This study demonstrated that pyrosequencing allows very rapid, low-cost survey of the transcriptome and that results are robust and unbiased. However, in our study, we found that 26% of all sequence reads were derived from only 25 highly expressed genes such as members of the Rubisco and LHC gene families, whereas over 5,000 genes were represented by less than 10 EST reads. Thus, if priority is on gene discovery and assembly of longer contigs rather than on identifying the most abundantly expressed genes it is required to normalize the cDNA population prior to sequencing to maximize coverage of all transcripts present in the sample

We have also conducted MPPS on non-normalized and normalized cDNA populations from various tissues of pea. Normalization of the cDNA population was performed using duplex-specific nuclease from Kamchatka crab. cDNA sequence information generated in this project was used in support of a proteomics study of pea and it was shown that identification of proteins by tandem mass-spectrometry using MPPS data was far superior to using existing EST data or sequence databases of related organisms, such as *Medicago*, soybean, or lotus.

Related references:

Andreas P.M. Weber, Katrin L. Weber, Kevin Carr, Curtis Wilkerson, John B. Ohlrogge (2007) Sampling the Arabidopsis Transcriptome with Massively-Parallel Pyrosequencing. *Plant Physiology* 144: 32-42.

Charting and Sequencing Structural Variation using High-Resolution Paired-End Mapping (HR-PEM)

Jan Korbel, Urban, Alexander Ekehart; Affourtit, Jason; Grubert, Fabian; Kim, Philip; Du, Lei; Carriero, Nicholas; Palejev, Dean; Godwin, Brian; Taillon, Bruce; Simons, Jan; Tanzer, Andrea, Carter, Nigel; Hurles, Matthew; Weissman, Sherman; Harkins, Tim; Gerstein, Mark; Egholm, Michael; Snyder, Michael

Yale University, New Haven, CT; 454 Life Sciences, Branford, CT; The Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; Roche Applied Science, Indianapolis, IN

Structural variants (SV), i.e. deletions, duplications, insertions, and inversions involving kilo- to Megabases of genomic DNA, were recently suggested to be responsible for a considerable amount of phenotype variation, and possibly, disease in humans (1-6). However, to date, most methods for identifying structural variants have resolutions in the order of 50–75 kb (7), and thus do not precisely identify the boundary sequences (i.e. *breakpoints*) of SVs (8). Furthermore, the majority of approaches used so far for cataloging SVs in the human genome do not detect copy-number neutral variation events such as inversions and balanced translocations.

We present a novel approach, High-Resolution Paired-End Mapping (HR-PEM), which makes use of 454/Roche sequencing technology, and combines computational analysis, high-throughput PCR assays, and amplicon-cocktail-sequencing to rapidly identify SVs at high resolution, and subsequently sequence across the breakpoints associated with these variants. The approach involves sequencing the ends of circularized 3 kb genomic fragments and mapping them onto the human genome reference sequence. The resolution of breakpoint assignments is ≈ 3 kb and thus well-suited for PCR validation. We have used HR-PEM to map and sequence SVs – i.e. simple deletions, insertions, and inversions, as well as more complex structural rearrangements – in two individuals in order to generate a precise map of SVs and their associated breakpoints. From 21 million and 10 million paired-end sequences,

respectively, from each individual, several hundred SVs have been predicted so far, ranging from 2 kb to several Mb in size. A first pass PCR analysis indicates that at least 60% of the predicted SVs can be amplified in a single PCR band and analyzed using DNA sequencing. Our results reveal as yet unexplored aspects of structural variation in the human genome, and suggest mechanisms by which this layer of genomic variation has arisen.

References

1. Sebat *et al.* (2004) *Science* **305**, 525-8.
2. Iafrate *et al.* *Nat Genet* **36**, 949-51.
3. Tuzun *et al.* *Nat Genet* **37**, 727-32.
4. Redon *et al.* (2006) *Nature* **444**, 444-54.
5. Gonzalez *et al.* (2005) *Science* **307**, 1434-40.
6. Stranger *et al.* (2007) *Science* **315**, 848-53.
7. Coe *et al.* (2007) *Genomics* **89**, 647-53.
8. Korbelt *et al.* (2007) *PNAS* **104**, 10110-5.

Mutation detection in cancer: parallel sequencing and mass-spectrometric genotyping for sensitive and high-throughput analysis of human cancer mutation space

Roman Thomas

Max-Planck Institute for Neurological Research, Cologne, Germany

A new class of cancer therapeutics targeting specific signaling pathways activated by gene mutations has shown clinical success. However, the molecular heterogeneity of non-small cell lung cancer (NSCLC) challenges the successful clinical deployment of these agents. This is mostly due to the fact that only subsets of the tumors of a given subtype carry a specific genetic lesion (e.g., EGFR mutation). Such lesions cause dependency of the tumor cells on the activity of this oncogene and thus render the cells sensitive to agents targeting this oncogene or the downstream pathway. Thus, taking such concepts into clinical testing requires real-time and accurate measurement of somatic (tumor-specific) genetic lesions in the tumor, irrespective of the quality of the tumor specimen. We have pioneered the application of novel parallel sequencing approaches for in-depth quantification of mutant EGFR alleles in clinical specimens. Such methods afford sensitive mutation diagnosis in real-time to guide patient stratification for rational deployment of molecularly targeted therapeutics. Furthermore, we have solved the issue of cost and throughput in large-scale cancer gene mutation profiling by applying mass-spectrometric genotyping to the detection of oncogene mutations in 1,000 human tumor samples. In summary, novel methods for genetic profiling afford sensitive, accurate and cost-effective patient stratification on the basis of the genetic make-up of the tumors rather than histology-based classification of cancer.

**Chronic inflammatory diseases of barrier organs:
from genetic etiology to genomic pathophysiology**

Philip Rosenstiel

Institut für Klinische Molekularbiologie, Kiel, Germany

Mutation discovery using SOLiD sequencing for the identification of gene knockouts in model organisms

Edwin Cuppen

Hubrecht Institute, Uppsalalaan 8, 3584 CT, Utrecht, The Netherlands,
e.cuppen@niob.knaw.nl

Over the last years, we have established a universal approach - target-selected mutagenesis - to knock out genes in model organisms. Random DNA mutations are introduced in the germline of an organism using chemical (usually ethyl-N-nitrosourea, ENU) mutagenesis. These animals are used to generate a large F1 population that is screened for induced mutations in genes of interest, which subsequently are outcrossed. Currently, this platform is operational in our institute for making knockouts in zebrafish, medaka, C.elegans, and most recently, we have successfully generated multiple rat knockout models.

Traditionally, we have been using dideoxy resequencing for mutation discovery, but this approach can only be scaled to a limited extent as samples can not be pooled. Here, we will report on our initial experiments to use massively parallel sequencing (SOLiD) for the identification of rare alleles in PCR-generated mini-genomes of pools of animals.

Pair-end-ditag (PET) based Sequencing Approach for Transcriptome Analysis and Genome Annotation

Ruan Yijun, Zhao Xiao Dong, Ng Patrick, Liu Jun, Yao Fei, Han Xu, Ooi Hong Sain, Chiu KuoPing, Sung Ken and **Wei, Chia-Lin**

Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672

In the past years, high throughput sequencing has evolved into a completely new stage. The arrival of the next generation sequencing technologies and their numerous applications has opened doors to many different research areas that were not previously appreciated. One of the unmet challenges in the post-genomic era is to completely and precisely define all functional elements, including all the expressed transcripts, promoters/enhancers, transcription factor binding sites, and epigenetic features that are carried in the human genome. Toward such endeavor, we have developed a pair-end-ditagging (PET) approach to interrogate transcriptome and regulatory elements with high efficiency and specificity. This novel approach “shrinks” each DNA fragment into a 36 bp pair-end ditag (PET; 18 bp 5’ and 18 bp 3’) followed by multiplex GS20 pyrosequencing sequencing. When mapped to genome, these PETs can clearly mark the boundary of each DNA fragment to specific regions on genome. When applied to full length cDNAs, the PETs precisely demarcate transcript boundaries, inferring proximal promoter sites, and identifying novel genes or alternative variants. We also demonstrated the PET approach when applied to chromatin immunoprecipitated DNA can determine the global localization of transcription factor binding sites (TFBS) and epigenetic histone modification profiles. I will describe these approaches and their ability to integrate with the next generation of DNA sequencing. The focus will be our effort in the annotation of the embryonic stem cell genome, generation of comprehensive expressed genes, identification of gene structure variations, mapping key master regulator, epigenetic modifications and the construction of cis-regulatory networks.

Neanderthal Genomics

Johannes Krause¹, Richard E. Green¹, Janet Kelso¹, Adrian Briggs¹, Udo Stenzel¹, Kay Prüfer¹, Susan E. Ptak¹, Johann Visagie¹, Michael Ronan², JF Simons², L Du², Michael Egholm², Jonathan Rothberg², & Svante Pääbo¹

¹Max-Planck Institute for Evolutionary Anthropology, Leipzig, Germany,

²454 Life Sciences, Branford, CT, USA

Neandertals are a hominid group that is morphologically distinct from modern humans. They appeared in the European fossil record around 400.000 years ago and disappeared around 30.000 years ago and are believed to be our closest extinct relatives. Although Neandertals and modern humans overlapped in certain regions in time and geographic range, the relationship between us and them is unclear and contentious. Genetic comparisons between modern humans and Neandertals could both address the relationship between us and them and offer the possibility to identify genetic changes that happened specifically on the human lineage. Furthermore, it would allow positive selection to be detected on the human lineage using Neandertal as a closely related outgroup.

We are applying large-scale parallel pyrosequencing on the GS-FLX and GS-20 platform to DNA extracts of Neandertal bones with the ultimate goal of achieving 1-fold coverage of the Neandertal genome. Several lines of evidence, including comparison to a modern human sample sequenced in a similar way, indicate that the fossil extract sequence is of Neandertal origin. Comparison of three European Neandertal individuals, including the Neandertal type specimen, to the human and chimpanzee genomes reveals that modern human and Neandertal DNA sequences diverged on average about 800,000 years ago. The effective size of the ancestral population of the two groups was found to be small, like the current effective population size of modern humans, suggesting a bottleneck before the Neandertal/modern human population split.

PARTICIPANTS

Alves Junior, Leonardo
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Agogué, Dr. Hélène
Netherlands Institut
Den Burg
PO Box 59
1790 Texel
The Netherlands

Arnold, Dr. Walter
Zentrales Isotopenlabor
Universität Bielefeld
33594 Bielefeld
Germany

Baier, Markus C.
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Bau, Stephan
febit synbio GmbH
Im Neuenheimer Feld 519
69120 Heidelberg
Germany

Bauer, Dr. Eva
Universität Hohenheim
Fruwirthstr. 21
70599 Stuttgart
Germany

Baumgarth, Dr. Birgit
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Becker, HD Dr. Anke
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Bekel, Thomas
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Berninger, Philipp
Biozentrum
University of Basel
Department of Bioinformatics
Klingelbergstr. 50/70
4056 Basel
Switzerland

Bertelli, Dr. Claire
Institute of Microbiology
University of Lausanne
Bugnon 48
1011 Lausanne
Switzerland

Blom, Jochen
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Bölter, Dr. Jürgen
Roche Diagnostics GmbH
Vertrieb Applied Science-VA
Außendienst Nord
Sandhofstr. 116
68308 Mannheim
Germany

Böhm, Dr. Markus
Illumina Ltd
Chesterford Research Park
Little Chesterford
Nr Saffron Walden
Essex, CB10 1XL
United Kingdom

Braun, Dorothee
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Brink, Maaike
Netherlands Institut
Den Burg
PO Box 59
1790 Texel
The Netherlands

Brinkrolf, Karina
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Cherkasov, Dr. med. Dmitry
Institut für med. Mikrobiology
Uni Marburg
Hans-Meerwein Straße 2
D-35032 Marburg
Germany

Cuppen, Prof. Dr. E.
Functional Genomics Group
Hubrecht Lab
Uppsalalaan 8
3584 CT Utrecht
The Netherlands

Cuppens, Prof. Dr. Harry
Center Human Genetics /
Centrum Menselijke Erfelijkheid
Campus Gasthuisberg
Herestraat 49 - box 602
BE-3000 Leuven
Belgium

Diaz, Naryttza
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Distl, Prof. Dr. Ottmar
University of Veterinary
Medicine Hannover
Institute for Animal
Breeding & Genetics
Bünteweg 17p
30559 Hannover
Germany

Dörhöfer, Dr. Sabine
Roche Diagnostics
Scandinavia AB, Box 147
SE-16126 Bromma/Stockholm
Sweden

Dröge, Dr. Marcus
Roche Diagnostics GmbH
Customer Support- Customer Care
Genome Sequencing
Nonnenwald 2
82377 Penzberg
Germany

Dutilh, Dr. Bas E.
CMBI / NCMLS / Radboud University Nijmegen
Medical Center
Geert Grooteplein 28
6525 GA
Nijmegen
The Netherlands

Eggeling, Dr. Lothar
Institut für Biotechnologie 1
Forschungszentrum Jülich
52425 Jülich
Germany

Ehrenreich, Dr. Armin
General Mikrobiology
Institute for Mikrobiology and Genetics
Georg-August-University of Göttingen
Grisebachstr. 8
D-37077 Göttingen
Germany

Evers, Dr. Dirk
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Finkelburg, Dr. Britta
Applera Deutschland GmbH
Frankfurter Str. 129 b
64293 Darmstadt
Germany

Firnhaber, Christian
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Gaidatzis, Dimosthenis
Biozentrum
University of Basel
Department of Bioinformatics
Klingelbergstr. 50/70
4056 Basel
Switzerland

Gerlach, Dr. Gabriele
Universität Würzburg
Institut für Hygiene und Mikrobiologie
Josef-Schneider-Str. 2
Bau E 1
97080 Würzburg
Germany

Ghai, Dr. Rohit
Justus Liebig-Universität
Institut für med. Mikrobiologie
Frankfurter Str. 107
35392 Gießen
Germany

Giegerich, Prof. Dr. Robert
Technische Fakultät
Universität Bielefeld
33594 Bielefeld
Germany

Goesmann, Dr. Alexander
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Gottwald, Dr. Sven
Leibniz-Institut für Pflanzengenetik
und Kulturpflanzenforschung
Corrensstr. 3
06466 Gatersleben
Germany

Gradl, Dr. Georg
Sales Manager Sequencing Europe
Eurofins MWG GmbH
Anzinger Str. 7a
85560 Ebersberg

Graedler, Dr. Florian
Field Application Scientist
Illumina, Inc.
9885 Towne Centre Drive
San Diego
CA 92121
USA

Günther, Dr. Simone
Applied Biosystems
Frankfurter Str. 129 b
64293 Darmstadt
Germany

Hain, Dr. Torsten
Justus Liebig-Universität
Institut für med. Mikrobiologie
Frankfurter Str. 107
35392 Gießen
Germany

Hall, Dr. Kevin
Illumina, Inc
Chesterford Research Park
Little Chesterford
Nr. Saffron Walden
Essex
CB10 1XL
United Kingdom

Hardin, Dr. Susan H.
President and CEO
VisiGen Biotechnologies, Inc.
2575 West Bellfort, Suite 250
Houston, TX 77054
USA

Hausmann, Dr. Ludger
BAZ Institut für Rebenzüchtung
Geilweilerhof
76833 Siebeldingen
Germany

Henckel, Kolja
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Himmelbauer, Dr. Heinz
Max-Planck-Institute of Molecular Genetics
Innestr.73
D-14195 Berlin
Germany

Hüser, Dr Andrea T.
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Huson, Prof. Dr Daniel
Center for Bioinformatics (ZBIT)
Institute for Computer Science,
Tuebingen University
72076 Tuebingen
Germany

Janitz, Dr. Karolina
AnySeq Sp.2 O.O.
Limanouskiegs 24 A/16
60-744 Poznan
Poland

Jochmann, Nina
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Jungwirth, Britta
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Kaiser, Dr. Olaf
Roche Diagnostics GmbH
Customer Support- Customer Care
Genome Sequencing
Nonnenwald 2
82377 Penzberg
Germany

Kalinowski, Dr. Jörn
Universität Bielefeld
Institut für Genomforschung
& Systembiologie
33594 Bielefeld
Germany

Kianianmomeni, Arash
Abteilung für Zell- und
Entwicklungsbiologie der Pflanzen
Universität Bielefeld
33594 Bielefeld
Germany

Klauke, Dr. Bärbel
Herz- und Diabeteszentrum NRW
Ruhr-Universität Bochum
Georgstraße 11

Klein, Dr. Rolf
Institut für Immunologie und Genetik
am Klinikum Kaiserslautern
Hellmut-Hartert-Str. 1
67655 Kaiserslautern
Germany

Korbel, Dr. Jan O.
Yale University Center for Excellence in Genome
Science
266 Whitney Ave.
PO Box 208114
Yale University
New Haven, CT 06520 USA

Krause, Dr. Johannes
Department of Genetics
Max Planck Institute for Evolutionary
Anthropology
Deutscher Platz 6
04103 Leipzig
Germany

Krause, Lutz
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Krömeke, Holger
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Küster, PD Dr. Helge
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Laplace, Prof. Dr. Frank
Bundesministerium für Bildung
und Forschung (BMBF)
10115 Berlin
Germany

Liebl, Prof. Dr. Wolfgang
Georg-August-Universität Göttingen
Institut für Mikrobiologie und Genetik
Grisebachst. 8
37077 Göttingen
Germany

Linke, Burkhard
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Luerßen, Dr. Hartmut
Trait Genetics GmbH
Am Schwabeplan 1b
06466 Gatersleben
Germany

Lunstroo, Hans
Roche Diagnostics GmbH
Nederland BV
BA MD/AS
Postbus 1007
1300 BA Almere
The Netherlands

Matthiesen, Dr. Peter
Roche Diagnostics GmbH
Customer Support- Customer Care
Genome Sequencing
Nonnenwald 2
82377 Penzberg
Germany

Mellmann, Dr. Alexander
Inst. Hygiene
Universität Hospital Münster
Robert Koch Str. 41
48149 Münster
Germany

Menzel, Dr. Uwe
Fritz-Lipmann-Institut, Jena
Beutenbergstr. 11
07745 Jena
Germany

Mitra, Suparna
Algorithms in Bioinformatics
ZBIT Center for Bioinformatics
Universität Tübingen
Sand 14
72076 Tübingen
Germany

Nematollahi, Ghazaleh
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Neuweger, Heiko
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Nordsiek, Gabriele
Helmholtz-Zentrum für Infektionsforschung
Inhoffenstraße 7
38124 Braunschweig
Germany

Nowak, Dr. Jacek
National Research Council
Plant Biotechnology Institute
Saskatoon, SK
S7N 0W9
Canada

Nübel, Dr. Ulrich
Robert-Koch-Institut
Burgstr. 37
38855 Wernigerode
Germany

Nürnberg, Prof. Dr. Peter
Institut für Genetik
Universität zu Köln
Zülpicher Str. 47
50674 Köln
Germany

Obermeier, Dr. Christian
Justus-Liebig Universität Gießen
Heinrich-Buff-Ring 26-32
35392 Gießen
Germany

Oehm, Sebastian
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Paulin, Dr. Lars
Institute of Biotechnology
Street address: Viikinkaari 9
00014 University of Helsinki
Finland

Peterson, Dr. Hans-Peter
Forschungszentrum Jülich GmbH
Projekträger PTJ
52425 Jülich
Germany

Platt, Dr. Darren
Head of Informatics
DOE Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598,
USA

Pühler, Prof. Dr. Alfred
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Rakin, Dr. Alexander
Max von Pettenkofer-Institut, LMU
Pettenkoferstr. 9 a
80336 München
Germany

Rahmann, Dr. Sven
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Reinecke, Dr. Frank
QIAGEN GmbH
Research & Development
Qiagen Str. 1
40724 Hilden
Germany

Reinhardt, Dr. Richard
MPI Molecular Genetics
Innestrasse 63-73
D-14195 Berlin-Dahlem
Germany

Richardson, Dr. Paul
Head of Genomics
Technologies Program
DOE Joint Genome Institute
2800 Mitchell Drive
Walnut Creek, CA 94598
USA

Richter, Daniel
Algorithms in Bioinformatics
ZBIT Center for Bioinformatics
Universität Tübingen
Sand 14
72076 Tübingen
Germany

Rhodes, Dr. Michael
Applications Manager
High Throughput Discovery
Applied Biosystems
850 Lincoln Centre Drive
Foster City CA 94401
USA

Rodriguez, Dr. Verenice Ramirez
Laboratorio Nacional
de Genómica para la
Biodiversidad
CINVESTAV Campus
Guanajuato
Km 9.6 Carr. Libramiento Norte
Irapuato-Leon
Irapuato , Guanajuato
México

Rosenstiel, Dr. Philip
Institut für Klinische Molekularbiologie
Campus Kiel
Schittenhelmstraße 12
24105 Kiel
Germany

Rückert, Dr. Christian
Universität Bielefeld
Institut für Genomforschung
& Systembiologie
33594 Bielefeld
Germany

Sakwa, Dr. James
National Genomics Platform
LIFElab
ECOBio Innovation Center
Durban
South Africa

Scharfe, M.
Helmholtz-Zentrum für Infektionsforschung
Inhoffenstraße 7
38124 Braunschweig
Germany

Schlüter, Dr. Andreas
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Schneiker, Dr. Susanne
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Schoen, Dr. Christoph
Universität Würzburg
Institut für Hygiene und Mikrobiologie
Josef-Schneider-Str. 2
97080 Würzburg
Germany

Schürmann, Klaus-Bernd
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Schulte, Dr. Daniela
Leibniz-Institut für Pflanzengenetik
und Kulturpflanzenforschung
Corrensstr. 3
06466 Gatersleben
Germany

Schuster, Prof. Dr. Stephan C.
310 Wartik Building
Penn State University
University Park
PA 16802,
USA

Sczyrba, Dr. Alexander
Technische Fakultät
Universität Bielefeld
33594 Bielefeld
Germany

Selbitschka, Dr. Werner
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Selbmann, Dr. Peter
AGOWA GmbH
Glienicke Weg 185
12489 Berlin
Germany

Serrania, Javier
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Siebelt, Nicole
febit synbio GmbH
Im Neuenheimer Feld 519
69120 Heidelberg
Germany

Srinivasan, Dr. Maithreyan
Director, Sequencing and
Protein Sciences
454 Life Sciences/
Roche Diagnostics
20 Commercial Street
Branford, CT 06405
U.S.A.

Stangier, Dr. Kerstin
GATC Biotech AG
Global Key Account Manager
Jakob-Stadler-Platz 7
D-78467 Konstanz
Germany

Stein, Dr. Nils
IPU Gatersleben
Corrensstr. 3
06466 Gatersleben
Germany

Stoye, Prof. Dr. Jens
Technische Fakultät
Universität Bielefeld
33594 Bielefeld
Germany

Strack, Dr. Volker
Key Account Manager Sequenzierung
Roche Diagnostics GmbH
Sandhoferstr. 116
68305 Mannheim
Germany

Strittmatter, Dr. Axel
Georg-August Universität
Göttingen
Grisebachstraße 8
37077 Göttingen
Germany

Sun, Dr. Jibin
Helmholtz-Zentrum für
Infektionsforschung
Inhoffenstr. 7
38124 Braunschweig
Germany

Tanzi, Dr. Raimo
Appliedbiosystems
Via Per Rovido 2
20090 Buccinasco
Italy

Tauch, Dr. Andreas
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Taudien, Dr. Stefan
Leibniz Institute for Age Research
Fritz Lipman Institute (FLI)
Genome Analysis
Beutenbergstr. 11
07745 Jena
Germany

Thomas, Dr. Roman
Max-Planck-Institut für
neurologische Forschung
GIMax-Planck-Institut
für neurologische Forschung
Gleueler Str. 50
50931 Köln
Germany

Troggi, Dr. Michela
Biology Dept. Istituto Agrario
San Michele all'Adige
Italy

Trzeciok, Dr. Dietrich
Institut für Mikrobiologie und Genetik
Georg-August-Universität Göttingen
Grisebachstr. 8
37077 Göttingen
Germany

Tümmler, Prof. Dr. Burkhard
Medizinische Hochschule Hannover
Klinische Forschergruppe
Carl-Neuberg-Str. 1
30625 Hannover
Germany

van den Bergh, Marianne
Agro Business Park 90
6708 PW Wageningen
The Netherlands

van Nimwegen, Dr. Erik
Biozentrum
University of Basel
Department of Bioinformatics
Klingelbergstr. 50/70
4056 Basel
Switzerland

Vorhölter, Dr. Frank-Jörg
Lehrstuhl für Genetik
Universität Bielefeld
33594 Bielefeld
Germany

Weber, Prof. Dr. Andreas
Institute for Plant Biochemistry
Heinrich-Heine-University
Geb. 26.03.01
Universitaetsstrasse 1
D-40225 Duesseldorf
Germany

Weber, Dr. Tilmann
Universität Tübingen
Lehrstuhl für Mikrobiologie/Biotechnologie
Auf der Morgenstelle 28
72076 Tübingen
Germany

Wei, Dr. Chia-Lin
Genome Institute of Singapore
60 Biopolis Street, 02-01,
Genome,
Singapore 138672

Weidner, Dr. Stefan
CeBiTec
Universität Bielefeld
33594 Bielefeld
Germany

Weisshaar, Prof. Dr. Bernd
Universität Bielefeld
Fakultät für Biologie
Lehrstuhl für Genomforschung
3594 Bielefeld
Germany

Wiezer, Dr. Arnim
Georg-August Universität
Göttingen
Grisebachstraße 8
37077 Göttingen
Germany

Wittendrop, Jon
Keygene NV
Agro Business Park 90
6708 PW Wageningen
The Netherlands

Zavolan, Dr. Mihaela
Biozentrum
University of Basel
Department of Bioinformatics
Klingelbergstr. 50/70,
4056 Basel
Switzerland

Zimmermann, Dr. Wolfgang
AGOWA GmbH
Glienicke Weg 185
12489 Berlin
Germany