De Novo Sequencing of Genomes
Re-Sequencing of Genomes and Comparative Genomics
Transcriptome Analysis
Ultra Deep Amplicon Sequencing
Targeted Re-Sequencing by Sequence Capture
Re-Sequencing of Genomic Fragments
Human Data Analysis and Data Mapping
Customised Bioinformatic Services
NGS Library Generation Services

www.eurofinsdna.com
**De Novo Sequencing of Genomes (Roche/454 Sequencing)**

The sequences of unknown DNA viruses and prokaryotic or eukaryotic genomes can be determined using Roche GS FLX with Titanium-series chemistry and latest FLX+ technology. It is possible to analyse, study and characterise genomes or metagenomes of interest in short turn around times (TAT). When sequencing is performed on Roche GS Junior™ an express delivery time of 1-3 weeks is possible.

**Whole genome sequencing of viral, prokaryotic and eukaryotic genomes**

- Fragmentation by nebulisation or shearing of the genomic DNA
- Preparation of non-cloned shotgun library and clonal amplification by emulsion PCR (emPCR)
- Optional, generation and sequencing of long paired end libraries (Fig. 1)

![Fig. 1: Generation of a long paired end library](image)

- Ultra high throughput sequencing with GS FLX Titanium-series chemistry to approx. 20 fold coverage
- Standard or enhanced data assembly with appropriate assembly software (gsMapper, MIRA or Celera)
- Mapping of long paired end sequences and scaffolding of the contigs (see Fig. 2, next page)
- De novo sequencing of complex genomes

---

**The Best Solution for Your Application**

Powered by the Genome Sequencer™ FLX/FLX+, GS Junior™, Illumina HiSeq 2000™ and PacBio RS™, a multitude of well known or new next generation sequencing (NGS) applications are provided. Optimum results are achieved with our new, shortened and more flexible service times including optional express delivery (e.g. for 1/16 lane). We can combine all NGS technology services with classical Sanger sequencing.
- Closing of gaps by semi-automated primer design on adjacent contig ends and sequencing of
  - PCR amplicons of the genomic region of interest, or
  - Sequencing of long range PCR by primer walking
- Basic and complex bioinformatic analysis such as reference guided ORF finding, annotation or sample comparison (Synteny plot)

Re-Sequencing of Genomes and Comparative Genomics

Genome comparison or re-sequencing of genomes using the NGS technologies is the method of choice to identify genetic variations including single base mutations (SNPs), inserted and deleted genes or heterozygous SNPs. Re-sequencing is used for production strain optimisation, metabolic engineering or mutation analysis. For well known reference sequence projects we recommend sequencing on HiSeq 2000 using at least 2 different shotgun libraries with different fragment lengths. Optional our long jumping distance libraries (LJD) with variable jumping sizes (3 kbp, 8 kbp, 20 kbp, 40 kbp) can be combined.

- Fragmentation of the genomic DNA by nebulisation
- Optional, preparation of 2 shotgun libraries (i.e. with 200 bp and 400 bp fragment size) per sample
- Shotgun sequencing on HiSeq 2000 to 30–50 fold sequence coverage
- Optional, in combination with sequencing of different LJD libraries
- Mapping of the sequencing data onto a reference sequence using appropriate software (= reference guided assembly)
- Optional, professional data mapping and analysis using Genomatix services can be ordered
- Sequence reads which are not present on the reference genome (e.g. due to phage insertions or plasmids) are automatically sorted and will be delivered
- Optional, closing of gaps by designing primer pairs for adjacent contig ends and sequencing PCR amplicons of the genomic region of interest
- Genome comparison by SNP analysis (Fig. 3)
- Additional bioinformatic services available e.g. annotation of contigs and sequencing of PCR amplicons of the genomic region of interest
Transcriptome Analysis

The transcriptome analysis service covers the construction and sequencing of random primed, 3’-fragment and/or 5’-fragment cDNA libraries, optional with normalisation. We also offer preparation and sequencing of miRNA or snRNA (small non coding RNA) libraries. New and unknown expressed sequence tags (ESTs) or new rare transcripts can be identified. Expression profiling and differential gene expression studies can be performed. Depending on the application, different sequencing strategies are carried out either with long read technology (GS FLX/FLX+, GS Junior) or massively short read technology (HiSeq 2000).

- Generation of random primed cDNA libraries (no nebulisation step is needed)
- Generation of 3’- or 5’-fragment and miRNA cDNA libraries (no nebulisation step is needed)
- Generation of normalised cDNA libraries (Fig. 4, 5)
- Generation of the Illumina mRNA library (fragmentation of RNA with zinc protocol)
- Ultra high throughput sequencing of above cDNA libraries with GS FLX/FLX+ or HiSeq 2000

![Fig. 4: Generation of a normalised cDNA library](image)

![Fig. 5: Kinetics of Normalisation](image)

- State-of-the-art clustering and assembly service (Fig. 6)
  - Overall clustering and assembly statistics including the number of contigs and singlets, number of reads per cluster, information on contig lengths and allocation of reads to contigs/clusters, etc.
  - FASTA files of all singlets and contigs
  - BLASTn and BLASTx analysis and filtered output files in standard GFF format
Amplicon Sequencing

Amplicon sequencing by the GS FLX/FLX+ technology refers to ultra deep sequencing of PCR products, exons or multiplexed samples for analysing genetic variations e.g. for SNP or mutation detection, variance analysis or for identifying and qualifying methylation patterns.

- We recommend providing amplicons which include PCR incorporated specific adaptors (A and B primer sequences)
- Alternatively PCR products may be sent without adaptor (less efficient setup)
- Clonal amplification by emulsion PCR (emPCR)
- Sequencing of samples using A/B-primer and optional pooling of samples with specific sequence tags (MID=Multiplex Identifiers, Fig. 7)
- Ultra deep sequencing with the GS FLX/FLX+ system
- Clustering of the sequence reads, sophisticated variance analysis
- Further bioinformatic analysis, if required

Fig. 6: De novo cDNA assembly of GS FLX reads

Fig. 7: Scheme of the generation of ready-to-sequence bar coded amplicons
Targeted Re-Sequencing by Roche NimbleGen Sequence Capture (Certified Service Provider of Roche NimbleGen)

Sequence Capture enables the selective enrichment of genomic regions from full-complexity genomic DNA. The method is applicable to any target region of interest, like large contiguous or non-contiguous genomic regions (whole exomes; Fig. 8). Eurofins MWG Operon combines sequence capture with powerful next generation re-sequencing solutions.

As a Certified Service Provider (CSP) of Roche NimbleGen we offer customised Sequence Capture array service with 3 different Roche NimbleGen array layouts and EZ in-solution capturing. The 385 K, the 2.1 M and the 12x 135 K array capture up to 5 Mbp and 30 Mbp target sequence or 12x 1.5 Mbp identical target sequence, respectively. The arrays contain overlapping high density oligonucleotide probes (75 - 100 bp) that generate redundant coverage of the target region of interest.

Latest EZ in-solution Choice and Choice XL customised version can be ordered with human target sizes of 100 kbp - 7 Mbp and 7 Mbp - 50 Mbp, respectively. For capture of non-human genomic regions Eurofins MWG Operon offers EZ Developer with target sizes up to 50 Mbp.

For targeted re-sequencing of human exomes we offer the Eurofins MWG Operon and Genomatix whole human exome array technology, which provides the largest and most complete coverage of the human exome available on the market. The EZ in-solution capture for the Roche NimbleGen whole human exome is an excellent way to economically analyse large amounts of human samples in parallel. Alternative enrichment methods are available on request.

- Customer either defines the target region and respective capture area or selects an established exome design (on array or in-solution)
- After design approval, fragmentation of the genomic DNA sample and hybridisation to the capture probes
- Removal of unbound fragments and elution and amplification of target enriched genomic fragments
- Available NimbleGen Capture sizes are: up to 5 Mbp, up to 30 Mbp, 12x 1.5 Mbp, 100 kbp - 7 Mbp, 7 - 50 Mbp
- Ultra high throughput sequencing of enriched samples either with long read technology (GS FLX/FLX+, GS Junior) or massive short read technology (HiSeq 2000)
- Bioinformatic analysis, e.g. mapping to a reference sequence, SNP analysis (in cooperation with Genomatix Software GmbH)

![Fig. 8: NimbleGen Sequence Capture in-solution (left) or on array (right)](image-url)
High throughput re-sequencing of genomic fragments is the sequencing of specific genomic regions such as exons or genes of interest in order to find a correlation between sequence deviations (e.g. SNPs or InDel mutations) and phenotypes. Based on Sanger technology, this service is performed in three steps:

**Phase I: Establishment of the PCR**
- Primer design and synthesis for all exon regions
- Establishment of PCR amplification
- Quality check by double stranded sequencing of the test samples

**Phase II: PCR and Sequencing**
- High throughput PCR of all DNA samples
- High throughput purification and Sanger sequencing of PCR products

**Phase III: Bioinformatic Analysis**
- Identification of all differences between reference and samples (Fig. 9)
- Homozygous and heterozygous substitutions
- Detection of insertions/deletions

---

**Customised Bioinformatic Services**

Analysis of data from a sequencing project sometimes require individual bioinformatic solutions or industrial scale bioinformatics. We offer not only clustering and assembly services for Roche and Illumina sequencing data, standard BLAST analysis and annotation services. In addition, we can support specific requests with customised bioinformatic solutions.

Eurofins MWG Operon offers
- Sequence clustering and assembly services for GS FLX/FLX+ and HiSeq 2000 data
- De-Novo assembly of genome data with different assembly software tools
- Genome annotation services and prediction of protein function
- Standard BLAST analysis for single reads and assembled contigs
- BLAST analysis of genomic and cDNA sequencing data
- Mapping and cross-mapping of sequencing data on reference sequences
- Read count reporting (expression profiling) for transcriptome analysis projects
- SNP/InDel analysis (variant calling) for re-sequencing projects

---

Fig. 9: Mutation surveyor (screenshot from sequencing data analysis)
Eurofins offers their bioinformatics services in an exclusive worldwide cooperation with Genomatix Software GmbH. Genomatix is a leading company in complex analysis and interpretation of next generation sequencing data and state-of-the-art solutions for your individual project. Founded in 1997, Genomatix has published more than 180 peer reviewed scientific papers with more than 5,000 citations. Over 35,000 researchers worldwide currently apply Genomatix tools and databases.

Genomatix Mining Station (GMS)
The Genomatix Mining Station (GMS) delivers an integrated environment for profound first level analysis of next generation sequencing data. Proprietary algorithms and workflows perform all mapping related tasks. Assembly, genome indexing, genomic positioning, SNP detection, CNV and structural variations, alternative splicing and transcript variants are addressed by GMS with highest scientific rigor (Fig. 10).

Genomatix Genome Analyzer (GGA)
The Genomatix Genome Analyzer is an integrated solution for comprehensive second level analysis addressing deep biological data analysis and interpretation. More than 3 terabyte of structured, pre-analysed background data shed light on biological context, intra- and cross species. ChIP-Seq, RNA-Seq or genotyping experiments, gene regulation analysis, epigenetics, comparative genomics, pathway and literature mining – the GGA produces results of highest relevance (Fig. 10).

Genomatix Software Suite
A well-established software bundle, the Genomatix Software Suite performs a number of tasks
- Conducts a scientific analysis of genomic data, gene regulation and expression
- Generates and evaluates networks and pathways
- Performs extended literature searches and sequence analyses and extraction
- Visualises comprehensive genome annotation.

Fig. 10: Screenshot genome mining station and genome analyzer software

NGS Library Generation Services

Eurofins MWG Operon routinely prepares a multitude of library types for sequencing on GS FLX/FLX+ or HiSeq 2000. We are able to provide you with high quality genomic DNA and cDNA libraries ready to sequence on the appropriate sequencing instrument.

- Genomic shotgun libraries to be sequenced on GS FLX/FLX+
- Long paired end (LPE) libraries adapted to be sequenced on GS FLX/FLX+ and mate pair libraries to be sequenced on HiSeq 2000
- Long jumping distance libraries (LJD) for sequencing on HiSeq 2000
- Random primed, 3'-fragment and 5'-fragment cDNA libraries optional with normalisation are available for transcriptome analysis on both, GS FLX/FLX+ or HiSeq 2000
- Generation of cloned shotgun libraries i.e. mechanically sheared shotgun fragments of large constructs or bacterial genomes cloned into plasmids
Advantages of the Next Generation Sequencing Technology

Utilising the respective advantages of each technology – pyrosequencing with GS FLX/FLX+ (long read technology), sequencing by synthesis with HiSeq 2000 (short read technology), single molecule realtime sequencing with PacBio RS and Sanger sequencing with ABI 3730 XL – adds strength and a higher degree of flexibility than any technology would contribute on its own.

PacBio RS™ Sequencer

- Latest technique
- 30 - 40 Mbp of data per SMRT cell
- Average read length of 1,500 - 2,500 bp
- Max. read length of up to 4,500 bp
- Strobe sequencing technology will be available soon

Roche Genome Sequencer™ FLX/FLX+

- Good representation of AT-rich areas as no cloning step is needed
- Trouble-free sequencing of regions with high GC content – no "full stops"
- Throughput with Titanium-series chemistry
  - Generation of 400 - 500 Mbp per run or 1 Gbp per day
  - About 1 million reads per plate with an average length of 350 - 450 bp
- Reliable sequence quality
  - Consensus accuracy > 99.99 % at a 20-fold sequence coverage
  - Single read accuracy > 99.5 %
- Multiplexing of several samples per lane possible (on routine 12 - 24, please ask for more)
- New FLX+ chemistry version offers:
  - 600 - 800 bp read length
  - About 700 Mbp per run

Illumina HiSeq 2000

- Good representation of AT-rich areas as no cloning step is needed
- Paired end sequencing possible: 2x 50 bp and 2x 100bp
- Throughput per channel
  - 80 to 150 million clusters with read lengths of 50 or 100 bp
  - Generation of up to 15 Gbp per channel (1x 100 bp)
  - Generation of up to 30 Gbp per channel (2x 100 bp)
- Reliable sequence quality
  - 85-95 % of bases of single reads display 99.9 % accuracy
- Multiplexing of several samples per channel possible (4 - 24 on routine, please ask for more)
- Higher cluster density with chemistry version 3.0

Sanger Technology

- System of choice for finishing projects
  - Gap closure with primer walking strategy
  - Sequencing of amplified PCR products of interest
- Read lengths of up to 1,100 bases
- Gold standard for single read approaches
  - Re-sequencing of selected PCR products or plasmid clones
  - Sequencing of specific regions of large constructs like BAC, PAC, Cosmid or Fosmid clones
1. Pyrosequencing with emPCR by Roche GS FLX/Roche Junior

Single molecules are clonally amplified by emulsion PCR (emPCR) in an oil-water emulsion (Fig. 11). Millions of clonally amplified molecules bound to beads are then sequenced by pyrosequencing. The strategy relies on detection of pyrophosphate release, generated by a cascade of enzymatic reactions on nucleotide incorporation (Fig. 12).

2. Sequencing by synthesis with Illumina HiSeq 2000

DNA fragments are amplified on a surface via bridge PCR (Fig. 13). The generated clusters are sequenced by synthesis using a technique called cyclic reversible termination (Fig. 14). As the name implies the method uses reversible terminators in a cyclic method that comprises incorporation of a single nucleotide, fluorescence imaging and cleavage of the blocking group.
Eurofins MWG Operon has been providing sequencing services for projects of all sizes since 1996.

**Expertise in de novo sequencing and re-sequencing of genomes**

We have sequenced over 100 microbial genomes De-Novo, such as E.coli strains, strains of Streptomyces, Chlamydomonas, Lactococcus, Salmonella, Clostridium, Mycoplasma, Pseudomonas, etc. In the first year after the implementation of the GS FLX system we finalised several dozen larger projects such as fungal genomes and metagenome projects. Since 2010 we have also gained excellent experience in de novo sequencing of large eukaryotic genomes. BAC sequencing projects with thousands of BACs have been performed. We have developed an advanced multiplex approach to sequence up to 48 BACs in parallel.

**Expertise in transcriptome analysis and amplicon sequencing**

We have carried out more than 100 large EST projects using Sanger technology. For ultra deep Transcriptome Analysis we use a new type of non cloned normalised and 3’-fragment cDNA libraries and miRNA libraries. This enabled us to sequence over 80 transcriptomes as well as hundreds of amplicon projects in the last year.

**Selected publications**


Young et al. *Elucidating the transcriptome of Fasciola hepatica - A key to fundamental and biotechnological discoveries for a neglected parasite.* Biotechnology Advances 2010, 28; 222–231


Horn et al. *Illuminating the evolutionary history of Chlamydia (Sequencing of the genome of Protoclamydia amoebiophila).* Science. 2004 Apr 30; 304 (5671): 728–30

GS FLX/FLX+, GS Junior and Genome Sequencer are trademarks of Roche, Illumina and HiSeq are trademarks of Illumina, Inc and ABI 3730 XL is a trademark of Applied Biosystems Group, PacBio RS is a trademark of Pacific Bioscience, Genomatix is a trademark of Genomatix Software GmbH.
Feel Free to Contact Us

We invite you to send us your request using our online request form at www.eurofinsdna.com, by email or contact us directly by phone.

Free Phone Numbers:
- Austria: 0800 296 562
- Belgium: 0800 77862
- Denmark: 8088 1262
- Finland: 0800 112 744
- France: 0800 903 807
- Ireland: 1800 555 056
- Italy: 800 785 950
- Luxembourg: 8002 6418
- Netherlands: 0800 0226215
- Norway: 800 138 44
- Sweden: 020 798 148
- Switzerland: 0800 562 013
- UK: 0800 0323 135

Email:
genseq-eu@eurofins.com

Phone:
+49 8092 8289-77

Our official business hours are:
8 a.m. – 6 p.m. CET