Executive Summary:
The METAEXPLORE project developed and applied advanced molecular tools that allowed the screening and selection of environmental habitats, followed by the cloning and sequencing of the metagenomes of microbial communities of selected soil and aquatic habitats. Thus a total of nine metagenomic libraries was produced from selected environmental habitats. Moreover, existing and newly generated metagenomic datasets were explored for promising gene sequences. The libraries produced were then subjected to educated activity- and sequence-based screenings for, and analyses and engineering of, target enzymatic activities. These desired activities were derived from industrial demand, and focus in particular on improved enzymes involved in the biodegradation of recalcitrant and xenobiotic molecules, including novel chitinases, laccases/ligninases, and aerobic and anaerobic dehalogenases (halogenases as a spin-off). Ammonia lyases were desirable targets too, but are more difficult to screen for. Upon recommendation by the midterm review panel, an emphasis was placed on beta-amino transferring enzymes as well. A strong focus was further placed on the mobilome, that is, the collective pool of mobile genetic elements (MGE, with an emphasis on plasmids) that are carried by a microbial community, as there is strong evidence that the frequency of occurrence of genes encoding the desired enzymatic activities is raised in this gene pool. Selected genes/operons found to encode the desired novel enzymatic functions, in particular novel chitinases, laccases, (de)halogenases and beta-selective amino transferases, were analysed at the molecular level and expressed in expression vectors in suitable hosts. Subsequently, the enzymes were characterized with respect to structure and function (kinetics), and selected ones used in directed evolution experiments to enhance or extend their activities to a practical level. Moreover, selected chitinases were produced at pilot scale, paving the way towards application. However, the project did not fully tackle the practical issues involved in industrial production. Finally, a set of tools that enhances the metagenomics-based exploration of natural habitats was developed to foster our capability to explore natural habitats for novel enzymes. To safeguard the developed expertise and genetic resources, databases (data warehouse) were established (containing the collective functional genes and operons with novel activities found in the project) and the metagenomic libraries produced stored for future explorations.
Project Context and Objectives:

Project objectives

General - The Metaexplore project develops and applies advanced molecular tools that allow the cloning and sequencing of the metagenomes of microbial communities of selected soil and aquatic habitats, followed by educated activity- and sequence-based screenings for, and analyses and engineering of, target enzymatic activities. These desired activities are derived from industrial demand, and focus in particular on improved enzymes involved in the biodegradation of recalcitrant and xenobiotic molecules, including novel chitinases, ligninases/laccases, and aerobic and anaerobic dehalogenases (halogenases as a spin-off). A strong focus will be placed on the mobilome, that is, the collective pool of mobile genetic elements (MGE, including plasmids) carried by a microbial community, as there is strong evidence that the frequency of occurrence of genes encoding the desired enzymatic activities is raised in this gene pool. Selected genes/operons found to encode desired novel enzymatic functions will be analyzed at the molecular level and expressed in optimized expression vectors in suitable hosts. Subsequently, the enzymes will be characterized as to structure and function (kinetics), and used in directed evolution experiments to enhance or extend their activities to a practical level. This will pave the way towards application; however this project will not fully tackle the practical issues involved in industrial production. Finally, a database is established that contains the collective functional genes and operons with novel activities found in the project.

List of objectives as in the work plan across the whole project:

- Prescreening of preselected terrestrial and aquatic habitats, using chemical and genetic tools (WP1)
- Construction of environmental metagenome libraries (WP1)
- Construction of environmental mobilome libraries (WP1)
- Storage, maintenance and preservation of the metagenome/mobilome libraries (WP1)
- Development of screens for the detection of chitinolytic activity (WP2)
- Functional screen of chitinolytic activity in metagenomic libraries to identify metagenomic clones carrying this activity (WP2)
- Develop genetic tools to identify and isolate putative chitinolytic genes and/or promising clones (WP2)
- Development of screens for the detection of laccase activity (WP2)
- Functional screen of laccase activity in metagenomic libraries to identify metagenomic clones carrying this activity (WP2)
- Develop genetic tools to identify and isolate putative laccase genes and/or promising clones (WP2)
- Screening of metagenomic libraries for enzymatic activities relating to the degradation of dehalogenases and/or ammonia lyases (WP3)
- Genetic screening of metagenomic libraries for novel halogenases (WP3)
- Validation of the expanded gene library for dehalogenases, ammonia lyases and halogenases (WP3)
- Sequencing of selected metagenomic library clones that encode enzymes with desired function and catalytic activity (WP4/5)
- Setup of a bioinformatics platform for storage and management of metagenomic data sets including mobilome nucleotide sequence data (WP5)
- Development and design of a data warehouse collecting information on target genes and enzymes (WP5)
- Implementation of novel methods for searching sequences encoding target functions (WP5)
- Development of expression system for metagenomic libraries in Trichoderma reesei or Saccharomyces cerevisiae, allowing expression of eukaryotic genes (WP6)
- Development of a vector system that can be used to manage and screen metagenomic libraries in different hosts (WP6)
- Creating libraries from environmental samples by first capturing metagenomic DNA with conventional cloning techniques into an E. coli vector (WP6)
- Transfer of libraries to screening vectors suitable for use in the screening host T. reesei (optionally in S. cerevisiae) (WP6)

List of achievements, in main terms, per work package (WP):

WP1 - The screening of selected environments and construction of (nine) metagenomic clone libraries for useful enzymes was successfully completed. As reported, the habitats sampled included soils and soil-related (chitin and/or lignin - treated soils, peat bogs, biofilters, rhizospheres) as well as aquatic (river bank sediments, wastewater, sponges) environments, to which the Consortium partners have access. Thus, nine habitats were selected for the production of high-quality environmental DNA and metagenomic library construction in fosmids in Escherichia coli. To these, existing libraries were added (Baltic sea, steam-exploded lignocellulose, collected fungi, park grass soil). Moreover, in a later stage (see report 2), the mobilome (i.e. the collective mobile genetic elements – MGE) was sampled by cell detachment and subsequent separation, massive cell lysis, DNA extraction and separation of plasmid from
chromosomal DNA. Mobilome DNA was preserved in microtiter plates at -80°C and subjected to high-throughput sequencing analysis. However, the approach yielded preferentially DNA from small plasmids, making a complete analysis of plasmid genes (including those of bigger plasmids) not feasible. As to host/vector systems to be used, a heterologous expression system based on the fungus Trichoderma reesei was developed and a bacterial plasmid-based system for expression outside of E. coli is under construction (see also WP6).

WP2 - Screenings of libraries were performed for the target enzyme activities in the libraries produced used activity- (activity screening) and gene sequence (genetic screening) - based approaches. Activity screening was directed towards gene functions for the degradation of both natural compounds (chitin and lignin), however was difficult for laccases/ligninases. We also selected for growth on the target compounds (i.e. chitin, lignin) or on products of the initial biodegradation. Genetic screening was based on selected genetic tools developed or adapted in the course of the work (hybridization and/or PCR based), i.e. specific family 18 and family 19 glycosyl hydrolase PCRs next to specific 2- and 3-domain bacterial laccase PCR. The library screenings yielded a pool of promising fosmid clones that was of use for further work under WP5.

WP3 – Screenings for dehalogenases/halogenases and ammonia lyase/epoxide hydrolase activities were completed. As previously reported, available libraries were screened by activity and growth selection and assays. Overall, positive hits were rare. In conformity with the recommendations of the midterm panel, the focus was then placed on metagenomic as well as genomic sequence-based assessments. In the shifted focus, specific amino transferases were included in the screens.

WP4 - The project has yielded a detailed characterization of key selected genes on fosmids, and the functions they encode. Main emphasis was placed on the fosmids yielding chinatinases chi18H8, 53D1 and G9_54, next to several laccases and biochem-relevant enzymes. Following extensive characterization of genetic backgrounds, cloning and expression, we were able to establish the organosolvent-tolerant chinatinase chi18H8, the salt-tolerant chinatinase 53d1 and the key streptomyctical putative chinatinase G9_54. In terms of laccases, one main bacterial one next to several fungal ones were expressed. In addition, several industrially important novel enzymes were also identified and characterized. The genes encoding these enzymes were mainly expressed in E. coli, Trichoderma reesei, Streptomyces. We also performed detailed biochemical analyses of some of the selected enzymes (amino acid sequence, folding, target active site, functioning, substrate specificity, kinetic properties) and expression (genes involved, regulation thereof, optimal host for expression).

WP5 – High-throughput sequencing was applied to several metagenomes in order to shed light on genes encoding target functions present in these. For this purpose, several chitin-treated and untreated (control) soils, next to sponges, were selected, on the premise that the former (soil) samples would yield data on the chitin-biodegradative populations, whereas the sponge data would describe sponge-directed diversity next to information on genes for halogenating enzymes. In addition, fosmids carrying the original metagenome fragments expressing desired functions were and are completely sequenced. This included the genes flanking the desired gene functions, and extended into a trial of meta-fosmidome sequencing, implying the direct sequencing of fosmid pools having hundreds of clones followed by bioinformatics-supported sorting out of the different fosmids. Moreover, collected plasmidomes (several partners) were subjected to high-throughput sequencing as well. An important asset in this work package was the bioinformatics tools and sequence databases that were developed at partner UB (WP5); the latter is meant to serve as a (sequence) data repository.

WP6 – In the project, the development and application of the different novel tools was completed. First, the fungal (Trichoderma reesii) host/vector system was successfully developed and is already being routinely applied. The project task aimed at the design of novel host/vector systems met with partial success; a plasmid of extremely broad host range, of the PromA group (pIPO2) was used as the basis, and the mobilization and replication functions, next to a multiple cloning site and a selectable marker, were cloned and combined into one small vector. This vector is theoretically able to hold up to 20 kb of exogenous DNA, and transfer to and replicate in a range of hosts. Thus, although not completely ready, we obtained proof-of-principle that this novel vector will work as a novel improved trans-host tool. In addition, the long PCR tool (reported earlier, and substituting the enhanced expression tool that was originally proposed) was successfully applied in analyses of the accessory genes present in the insertional hot spots in the plasmids that were screened from biofilters.

WP7 - This work package concerned the organisation and management of the project. Next to nine project meetings and numerous sub-project ones, one key meeting encompassed meeting with Industry (October 2013, Groningen, together with the MicroB3 project). In the latter meeting, possibilities for further exploration of mentioned enzymes and technologies were sought.

Project Results:
The 5-year METAEXPLORE project encompassed seven work packages, under which 17 partners worked together in varying constellations. Overall, the majority of objectives over the whole set of work packages (1 through 7) of the project has been achieved. Hereunder, an account of the achievements per work package (WP) are detailed.

WP1:

Using a suite of 28 environmental habitat types (each one replicated or divided in subsamples; see Table 1), screens for the desired enzymatic activities and functions/genes, i.e. chitinases, laccases, (de)halogenases and the mobilome, were performed across several partner labs (mainly 1, 3, 5, 6, 9 and 14). This task was thus accomplished, allowing us to make an educated choice as to the environments to be metagenomically explored for interesting enzyme activities. Moreover, nine habitats were selected for library construction, and libraries were thus produced and stored, in conformity with the project plans (see hereunder).

Table 1 – Environmental habitats sampled and their characteristics regarding selected enzyme groups and the mobilome

Sampled habitat | Relevant characteristics | Compounds /chemistry | Chitinases (18/19) activty pres/div | Laccases activ pres/div | (De)halog. presence | Mobilome Inc groups |
---|---|---|---|---|---|---|
Vredepeel (V) soil 1 Agricultural soil | Organic matter (OM) | -/+ | +/+ | +/+ | P1*, P7*, P9* |
V soil, chitin treated | Chitin | -/+ | +/+ | +/+ | P1*, P7*, P9* |
Oxyria digyna 1 Arctic plant rhizosphere | Plant-released compounds | -/+ | +/+ | +/+ | n.d. |
Spent mushroom substrate 1 Organic-rich | Mushroom waste | +/+- | +/+ | +/+ | P1*, P7, P9* |
Wood filter 1 Wastewater treatment | Wastewater organics | +/+- | +/+ | +/+ | n.d. |
Ephydatia uviatilis 1 Freshwater sponge | Lignin, chitin | +/+- | +/+ | +/+ | P1* |
Haliclona panicea 6 Marine sponge | Chitin | +/- | n.d. | + P1 |
Corticium candelabrum 6 Marine sponge | Chitin | +/- | n.d. | + P1 |
Petrosia ciformis 6 Marine sponge | Chitin | +/- | n.d. | + P1 |
Bog soil 14 Peaty soil | High OM | -/+ | +/+ | P7* |
Fen soil 14 Peaty soil | High OM | -/+ | +/+ | n.d. |
Biofilters*: LE / PC/Li/Kort/Koks 10 Filter soil for pesticide cleanup | Pesticides | -/+- | +/+ | +/+ | P1, P9 |
VTT compost 11 Compost, treated OM | | +/- | +/+ | +/+ | P1, P7, P9 |
Cayo (C) soil 5 Agricultural soil | Chitin (beta) | +/+- | +/+ | P7, P9 |
C soil, chitin treated | Chitin-treated chitin | +/+- | +/+ | P1, P7 |
Whitby (W) soil 5 Pristine | Organic matter | +/+- | +/+ | n.d. |
W, shell waste 5 Chitin-treated Chitin | +/+ | +/+ | P1, P7, P9 |
Sourhope (S) soil 5 Pasture | Organic matter | +/+- | +/+ | n.d. |
Cryfield soil 5 Pasture | Organic matter | +/+- | +/+ | n.d. |
Park grass soil | | | | | |
EA4/F48/E1/L01/ 7 Agricultural soil | Organic matter | +/+- | +/+ | +/+ | n.d. |
Uppsala (U) soil 4 Suppressive | Organic matter | +/+- | +/+ | +/+ | n.d. |
Canal oeste 15 Canal water | Organics | +/- | +/+ | P1, P7, P9 |
Arroyo el Gato B1 / B2 / B3 | Canal water | Organics | +/- | +/+ | P1, P7, P9, N, Q |
Musselbank 15 Water channel | Chitin | +/+- | +/+ | P1, P7, P9 |
Selva marginal 15 Natural forest | Lignocelluloses | +/+- | +/+ | P1 |
Canal Sarandi E1 / E2 / E3 15 Canal | Organics | +/+- | +/+ | P1, P7*, P9*, N |
Baltic sea sediment 4 Marine sediment | Organics, chitin | +/+- | +/+ | +/ P7 |

P = partner providing sample. Chitinases, laccases: +/-: present and diverse (diverse: > ~10 even types, evidenced by molecular screens); +/+-: present and low (< ~8 types) diversity; +/-NA: not detected by genetic screens. (De)halogenases: evidence for presence of
genetic identity (using a p7 probe / p6 primers) is shown. Chitinases: family-18 >> family-19 (Park grass).

Screening for chitinases - Partners 1 and 5 screened environmental samples provided by the Consortium for chitinolytic enzyme activity and for the presence of family-18 chitinase genes, e.g. using primers of Williamson et al. (2000) targeting the chiA gene. Partner 1 also applied PCR-DGGE to ascertain chiA gene diversity. Partner 5 screened for family-19 chitinases. Among others, they found Sourhope soil to reveal a remarkable diversity.

Additionally, partner 1 performed chiA gene-targeted pyrosequencing in collaboration with partner 3, in order to identify the diversity and novelty of chiA across the selected environments (soil, arctic rhizospheres, freshwater and marine sponges, woodchip, spent mushroom substrate). Furthermore, a soil microcosm experiment (with chitin amendment and pH biasing) was performed in order to enhance the chance to detect novel chitin-degrading enzymes (partner 1).

Screening for laccases - Partner 14 applied direct phenol oxidase activity measurements as a proxy for habitat laccase activity across samples. They also developed and applied a new tool for genetic screening for bacterial (2-domain) laccases, with a focus on acidobacterial ones. They found laccases in most of the screened habitats and also described their diversities. The data served to make educated guesses as to the most promising laccase-harbouring habitat to sample for metagenome preparation.

Screening for (de)halogenases - Partner 6 screened samples provided by the Consortium for the presence of (reductive) dehalogenases using three primer sets: one broad-range (dehalo F4 – dehalo R2) and two targeting specific groups of producers (RD2S-F and RD5r-R for Desulfitobacterium spp. and RRF2 and B1R for Dehalococcoides spp.; Table 2). They thus found putative reductive dehalogenases to be present in most environmental samples. Pyrosequencing of amplicons from positive samples will be performed to identify the diversity and novelty of reductive dehalogenases found. Pre-screening for halogenases was started but was hampered due to ongoing primer validation. This has now been accomplished, and full-scale screening of the samples was started and reaches completion (in progress).

Table 2 – Primers developed for detection of dehalogenases

Partner 10 screened a range of biofilters used in wastewater cleanup of pesticides for (chloroaromatics dehalogenase) activities and found several to be highly active.

Screening for mobilome sequences - Partner 9 screened DNA samples received from partners (1, 5, 7, 14) for IncP-1, IncP-7 and IncP-9 sequences by PCR-Southern blot and qPCR analyses. Further, the diversity of these plasmids in biofilter 5 was studied by cloning and sequencing PCR products amplified from total community DNA. Partner 3 assessed the biofilters used for cleaning pesticide contaminated water and selected it as the environment with greatest potential for mobilome gene mining. Partner 15 performed analyses on pooled plasmids from cultivated organisms. Partner 1 analyzed exogenously isolated plasmids from forest soil.

On the basis of the prescreening for diverse activities (chitinases, laccases and (de)halogenases), a total of nine habitats were selected, that is, five provided by partner 1 (chitin-amended and unamended soil, bog soil, sponge and Arctic plant rhizosphere), two by partner 5 and two by partner 6, for high molecular weight DNA extraction and metagenomic library construction in fosmids (listed in 1.4). Partner 14 furnished the DNA for one (bog soil) library. Partner 6, on the basis of the prescreening described in task 1.1 selected marine sponge samples for the libraries rather than soils. All libraries that were foreseen in our project plan have been constructed and are being stored as genetic resources for future work.

WP2:
Work package 2 encompassed screening of metagenomic libraries for enzymatic activity relating to the degradation of the recalcitrant natural compounds chitin and lignin, as a component of lignocelluloses, and the validation and preliminary analysis of selected promising clones as summarized in the following achievements:

Metagenomic libraries of the consortium were screened for chitinolytic and lignolytic/laccase activities using either activity based or genetic assays, or a combination of both. Libraries included nine novel, provided through the project, and other metagenomic libraries available to the consortium, i.e. from reference soil of Rothamsted, a eukaryotic (fungal) metagenome, a mesophilic fungal pool, a suppressive soil, a wood chip enriched compost. In addition, one mobilome/plasmid library, pools of isolates and metagenomic sequence data sets of consortium partners were screened. A suit of efficient screening assays, protocols and approaches have been developed and optimised for high throughput, robotic, and conventional screening of the large metagenomic- and mobilome/plasmid-libraries, for the identification of clone candidates from both enzymatic groups. Genetic pre-sieving has been employed in some cases when the number of enzymatically false positives was high, improving efficiency in hit rates. For the larger libraries, techniques of
pooling of clones were successfully employed.

Chitinolytic enzymes
Major efforts were placed on the development of suitable activity-based assays for chitinolytic enzymes. These include three different growth selection assays on agar plates, a method to detect the production of chitosan and a number of enzymatic microtiter plate assays, for example detecting chitobiosidase or N-acetyl glucosaminidase activity with fluorogenic chito-oligosaccharide substrates or chitodextrinase and chitin deacetylase, the latter for the identification of enzymes that may provide an alternative or a complementary approach for the biotechnological production of chitosan from chitin. Functional assays have also been developed to investigate the chitinolytic system of Nonomuraea sp. ATCC 39727, an uncommon actinomycete with biotechnological potential to produce bioactive molecules.

Genetic screening protocols that have been developed include PCR-based (e.g. ChiA, chitinase family 19, chitin deacetylase), robotic hybridization using probe and in silico mining of metagenomic sequence data and the application HMMs, for chitinase family 18 and chitin binding domain, developed within the consortium.
A diversity of clone candidates with chitinolytic activity have been identified through activity or genetic screenings. A set of promising clone candidates with high activity and/or novel sequence were selected for further sequence analysis, clone validation and preliminary characterization. The most promising novel chitinolytic enzymes or genes include:

- Chitinase Chi18H8, a novel family 18 group II from Swedish suppressive soil with high chitinase activity. Chi18H8 shows predominant chitobiosidase (MU-NAG2), some MU-NAG3 activity but also antifungal activity against four crop pathogens which makes it promising for applications as antifungal biocontrol agent as well as a novel chitinase for biotechnological applications. The gene was isolated from the fosmid clone, sequenced and subcloned into an expression vector for further characterization and protein production optimisation within work package 4.
- Family 18 chitinase of fosmid 53D1, from chitin amended Dutch agricultural soil, identified by PCR screening. Protein 53D1 showed chitobiosidase (MU-NAG2) activity and was subcloned for further characterization within work package 4.
- Two fosmids, D7 and G9, with high chitinase activity against MU-NAG2 and MU-NAG3 identified from a chitin-amended microcosm with test soil from UK were sequenced and sub-cloned. The A multidomain chitinase, G9_54, predicted to be a member of the less common Family 19 chitinases (with an additional putative alpha-L-rhamnosidase domain), was transferred to work package 4 for further characterization.
- A putative polysaccharide/chitin deacetylase, selected out of several interesting clone hits, from untreated Rothamstedt soil, identified genetically and characterized for sequence characteristics, codon usage and compatibility for expression in E. coli.
- A gene for a putative new family 19 chitinase identified by in silco analyses of S. meliloti mobilome/plasmid sequence data was chemically synthesised with optimized codon usage for E. coli and cloned into the expression vector pET22b+.
- Two clone candidates (fosmids) of a putative chitin deacetylase and a family 19 chitinase, respectively, were identified through in silico screening 454 sequence data of Baltic Sea sediment library.

Laccases
Also for laccase screening, a suite of assays and approaches have been developed across partners. For activity-based (enzymatic microtitre plate assays, growth selection assays on agar plates using substrates such as ABTS and 2,6-DMP) and genetic (PCR, full-length inverse PCR, hybridization or in silico screening sequence data using an HMM developed for laccase within the consortium) assays, protocols and approaches have been developed for identification of bacterial and fungal ligninolytic enzymes. Approaches include efficient robotic high throughput put screening. Some approaches resulted, for unknown reasons, in the discovery of incomplete laccase type genes. Out of the following approaches; functional laccase screening of a mesophilic fungal pool library, screening of a wood chip-enriched compost, screening of a bog soil metagenome and screening of a fungal metagenome, the latter approach was more successful.
These screening approaches have enabled the identification of a set of clone candidates with laccase activity. The most promising novel laccase enzymes_GENES selected for further clone validation and preliminary characterization include:

- Three novel fungal laccase genes (with conserved motifs common in all laccases: 3 multicopper oxidase domains each and the so-called L1–L4 conserved regions with the copper-coordinating amino acids) from a metagenome of 33 thermophilic and 64 mesophilic fungal strains. The genes were subcloned into the expression system Trichoderma reesei, developed within the consortium, and were shown to express activity on the substrate ABTS. Enzymes were further characterized within work package 4.
- A novel three-domain bacterial laccase affiliated to Acidobacteria, from a library of Slovenian bog soil. The gene was successfully expressed as a His-tagged recombinant enzyme and further characterized within work package 4.
Furthermore, a manganese peroxidase activity in Nonomuraea sp. ATCC 39727 was identified through a complimentary approach to metagenomic screening and Streptomyces coelicolor was shown to produce a laccase activity likely attributed to the previously described SLAC (small laccase from streptomyces).

Following clone candidate validation and preliminary characterization in terms of (limited) sequence content and enzymatic activity, the selected chitinolytic or laccase candidate clones and genes described above have been transferred further to WP4 for improved protein production and complete genetic- and enzymatic characterization.

WP3:
In WP3, DNA from the most promising environments was deeply screened for genes for target enzymes, using gene sequence (genetic screening) - based approaches.

Activity screening was used, directed towards gene functions for the degradation of anthropogenic compounds. Activity screens were done by selection for growth on the target compounds (e.g. aliphatic halogenated compounds) or on products of the initial biodegradation, dehalogenation / halogenation or deamination/deamidation reactions.

Genetic screening was based on genetic tools (hybridization and/or PCR based or in silico homology searches).

Metagenomic libraries were functionally examined (partner RUG2) for haloalkane dehalogenases and ammonia lyases using plate-based activity screens and growth screens (Figure 1). However, as described in the mid-term review report, no active clones could be discovered, probably due to the low abundance of target sequences in the libraries. In fact, the representation of low abundance genes in metagenomic libraries may be too low to use functional screening of such libraries as a routine method, and in case of target sequencing, sequence-based approached or strategies that include enrichment steps may be required. Consequently, enriched cultures were used for further study.

Using bioinformatics methods, (meta) genomic sequences were examined by partner RUG2 for the following classes of enzymes: haloacid dehalogenase, haloalkane dehalogenase, ammonia lyases, and aminotransferases. Haloacid dehalogenases were detected in Baltic sea sediment sequences provided by partner SH. These enzymes are well known, and their biochemical properties were not examined further. In these samples, halohydrin dehalogenases were not detected, but we did find haloalkane dehalogenase/epoxide hydrolase gene fragments and a fragment encoding a putative enigmatic haloalkane dehalogenase/adenine deaminase fusion. The properties of these novel enzymes were examined after expression of PCR-amplified genes in E. coli (see WP4).

As justified above, bioinformatics is more extensively used than foreseen. This is made possible by the growth of sequence information, both from isolated cultures and in terms of metagenomic datasets. The contribution by partner RUG2 to the deliverables remained similar, with the difference that validated libraries are replaced by validated expression constructs to be used for biocatalyst profiling and improvement by directed evolution in WP4 (Tasks 4.2 4.3).

Fig. 1. Enzyme reactions considered in this WP by partner RUG2. Enzymes: 1,2: dehalogenase; 3,4: ammonia lyase; 5,6: aminotransferase.

Fig. 2. Degradation of β-valine (partner RUG2). A) Schematic representation of the organization of the ORFs identified on an insert from strain S1 encoding β-valine degradation genes. The sequence was obtained from a whole genome sequencing of a bacterial culture enriched from an environmental sample using β-valine as nitrogen source. B) Proposed bacterial degradation pathway of β-valine. The numbers represent ORFs that are potentially relevant for β-valine degradation. ORF9 encodes a completely new type of CoA-dependent ammonia lyase. Its activity was confirmed in vitro by expressing the gene in E. coli and examining the activity by HPLC.

BLASTP searches were also used to discover new C-N lyase sequences. A promising hit was found in Chelativorans sp. BNC1, an EDTA degrading bacterium. Although the protein is annotated as an arginosuccinylate lyase, it shares about 77% sequence identity to an enzyme from Brevundimonas sp. TN3, called ethylenediamine-N, N'-disuccinic acid (EDDS) lyase. The novel EDDS lyase–like gene from strain BNC1 was synthesized and the encoded enzyme is under study.

Consequently, bioinformatics approaches were also used by partner RUG2 to explore organisms enriched from environmental samples with high biodiversity for genes involved in the metabolism of amino acids, focusing on β-amino acids, esp. β-valine, β-phenylalanine, β-glutamate and related compounds. Crude whole-genome sequences were obtained for 6 of these new isolates. Genome sequences were examined in collaboration with partner Bielefeld, which led to the identification of a new β-valine degrader and 2 strains utilizing β-glutamate. A pathway for β-valine degradation was established, identifying the deamination mechanism as being CoA dependent and dehydratase-like (Fig. 2).
WU worked on the development of a reliable and robust assay for screening metagenomic libraries for the presence of halogenases. Such an assay was not available and a requirement for screening large clone libraries. Two different potential screening methods have been tested:

1. using a colour change of eosin-methylene blue as a response to a redox change associated with halogenation
2. using the conversion of fluorescein (green) to eosin (red) to bromination of fluorescein

Twenty halogenase-positive and putative halogenating strains were compared to twenty strains that did not possess a FADH2-dependent halogenase (as determined by PCR screening). In summary, no consistent and reproducible results were obtained and therefore high-throughput screening for halogenases is currently not possible. This is due to the indirect measurement of activity and activities recorded may be caused by other redox reactions in the cell.

The latter is a general issue with catabolic reactions, opposed to the anabolic reactions studied for the other enzymes in MetaExplore as the final halogenated products are very diverse and mostly unknown.

Therefore, we have assessed PCR primers based on in silico analysis of known FADH2-dependent halogenase encoding genes and PCR primers previously published (table 3.1). The Halo-B4-F – Halo-B7-R primer set showed a preference for actinomycete-derived halogenases, while proteobacteria-derived halogenases were not consistently picked up. The Half – HalR1 and Half – HalR2 primer sets were deliberately biased towards FADH2-dependent halogenases from proteobacteria. They were indeed discovered from a metagenomic DNA preparation from the marine sponge Halichondria panicea, but the number of false positives was outnumbering the number of hits. Therefore, a PCR mix of two forward and two reverse primers to detect both classes of FADH2-dependent halogenases (tryptophan halogenases & phenol and pyrole halogenases) was successfully applied on metagenomic DNA preparations from marine sediments, and sponges.

Table 3: Overview of FADH2-dependent halogenase primers assessed.

With primer mix HA002 and HA005 (that performed best) we assessed marine sediments (Black Sea and Mediterranean) and marine sponges (Aplysina aerophoba and Halichondria panicea), fosmid libraries of the sponges Crambe crambe and Halichondria panicea, isolates of these two sponge species, and the environmental samples mentioned in WP1 for diversity of FADH2-dependent halogenases.

For all putative halogenases obtained, it was confirmed that they are halogenases by comparison to the NCBI database of microbial proteins and the Conserved Domains Database. Phylogenetic analysis of the Crambe crambe microbiome-associated halogenases was performed based on amino acid sequence. Three clades of C. crambe-derived putative halogenases could be distinguished of which especially clade III was only remotely related to other previously-reported sequences. Only four sequences were not classified into any of these clades. Overall the halogenase sequences clustered together with the other tryptophan halogenase sequences in the tree, confirming their annotation. The full halogenase sequences obtained from Aplysina aerophoba by Bayer et al. (Mar. Biotechnol. 15, 2013) were phylogenetically distant from the sequences obtained from Crambe crambe microbiota.

Identification of putative halogenases in the MetaExplore environmental sample collection using deep sequencing with the before mentioned primers yielded a high diversity of halogenases ranging from 39 different halogenases in the compost of spent mushroom substrate to 900 for the marine sponge Aplysina lacunosa. We complemented the metagenomics strategy with a microbial isolation-based strategy to obtain halogenases. For C. crambe, using the same primers for PCR screening, only one of the 107 isolates (isolate 23D8) was found to possess a putative halogenase gene. The halogenase sequence from the isolate Crambe 23D8 appeared very closely related to one of the metagenome clones (H11 Cr) and is likely to be the same gene. For H. panicea 1103 isolates yielded 20 different FADH2-dependent halogenases. In addition a SAM-dependent halogenase was obtained from the marine sediment sample.

In view of the conclusions of the mid-term review, partner 2 (RUG2) has focused on enzyme discovery for: 1) functional metagenomic library screening approaches for dehalogenases and lyases; 2) bioinformatics approaches, so sequence analysis of metagenomic sequences and sequences from enriched strains; and 3) enrichment approaches, in order to recover genes that escape detection due to low representation in metagenomic libraries. From the isolated strains (see below), genomic libraries were constructed, which were subjected to functional screening in a workflow that merges with screening for active clones in metagenomic libraries. Target enzymes were dehalogenases, ammonia lyases and aminotransferases.

Over the course of the project, Partner 2 (RUG2) has constructed gene libraries in the broad host range vector pLAFR3 of 4 enriched strains: a) Variovorax paradoxus isolated on β-phenylalanine; b) a Rhizobium sp. isolated on an α,α-disubstituted amino acid, 3) strain Pseudomonas SBV1 isolated on β-valine; 4) a 2,3-dichloropropanol degrading strain of Pseudomonas putida called MC4.

These and several (meta)genomic DNA libraries obtained earlier in the pZERO vector were transformed in E. coli and recombinants were
tested for growth on:
• azidoacetic acid
• N-methylaspartic acid
• N-butylaspartic acid
• β-phenylalanine (!)
• β-leucine
• β-valine (!)
• β-glutamate
• β-alanine
• α-methyl-β-alanine
• β-lysine
• n-butylamine
• D-aspartate
• D-para-hydroxyphenylglycine
• other primary amines

Positive hits that could be reproduced were never found, with the exception of transformed E. coli clones utilizing β-valine and β-phenylalanine (marked with “!”). These positives were, however, transformed with DNA libraries obtained from enriched cultures, not with DNA cloned directly from environmental samples. This observation is in agreement with the low representation of uncommon genes in metagenomic libraries.

Enrichment approaches were done by RUG2 on the following compounds and yielded the strains listed below. Growth spectra were tested. However, quite some compound appeared recalcitrant and never gave growth.
• β-valine. A strain of Acidovorax was obtained, called MG01. A second isolate on this compound was also obtained, called M7A. This strain appeared as a mixture. Strain M7A belongs to Variovorax paradoxus and also grows on β-Phe, β-Tyr, β-Leu, and β-Glu. A putative β-Phe aminotransferase was indentified by BLAST searches; the putative protein was 84% identical to a similar aminotransferase recently studied in our lab. Strain M7B belongs to Herbaspirillum, which occurs also as an endophytic diazotroph.
• β-tyrosine
• N-Phe-2-oxazolidinones
• the triazole ipconazole: slow degradation by fungi.

The enrichment cultures were subsequently used for discovery of specific genes using three approaches: 1) construction of gene libraries in a broad-host range vector with functional screening by complementation to phenotypically negative strains; 2) enzyme assays to identify and characterize the target protein; 3) whole genome sequencing and bioinformatics analysis; 4) proteomics, looking at different expression profiles on different growth conditions. In our hands and in these cases, approaches 2) and 4) worked best, and led to the identification of new dehalogenases, lyases and aminotransferases.

Thus, although in a strict sense the approach deviated from the task described in the DoW, several organisms and genes could be obtained by partner RUG2 and used for further characterization in WP4.

WU further analysed the isolates and clones that were positive for the presence of halogenases by whole genome sequencing of selected isolates/clones (see report for Metaexplore report 3).

As halogenases differ in the reactions catalysed when compared to dehalogenases, ammonia lyases and aminotransferases in that they are anabolic enzymes involved in the biosynthesis of halogenated secondary metabolites instead of catabolic genes involved in degradation pathways. This characteristic has a major impact on the clone/isolate selected for continuation in WP4 as the potential substrate for the halogenase must be known. Since this information is not available for most halogenases that were identified we selected a marine sediment-derived halogenase for further analysis in WP4. This enzyme was characterized as a SAM-dependent halogenase and halogenation of these enzymes takes place in the first step of the pathway, using S-adenosyl methionine (SAM) as substrate. In order to optimise gene expression in E. coli we used a codon-optimised version of the gene for E. coli.

WP4:
This work package analyzed the clones containing genes/operons for the enzymes selected under WP2 and WP3, and the enzymes themselves, so as to understand the genetic background, next to the protein structure (amino acid sequence, homology models, tertiary and quaternary structure, active site, etc.), function (metabolic role, substrate specificity, kinetic properties) and expression (genes involved, regulation thereof, optimal host for expression). In the analyses, we used selected fosmids and genes/activities with target properties. The analyses were primarily performed by partners who dealt with the specific enzyme classes, as in WP2 and WP3. Partner RUG2 expressed a range of representative genes identified by bioinformatic analysis of (meta)genome sequences. This
includes:
- a CoA ligase involved in β-valine activation, and catalyzing the first step in β-valine metabolism.
- 2 novel ammonia lyases involved in deamination of β-valinyl-CoA.
- 1 novel aspartate ammonia lyase
- 3 aminotransferases involved in β-phenylalanine aminotransferase reactions
- 1 haloalkane-dehalogenase-epoxide hydrolase like enzyme
- 1 fusion enzyme of a haloalkane dehalogenase-adenine deaminase.

See Fig. 1 for the reactions catalyzed by these enzymes. Further details about these enzymes are included in the deliverable reports. In all cases, expression could be achieved in E. coli using standard conditions. e.g. pBAD vectors with induction by arabinose and pET vectors with induction by IPTG.

Partner RUG2 also established that the DYG type α/β fold hydrolase (Table 4) resembles a hybrid between an epoxide hydrolase and a dehalogenase. The putative gene that was expressed comes from Xanthobacter autotrophicus and protein was again produced in E. coli. Although these enzymes resemble dehalogenases the enzyme has epoxide hydrolase activity. Molecular modelling could not yet identify a well-converted native substrate for the enzyme. These hybrid enzymes are quite common and represent a class of epoxide hydrolase of which the natural function remains to be established.

Partner WU obtained a variety of different novel halogenase genes (see WP3). A halogenase (designated as marine sediment-derived halogenase MDSH) originally isolated from a marine sediment was selected for further genetic analysis and gene expression. The E. cloni 10G Chemically Competent Cells, a modified E.coli strain produced by Lucigen, was used as the heterologous expression platform of MDSH. The on-line GeneOptimizer software (Life technologies) was used to adjust the codon usage of the MDSH, for expression in the E. cloni 10G platform, to optimize the GC content of the salL, to eliminate killer motifs and the formation of secondary RNA structures. The optimized MDSH gene was ordered as a ready to use linear DNA fragment, using the GeneArt Gene Synthesis service Lifetechnologies). The 882bp long, optimized MDSH DNA fragment was designed with 3’ and 5’ extensions, which are required for ligation with the pRham C-His DNA vector. The Expresso Rhamnose Cloning and Expression System (Lucigen), was employed for the transformation of E. cloni 10G. The system is based on the linearized pRham C-His DNA vector, which contains the tunable rhamnose promoter for controlled protein expression. Soluble MDSH protein production was optimised by adapting the incubation temperature from 37 °C to 25 °C.

Biocatalyst profiling (partners 2, 5, 10, 11, 13) by high-throughput screening, will be characterized with emphasis on the following characteristics: substrate range, kinetic parameters (kcat, Km, substrate and product inhibition), and operational stability (pH and temperature stability, (co)solvents). The best enzymes will be made available to industrial partners with whom links exist through ongoing collaborations, and investigated further in respect to their overall and active site structure.

In agreement with Task 4.2 using enzymes recombinantly expressed in E. coli and purified by affinity tag chromatography, partner RUG2 has carried out activity assays and selectivity profiling with novel α/β-hydrolase fold enzymes, aminotransferases and ammonia lyases.

Table 4. Overview of biocatalyst profiling experiments.

<table>
<thead>
<tr>
<th>Enzyme Function</th>
<th>Properties</th>
<th>Other substrates</th>
</tr>
</thead>
<tbody>
<tr>
<td>a CoA ligase involved in β-valine activation</td>
<td>catalyze the first step in β-valine metabolism. See Fig. 2.</td>
<td>CoA ligase homolog very specific, only known to accept β-valine</td>
</tr>
<tr>
<td>2 novel ammonia lyases involved in deamination of β-valinyl-CoA.</td>
<td>catalyze the second step in β-valine metabolism. See Fig. 2.</td>
<td>enoyl hydratase homolog with structural alterations allowing acceptance of amino instead of hydroxyl very specific, only known to accept β-valinyl CoA</td>
</tr>
<tr>
<td>1 novel aspartate ammonia lyase</td>
<td>aspartate deamination to form fumarate aspartase homolog thermostable</td>
<td></td>
</tr>
<tr>
<td>3 novel aminotransferases of class I PLP enzymes</td>
<td>bacterial metabolism of β-amino acids similar to established β-phenylalanine amino- transferases from Variorovorax paradoxus and Mesorhizobium sp. Subjected to homology modeling and docking simulations.</td>
<td>like other β-phenylalanine aminotransferases: several other β-amino acids can be converted, but highly selective for β-amino acids and R-enantiomers</td>
</tr>
<tr>
<td>1 new haloalkane-dehalogenase-epoxide hydrolase like enzyme</td>
<td>unknown so-called DYG type α/β-fold hydrolase. Well expressed in E. coli, but no clear activity detected with various epoxides and halo-alkanes. no detectable activity with several substrates tested.</td>
<td>1 fusion enzyme of a haloalkane dehalogenase-adenine deaminase. unknown fusion enzyme, iden-tified by sequence analysis. Deamination and dehalogenase activity confirmed. dehalogenase activity only detected with 1,2-dibromoethane.</td>
</tr>
<tr>
<td>EDDS lyase from Chelativorans sp. BNC1</td>
<td>deamination of ethylenediamine to fumarate aspartase homolog addition of 1,3-</td>
<td></td>
</tr>
</tbody>
</table>
diaminopropane, 1,2-diaminopropane, 1,3-diamino-2-propanol or 2,3-diaminopropionic acid to fumarate.

The putative EDDS lyase gene was purified from the EDTA degrader Chelativorans sp. BNC1 (partner RUG2). Kinetic and 1H NMR spectroscopic studies showed that this protein catalyzes the reversible addition of ethylenediamine to fumarate to give ethylenediamine-N,N′-disuccinic acid (EDDS), thereby identifying it as a functional EDDS lyase. The enzyme also catalyzed the addition of 1,3-diaminopropane, 1,2-diaminopropane, 1,3-diamino-2-propanol or 2,3-diaminopropionic acid to fumarate. The substrate specificity of this EDDS lyase is limited to small diamines, with no measurable activity towards monoamines.

Partner RUG2 also characterized thermostable variants of limonene epoxide hydrolase and of haloalkane dehalogenase. A 12 fold mutant of limonene epoxide hydrolase which is 35°C more stable and has a higher temperature optimum, was still highly enantioselective. An 8-fold mutant of haloalkane dehalogenase LinB, which was 25°C more stable, was more active than wild-type at its optimum temperature, and also more resistant to some organic solvents. This allowed addition of solubility-enhancing cosolvents and thereby conversion of the highly recalcitrant β-hexa-chlorohexane isomer and also enantioselective conversion of a chiral bromocarboxylic acid ester. The enhanced stability thus widened the biotechnological scope of the enzyme.

These enzymes are available as purified enzymes that can be produced in hundreds of milligrams and can be delivered freeze-dried to any partner interested.

Using HPLC, it was tested by partner WU whether the marine sediment-derived halogenase (MSDH) was capable of catalyzing the reaction from S-adenosyl methionine (SAM) to 5-chloro dideoxyadenosine (5-ClDA). Incubation times up to 6 hours were tested. For all tested incubation times (partial) conversion from SAM to 5-ClDA was observed while controls to which no SAM or no enzyme was added remained negative. The MSDH was also capable of catalyzing the reverse reaction from 5-ClDA to SAM. This experiment proved the halogenase function of MSDH.

Directed evolution (partners 2, 6) Directed evolution tools were used to generate novel variants with improved characteristics. Partner 6 did not participate in this task. Partner 2 (RUG2) developed a novel strategy for design of mutant libraries for directed evolution of enzymes aimed at enhancing thermostability. The strategy is called FRESCO: Framework for Rapid Enzyme Stabilization by Computational libraries (Wijma et al., 2014). First, energy calculations on all possible point mutants and design of all possible disulfide bonds are used to generate a series of mutant enzymes. Next, molecular dynamics simulations and scoring for local flexibility (RMSF) are used to rank promising variants. After using these high-throughput computational tools, the highest ranking variants are expressed and tested in the laboratory. Experimentally confirmed beneficial mutations are combined, again after testing multiple combinations by molecular dynamics simulations. This gave a spectacular stability increase of different enzymes in only two or three rounds of laboratory testing. For further details, see deliverable report D4.3.

The methods were developed with limonene epoxide hydrolase, and were subsequently successfully applied to 4 other enzymes: haloalkane dehalogenase (LinB, Floor et al., in press), adrenodoxin of the P450scc system, peptide amidase for peptide deprotection and activation (Arif et al., in press; Wu et al., in preparation), and halohydrin dehalogenase (HheC, Arbanejad et al., in preparation). The latter 3 examples concern work done in other projects, and indicate the general relevance of the work for enzyme engineering. Partner RUG2 also explored a strategy (CASCO, Catalytic Sites by Computation) aimed at tailoring enzyme selectivity, and in which most experimental screening is replaced by computational methods. The strategy was applied to the redesign of a thermostable variant of limonene epoxide hydrolase for conversion of cyclopentene oxide to optically active diols, which requires different substrate binding modes in the active site (Fig. 4). First, active site redesign towards binding of transition state models is performed with Rosetta software, which produces a range of promising designs that should bind substrate in a reactive position for the desired selectivity. Next, the reactivity of the bound substrates is estimated by calculating the frequency of occurrence of near-attack conformations during molecular dynamics simulations.

Twenty-nine epoxide hydrolase variants were computationally designed and tested experimentally. Of the variants that were designed
to be (S,S)-selective, 63% was correctly predicted, and of the variants designed to be (R,R)-selective 85% was correct. Both for (S,S)-dial and (R,R)-dial enantioselectivities exceeding E=80 were observed. The final enzyme variants obtained were as enantioselective as those obtained by a directed evolution experiment, but with strongly reduced experimental screening. A publication about this work is in preparation Wijma HJ, Floor RJ, et al. manuscript in preparation).

The methods developed as part of this task are also offered and applied in the context of the BE-Basic project to an industrial partner (DSM). Furthermore, they are applied in a EU FP7 project called Kyrobio, aimed at developing enantioselective catalysts for preparation of chiral compounds.

Details of this results of this task and dissemination are given in deliverable report D4.3.

References (WP4)

Production (partners 13, 17):
An extended description of the work for the chi18H8 producer organism is given, whereas similar developmental work was done for 53D1 and G9_54. The protocols previously optimized at flask level for Chi18H8 production and solubilization from inclusion bodies were applied by Partner 13 (UNINS) and Actygea to the recombinant protein production in 3 L bench-bioreactor and in 30 L pilot-scale industrial bioreactor. The C-terminus His6-Tagged Chi18H8 was produced by E. coli BL21 StarTM(DE3)/pET24b(+):chi18H8 in high amount but accumulated into inclusion bodies as an inactive form (up to 55 mg/Lculture and 9.42 mg/gcells). By inclusion bodies treatment with 10 mM lactic acid, it was possible to recover Chi18H8 in active form (specific activity on 4-MU-(GlcNAc)2 up to 40.7 U/mg) and with a high solubilization yield (> 80 %). The electrophoretic analysis of the solubilized protein revealed that it was ca. 70 % pure: this purity level is appropriate for the majority of biochemical and functional studies of the protein, and also for future applications, for example as biocontrol agent. Partner 13 and Actygea successfully scaled up this production and recovery process in 3 L bench-bioreactor (Partner 13) and in a 30 L industrial bioreactor (Actygea subcontractor), demonstrating the possibility to produce high amount of the recombinant enzyme for future exploitations. Table 5 reports the results.

<table>
<thead>
<tr>
<th>Cell weight (gcells/Lculture)</th>
<th>Chi18H8 production (mg/gcells)</th>
<th>Solubilization yield (%)</th>
<th>Specific activity (U/mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flask</td>
<td>9.38</td>
<td>&gt; 80</td>
<td>40.7</td>
</tr>
<tr>
<td>3 L bench-bioreactor</td>
<td>5.3</td>
<td>91</td>
<td>65 52.1</td>
</tr>
<tr>
<td>30 L industrial bioreactor</td>
<td>6.2</td>
<td>12.9</td>
<td>65 48.4</td>
</tr>
</tbody>
</table>

Table 5. Comparison of Chi18H8 production and solubilization at flask, 3 L bench-bioreactor and 30 L industrial bioreactor levels.

Actygea repeated the 30 L fermentation three times and the results were reproducible. They also developed and scaled up a further purification method based ion exchange chromatography in order to obtain milligram amounts of pure chitinase for its structural characterization. The purification method proved to be effective in purifying the chitinase to apparent homogeneity also at pilot-scale level.

WP5:
Work-package 5 involved high-throughput sequencing of metagenomic community DNAs from selected habitats and sequence analyses with the aim to identify sequences encoding enzymes of interest featuring biotechnological value. In particular, development of a metagenomics analysis platform was a key task within this workpackage to allow management, handling, processing and interpretation of large metagenome datasets with respect to the Metaexplore objectives. Moreover, platform projects provided access to metagenome datasets for project partners and allowed collection of downstream results and storage of these results in a structured way. Thus, relevant information on genes and enzymes analysed in this project was documented and is accessible. An important focus
within this work-package is the plasmid metagenome, the collectivity of plasmids (mobilome) in a given habitat, as it was assumed that the frequency of hitherto unknown target biodegradative genes may be enriched in the mobilome as compared to that in the corresponding metagenome.

Mobilome sequencing and sequence analyses
Partners 1, 3, 8 and 15 worked on this task. Partner 1 (RUG-ME) has sequenced several novel plasmids, of which some were remotely similar to the highly-promiscuous PromA group of plasmids. Hence, novel plasmid backbone genes, such as rep, mob and tra regions, have become available. Partner 3 did a deep analysis of a plasmidome obtained directly from a wastewater treatment station. Partner 15 (UNLP) has built up a collection of plasmid replicons (several of them with more than 100 kb) recovered from a target BIOFILTER fed with pesticides. Plasmid purification was performed and plasmids were sequenced using Ion Torrent and the Illumina Technologies. As a result, a total of more than 20 Mb of non-redundant DNA sequence was obtained and automatically annotated via the GenDB pipeline available at CeBiTec (partner 8, UB) for in silico analysis.

Mobilome in silico analysis to search for chitinases
An in silico screening for chitinases was performed using the mobilome data set, together with a data set of another mobilome from the soil bacteria Sinorhizobium meliloti available at the laboratory of partner 15 (UNLP). Partner 15 identified 6 genes with significant sequence similarities to regions of already described chitinases, with amino acid identities between 58-98%. These sequences were sent to Partner 4 (SH) and Partner 5 (UW) for further analysis. After the bioinformatics analysis the in silico translated product S. meliloti_1914 was selected as a putative family-19 chitinase for further studies. It has a Glyco_hydro_19 domain and a PG_binding_1 domain, with overall sequence similarities to polypeptides present in a Desulfobacterium-like isolate (58%), Glaciecola (60%), Nitrosonomas (60%), and Ochrobactrum (42%). The N-terminal sequence of the coding sequence for S. meliloti_1914 was truncated (ca. 120 bp). The sequence was then completed by Partner 15 using inverse PCR (digestion of total DNA from the host bacteria, unimolecular self ligation to generate circular structures, PCR with specific primers, and sequencing). The sequence of the complete ORF was used for chemical gene synthesis (Genscripts, USA) and further cloning into the expression vector pET22b+ with optimized codon usage for expression in E. coli. Plasmid DNA, already available, is shared with Partner 5 (UW) for protein expression and biochemical analysis.

Mobilome in silico analysis to search for laccases
The in silico screening to search for new laccases was performed using the biofilter plasmidome-sequence data set. Partner 15 identified 10 genes with significant homology to regions of already described laccases. The search was performed using different tools, including hidden Markov models, in collaboration with Partner 14 (UL). Out of genes found, three of them were identical to E. coli genes encoding laccases. The remaining (seven) genes encoded for novel polypeptides that shared 66 to 96% amino acid sequence identities to reported laccases. A full-length gene with sequence similarity to a putative laccases from Ochrobactrum / Pseudomonas was selected for chemical gene synthesis with an optimized codon usage as performed with the chitinase described above. Plasmid DNA was shared with Partner 14 (UL) for protein expression and biochemical analysis.

Bioinformatics platform projects for the analysis of mobilome data (UB)
An MGX project was set up for mobilome sequence data from a pesticide contaminated biofilter environment (detailed description in report 3). MGX is an advanced Metagenomics analysis platform that in principle provides the same functionality as MetaSAMS developed in the frame of Metaexplore. MGX also was implemented by partner UB (8). Compared to MetaSAMS, MGX allows processing of larger datasets such as those generated on Illumina sequencing systems. Taxonomic and functional profiles were calculated within MGX. Obtained results are described by partner UNLP (see below). In a second step, assembled contigs from the biofilter sequence dataset were imported into the annotation platform GenDB for prediction of genes and functional annotation of genes. Results are available via the web-front-end of GenDB for project partners (in particular UNLP). Access for a user group has also been provided to the corresponding MGX project.

Global analysis of the mobilome of the pesticide containing biofilter using high-throughput sequencing (UNLP, UB)
A high molecular weight (HMW) plasmid DNA sample obtained from 35 plasmid-containing isolates previously recovered from the biofilter used for pesticide removal (Partner 15) was sequenced. Sequencing was performed using in a first approach the Ion Torrent technology, and later on the MiSeq Illumina technology. The obtained reads were assembled using Newbler 2.6. Assembled sequences were imported into the genome annotation system GenDB version 2.2 (Meyer et al., 2003). After the automatic annotation, the sequence information was refined manually. Protein identity values were predicted using BLAST. We obtained a total DNA sequence information of ca. 20 Mb. Plasmid enriched sequence data were filtered in order to remove putative chromosomal contamination using Pseudomonas putida, Pseudomonas fluorescens, Bordetella bronchiseptica, Ochrobactrum anthropi and Sinorhizobium meliloti.
chromosomes as reference genomes for filtering purposes, yielding ca. 8.6 Mb of plasmid-enriched non-redundant DNA sequence information which contained 11,056 ORFs with an average GC content of 54.2%.

A) In silico analysis of plasmid-encoded replication, mobilisation and stabilisation functions (UNLP, UB).

Plasmid replication functions were analyzed by comparison of amino acid sequences deduced from all the reads to the domains and reference proteins from the Pfam database. Analysis was carried out using the MGX platform. The protein families related to plasmid replication (rep), stabilisation/partition and mobilisation (mob) functions were analyzed. Out of 2,055,622 matches against the Pfam database 368,055 represented putative plasmid-related functions, which indicates that at least 18% of sequences represent plasmid-related genes. In comparison to other metagenomes, genes encoding plasmid replication, mobilisation and stabilisation and/or partition functions were highly overrepresented in our plasmid metagenome, with also a high abundance of transposable elements. The variety of rep Pfams identified in the dataset (13 families) provides insights into the diversity of the plasmids present in the biofilter isolates of which Rep_3 (Pfam01051) and RepA_C (Pfam04796) were the most abundant ones (see WP6, task 6.4). These Pfams were assigned to 48 different rep genes. In addition, 6 other putative rep genes were identified which do not strictly belong to any rep Pfams. Most of the identified rep sequences in the dataset correspond to genes present on plasmids, the known hosts of which comprise different species belonging to the alpha-, beta- and gamma-Proteobacteria, Firmicutes and Actinobacteria, among others. The results were in agreement with the 16S rRNA gene sequences previously sequenced for the plasmid-containing bacterial isolates (previous Report 2). Some of the plasmids that encode the identified Rep proteins also contain degradative genes, such as pCAR1, and resistance proteins. Percentages of identity of the Rep proteins identified in the dataset against the NCBI protein database were variable, including 7 proteins with an identity lower than 50%, which indicates that novel Rep proteins are present in the plasmid collection from the biofilter.

Genes associated to the two main mechanisms involved in plasmid stabilization/partition (i.e. post-segregational killing and active partitioning systems) are highly represented in the dataset (not described in detail here). Mobilisation (mob) and conjugative functions were the most abundant plasmidic functions comprising ca. 12% of the hits obtained for the deduced amino acid sequences present in the biofilter plasmidome. The quantity and diversity of the conjugal protein families identified by the Pfam analysis indicates the presence of both conjugative and mobilizable plasmids.

B) Presence of antibiotic resistance genes (ARGs) in the plasmid metagenome (UPLP)

To characterize the biofilter resistome, we searched the plasmid metagenome for signatures of known ARGs, using the Antibiotic Resistance Database (ARDB). The abundance and types of the 174 ARGs identified by Partner 15 were relatively lower than expected.

C) Presence of metal resistance genes in the plasmid metagenome (UNLP)

The collected DNA sequences of the plasmid metagenome also showed different types of heavy-metal resistance genes, including genes with similarities to those encoding resistance to zinc, cadmium, mercury, nickel, copper, tellurite, and cobalt.

D) Presence of genes and IS elements associated to degradation of xenobiotic compounds in the plasmid metagenome (UNLP)

In order to identify genes related to the degradation of pesticides and other xenobiotic compounds, a preliminary in silico search was carried out. Since the degradation of these compounds often comprises complex pathways, it was difficult to identify the complete pool of enzymes that are involved in corresponding pathways. Furthermore, many xenobiotic degradative enzymes show sequence similarity with enzymes from the general cell metabolism. Moreover, genes encoding enzymes related to the degradation of methyl viologen and gamma-hexachlorocyclohexane could be identified. The IS1071 insertion sequence has been found flanking genes involved in degradation of natural and man-made toxic compounds. For this reason, PCR identification of IS1071 has been used as indicator for detection of catabolic genes. In our collection of plasmid-containing isolates, seventy-two percent of the isolates showed positive amplification signals with amplicons having 94% nucleotide sequence identity to the IS1071 sequence. The flanking regions of this element are potential targets to identify new degradative genes.

E) Search for genes encoding specific enzymes (UNPL).

E.1) Analysis of laccase/MCO genes

The in silico screening was done using the biofilter plasmidome dataset towards the isolation of genes encoding for new/improved laccase/MCO activities. Partner 15 identified 10 genes with significant sequence similarity to regions of already described laccase genes. Out of these, three are completely identical to genes of E. coli laccases. The remaining seven genes feature more distant sequence similarity with the known laccases (66-96%) thus probably encoding novel enzymes. Also, fasta files containing 13,217 sequences (nucleotide or amino-acid) were sent to partner 14 (University of Ljubljana, Slovenia) for further evaluation. The aim was to refine the search for novel laccases using hidden Markov models (pHMMs) previously developed by partners UL and UB. With this
The Acidovorax MG01 project comprises 5 reference genomes, the Pseudomonas SBV1 project consists of 4 reference sequences. Unique features between related strains were determined by means of the comparative genomics tool EDGAR developed by partner UB. Comparative analyses, closely related reference genomes were also imported into the corresponding GenDB projects. Differences and Variovorax M7A. Sequence reads for these isolates were assembled and contigs were imported into the annotation platform GenDB for Partner RUG-2 (2) provided genomic sequences for three microbial isolates, namely Acidovorax MG01, Pseudomonas SBV1 and Paracoccus selected. Gene synthesis and cloning in the expression vector pET22b+ with the optimized codon usage for gene expression in E. coli as a His-Taged protein was performed at Genscripts (USA).

E.2) Analysis of chitinase genes

The in silico screening for chitinase genes was done exploring the same plasmid mobilome from the biofilter, and a mobilome dataset from the soil bacterium Sinorhizobium meliloti also available at Partner 15. Six genes with significant sequence similarities to regions of already described chitinases (amino acid identities between 58-98%) could be identified. These sequences were sent to Partner 4 (SH) and Partner 6 (UK) for further analysis. After that, the translated product of the ORF S. meliloti_1914 was chosen as a putative Family 19 chitinase for further studies. It appeared to have a Glyco_hydro_19 domain and a PG_binding_1 domain, with significant sequence similarity to chitinases from Desulfobacterium-like bacteria (58%), Glaciecola (60%), Nitrosomonas (60%) and Ochrobactrum (42%). The N-terminal sequence of the coding sequence for S. meliloti_1914 appeared to be truncated (ca. 120 bp), thus being then completed by Partner 15 using inverse PCR (digestion of total DNA from the host bacterium, unimolecular self ligation to generate circular structures, PCR with specific primers, and sequencing). The sequence of the complete ORG was used for a chemical gene synthesis (Genscripts, USA) and further cloning in the expression vector pET22b+ with optimized codon usage for expression in E. coli. Plasmid DNA, already available, will be shared with Partner 5 (UK) for protein expression and biochemical analysis.

High-throughput metagenomic sequencing of soils (amended with chitin; RUG, UW, UB) and sponge, and preliminary analysis. Several soil metagenomes were sequenced, including two chitin-treated ones, i.e. one in NL (including a non-chitin control, and with respect to phylogenetic and chiA GH family 18 genes; partner 1), one in UK (total metagenomic sequencing; partner 5). The NL soil revealed a great diversity of potentially chitin-active organisms coming up under the influence of chitin, in particular oxalobacteriaceae and actinobacteria. The chiA GH family 18 based sequencing effort revealed unanticipated huge diversity of sequence types, providing access to enormous chitinase sequence space. The soil metagenome of the o-chitin amended test soil microcosm (UK, partner 5) was sequenced in collaboration with partner UB (8) on the MiSeq platform. 8,862,563 sequence reads were obtained and analysed in the MGX metagenomics analysis platform for taxonomic and functional profiles. Moreover, specific searches for genes of interest were carried out. Assembled contigs from metagenome sequence reads were imported into the GenDB annotation platform for automatic gene prediction and functional annotation. Access to the MGX- and GenDB-project was provided for a defined user group. The obtained sequence dataset has a size of 29.6 Mb and 10,839 putative genes were predicted and automatically annotated whereas 8,777 further genes need manual attention.

Preliminary Pfam searches on the test soil metagenome yielded sequences or sequence parts with the following Pfam assignments (number of sequences in brackets): CBM_14, CBM_21, CBM_21, ChitinaseA_N (2), Cu-oxidase (6), Cu-oxidase_2 (6), Cu-oxidase_3 (6), Cu-oxidase_4, Dyp_perox (2), GSHPx (2), Glyco_hydro_18 (2), PG_binding_1 (10) and fn3 (5). Among these sequences is a gene showing 71% identity to a polysaccharide deacetylase from Rhodomicrobium vannielii ATCC 17100, and a sequence showing 51% identity to a sugar hydrolase from an uncultured bacterium BLR5. The sponge sequencing effort (triplicate A. aerophoba, next to Dysidea, sponges) is under analysis, but preliminary data already indicate that the bacterial richness amounts to hundreds of species, including numerous with putative chitin-active functions.

Bioinformatics analyses of metagenome sequence datasets for (chitin-treated) soil environments (RUG-ME, UB). Two sequence datasets from partner 1 (RUG-ME) representing fosmid pools obtained from chitin-treated soil environments were transferred to partner 8 (UB) and imported into the metagenomics analysis platform MGX for taxonomic and functional profiling. These datasets comprise 429,267,892 and 64,171,634 reads. Transferred raw data were quality controlled and filtered for fosmid vector sequences and E. coli chromosomal background. Datasets were searched for putative chitinase genes applying BLAST analyses against defined databases representing reference genes and/or proteins. Moreover, candidate genes were taxonomically assigned by applying different taxonomic classifier modules implemented in MGX (MetaCV, Kraken and the Lowest-Common-Ancestor approach – LCA). A number of candidate genes were identified within both datasets, both of family 18 and family 19 glycosyl hydrolases. Access to the corresponding MGX projects was provided via the MGX web-front-end for project partners.

Screening of microbial genome sequences for the presence of target genes (RUG2, UB). Partner RUG-2 (2) provided genomic sequences for three microbial isolates, namely Acidovorax MG01, Pseudomonas SBV1 and Variovorax M7A. Sequence reads for these isolates were assembled and contigs were imported into the annotation platform GenDB for automatic gene prediction and gene annotation. Partner RUG-2 screened the genome sequences of these strains for target genes. For comparative analyses, closely related reference genomes were also imported into the corresponding GenDB projects. Differences and unique features between related strains were determined by means of the comparative genomics tool EDGAR developed by partner UB (8). The Acidovorax MG01 project comprises 5 reference genomes, the Pseudomonas SBV1 project consists of 4 reference sequences.
and the Variovorax M7A project has 6 reference genomes representing related species. Further results of this genomic approach are presented by partner RUG-2 elsewhere in this report. GenDB projects for the described target organisms are available for project partners via the GenDB web-front-end (see also Table 5.1 of task 5.5).

Sequencing of fosmid clones encoding candidate genes (RUG-ME, UW, UB)

Single promising fosmids (selected by genetic screens for the presence of GH family 18 genes, from fosmid pools from chitin-treated soils were sequenced (outsourced), after which the sequences were filtered for quality and robustness, assembled and annotated using existing pipelines (partner 1). Five such fosmids were thus produced, see further below. Moreover, a suite of mixed fosmids (meta-fosmidome) delivered by partner 1 (RUG-ME) to partner 8 (UB) were sequenced and analysed as reported above. The fosmids identified to have chitin-active enzyme activity by partner 5 were sequenced by partner UB (8) as planned. These were fosmids D7 and G9, sequenced on the MiSeq system and annotated using the GenDB annotation software by partner UB (8).

The GenDB project for fosmid sequences has been made available for project partners (UW) via its web-front-end. The data on fosmids D7 and G9 are reported in the foregoing (WP4.1).

We are currently characterising G9_54 due to its predicted multidomain structure and possible bi-functionality. We are also cloning G9_52 family 18 chitinase. PCR products have been ligated into pET22b vector for expression trials.

Sequencing of fosmid clones assumed to encode laccase candidate genes (UL, VTT, UB)

Six candidate fosmid clones were provided by partners UL and VTT. It was assumed that these clones contain laccase genes. Fosmids were sequenced at UB and obtained sequences were assembled and imported into the annotation platform GenDB for automatic and manual annotation. The corresponding GenDB-project is available for project partners. Contigs representing the fosmid-inserts (42, 35, 34, 40, 39 and 38 kb in length) are accessible in GenDB. An obvious laccase gene was identified on one fosmid (81, fosmid-clone 19).

Detailed results concerning the identified laccase gene and other genes of interest on these fosmid clones are described elsewhere in this report by partners UL and VTT.

Implementation of the metagenomics analysis platforms MetaSAMS and MGX

Metagenomics in general aims at exploring microbial communities concerning their composition and functioning. Application of high-throughput sequencing technologies for the analysis of environmental DNA-preparations can generate large sets of metagenome sequence data which have to be analyzed by means of bioinformatics tools to unveil the taxonomic composition of the analysed community as well as the repertoire of genes and gene functions. A bioinformatics software platform is required that allows the automated taxonomic and functional analysis and interpretation of metagenome datasets without manual effort. To address current demands in metagenome data analyses, the novel platform MetaSAMS was developed. MetaSAMS automatically accomplishes the tasks necessary for analysing the composition and functional repertoire of a given microbial community from metagenome sequence data by implementing two software pipelines: (i) the first pipeline consists of three different classifiers for taxonomic profiling of metagenome sequences and (ii) the second functional pipeline accomplishes region predictions on assembled contigs and assigns functional information to predicted coding sequences. Moreover, MetaSAMS provides tools for statistical and comparative analyses based on the taxonomic and functional annotations. The capabilities of MetaSAMS were demonstrated for different metagenome datasets obtained within the Metaexplore consortium. The MetaSAMS web interface is available at https://metasams.cebitec.uni-bielefeld.de. In the second phase of the project the advanced metagenomics analysis platform MGX was developed. This system, in principle comprises the same functionality as compared to MetaSAMS. However, MGX allows efficient processing of larger metagenome datasets achieved by an improved data management concept. Moreover, the platform implements extensive dialog fields for the collection of meta-data and provides flexible analysis pipelines covering state-of-the-art methods for taxonomic and functional profiling of microbial communities. In addition, user-defined analysis pipelines can be composed and implemented in MGX. This new metagenomics platform has been applied to analyse metagenome datasets within Metaexplore.

GenDB (microbial genome annotation platform), MetaSAMS and MGX projects were set-up for Metaexplore partners. Analysis results can be explored and retrieved via web front-ends of the analysis platforms. Meta-data for metagenomes and information on target enzymes was and can be added in a structured way to the corresponding projects via dialog fields provided in GenDB, MetaSAMS and MGX and therefore the set of platform projects represents a data warehouse environment for Metaexplore project partners.

Sequence-based screening of metagenome datasets for the presence of genes of interest

Plasmid metagenomes (mobilome) from pesticide contaminated biofilters (Kortrijk biopurification system) were sequenced and analyzed within the metagenomics platforms MetaSAMS and MGX. Novel plasmid replicons were identified that may serve as vector plasmids in downstream applications. Accessory determinants such as antibiotic resistance and metal resistance genes as well as putative xenobiotic degradation genes are present within the biofilter plasmid metagenome. Putative laccase, dehalogenase and chitinase genes that are in the focus of the Metaexplore project were identified. Moreover, different IncP-1 plasmids harboring
degradation (catabolic) genes are present in the biopurification system analysed.

Development of Hidden-Markov-Models for the identification of laccase and chitinase genes
Hidden-Markov-Model (HMM) based search-routines for the identification of sequence reads and contigs representing a specific gene encoding an enzyme of interest were implemented in the analysis platforms. Using this feature, variants of target genes can be identified. This approach was applied using the example of laccases.
Laccases have been used in various fields ranging from processes in wood and paper industries to environmental applications. Although a few bacterial laccases have been characterized in recent years, prokaryotes have largely been neglected as a source of novel enzymes, in part due to the lack of knowledge about the diversity and distribution of laccases within Bacteria. Within the Metaexplore project, genes for laccase-like enzymes were searched for in over 2,200 complete and draft bacterial genomes and four metagenomic datasets, using custom profile Hidden Markov Models for two- and three-domain laccases. More than 1,200 putative genes for laccase-like enzymes were retrieved from chromosomes and plasmids of diverse bacteria. In 76% of the genes, signal peptides were predicted, indicating that these bacterial laccases may be exported from the cytoplasm. Moreover, several examples of putatively horizontally transferred bacterial laccase genes were described. Many metagenomic sequences encoding fragments of laccase-like enzymes could not be phylogenetically assigned, indicating considerable novelty. Laccase-like genes were also found in anaerobic bacteria, autotrophs and alkaliphiles, thus opening new hypotheses regarding their ecological functions. Bacteria identified as carrying laccase genes represent potential sources for future biotechnological applications.
The HMM-based approach for the identification of target genes was also developed for chitinase genes and genes encoding other enzymes of interest.

WP6:
Under WP6, the Metaexplore project developed and applied advanced molecular to allow the cloning and sequencing of the metagenomes of microbial communities in environmental habitats to find genes that can be used to make special enzymes that can be used for industry and bioproduction for degradation or converting of substances that are otherwise difficult to work with such as chitin and lignin. Several of the objectives of the tasks under WP6, i.e. WP6.2 WP6.3 and WP6.4 have been achieved to a major extent, whereas other tasks, i.e. WP6.1 and WP6.5 have suffered from technical difficulties that impede their completion / success.

Gene fish - Concerning WP6.1 unfortunately no cassette recombination was detected in the chromosome of the Genefish strain from the four independent recombineering experiments achieved. It is true that recombineering is a useful method for up to 1.5kb DNA exchange but experiments with longer sequences have never been reported in the literature yet. Actually the 1.5kb long internal positive control used in our experiments showed the expected recombination frequencies obtained after lambda Red protein induction (mean value compared to control plasmid transformation frequency was 9x 10-5 ± 3.6x 10-5).

Vectors beyond E. coli - Concerning WP6.2 to develop vectors that allow transfer beyond E. coli, partner 1 (RUG-ME) adapted the extremely promiscuous highly-transferable broad host range plasmid pHPIO2 (PromA group of plasmids) to allow insertion of exogenous DNA and selection upon transfer to diverse hosts. The mob (mobilization), transfer (tra) and stability/maintenance regions of the plasmid were cloned, together with a multiple cloning site and a selectable tetracycline resistance, to yield a new vector coined pIPT. Tra genes are to work in trans from another introduced plasmid, and hence a derived form lacking the tra region was designed and is now under construction. Moreover, the already broad host range of plasmid pHPIO2 was confirmed and even extended to include Sphingobacterium spp. and related bacteria. The vector will need to show proof-of-principle, i.e. it should be easily transferred among diverse bacteria when carrying exogenous DNA, before it will be ready for use. Moreover, a new system based on a broad host range BAC vector and an acidobacterial strain was established to express environmental traits which might not be expressed in E. coli (Partner 12). Thus, Edaphobacter sp. IGE012 from the phylum of Acidobacteria was tested for its ability to host metagenomic libraries. High molecular DNA extracted from a highly contaminated sediment with aromatic hydrocarbons was cloned in BAC SBO HPT22 vector (Lucigen USA) and transferred to Edaphobacter sp. IGE012. Randomly selected clones showed high stability of the recombinant BACs. BACs were efficiently extracted using alkaline lysis methods. Extracted BACs were of good quality and quantity for enzymatic restriction digestion analysis and next generation sequencing. Some clones from the library showed a yellow color proving the expression of a functional trait from the sediment related to the synthesis of secondary metabolites. The insert sequencing is in progress. On the other hand, Edaphobacter sp. IGE012 was used as host for the laccase gene of Bacillus subtilis CotA using a transposition cloning vector (pUTminiTn5::Gfp). Laccase related genes were not detected in the genome of IGE012. Then, a functional screening for laccase activity based on the ABTS test was conceivable to detect recombinants. Recombinant clones were expressing the GFP protein and CotA marker was detected by PCR. Surprisingly, when the laccase activity was induced in the growth medium, not only recombinant clones expressed the activity but also the wild type. Therefore, a measurement of both activities is needed in order to check out if the activity was increased comparing to the wild type or not.
With respect to WP6.3 we did achieve this aim early on in the project, and the Trichoderma reesii expression vector was already successfully used in gene cloning experiments, see elsewhere in this report.

Mobilome sequences for novel vectors - WP6.4. has also been achieved with respect to the collection of mobilome sequences that are available (several partners involved). Partner 3 focussed on the mobilome, i.e. plasmid pools from environmental habitats. To get genes from bacterial species found in water or soil to ones used in industrial processes, plasmids can be used. Those plasmids found in the water or soil species are useful because they must be able to function in their original hosts. To investigate suitable plasmids, the DNA or “mobilome” that is transferred from one species to another was sequenced and in it hundreds of new small plasmids were found. Small plasmids are good at moving between species but may not be very good at taking useful genes with them, so to find out if there were also large plasmids UCPH found that using a technique called electrophoresis allowed to find suitable large plasmids in the mobilomes and that these also could be transferred from species to species by Pfam analysis. To see if tools for extracting special enzymes from the environment are not totally dependent on finding pre-existing plasmids from water or soil, UCPH synthesized an entire plasmid (pX1.0) in vitro. The plasmid was tested to see if it could transfer genes from one species to another. The plasmid was able to transfer by conjugation, and this showed that it is possible to make such tools by de novo processes if needed. Using high throughput sequencing, it was possible to find out about genes that are transferred from species to species or strain to strain without having to grow the bacteria in a laboratory. Therefore UCPH used these techniques to see if this type of activity was taking place in and around farms in biofilters for water clean-up. From the sequenced samples from the biofilters, lots of plasmid related genes were found that also encoded stabilization or mobilization genes were found. Some of them were similar to the previously synthesized pX10 plasmid in a family known as IncW. Some of these plasmids were shown to be able to transfer from species to species. The plasmids that were able to do so showed the presence of both antibiotic and metal resistance genes. This indicated that the species present in agricultural processing biofilters can confer antibiotic resistance from strain to strain and has importance for our understanding of how bacteria become resistant to antibiotics and also how bacteria can become resistant to metal pollution in areas that have been involved in mining or other industrial activity. The initiative showed that not only is it possible to find new tools for finding special genes or enzymes of interest for industrial activities, but also that there is a great deal more information to be found by applying these new techniques to areas of interest, and that they can also help in understanding how bacteria may play a role in remediation of metal contaminated land or how antibiotic resistance can be transmitted in the environment.

Furthermore, the mobilome sequencing has yielded multiple novel (rep, par, other) plasmid backbone sequences from the different mobilomes. These were sufficiently different from known such genes to offer possibilities for future use, developing novel vectors that broaden the host range of metagenomics (labs of partners 1, 3, 9 and 15). The sequences are available in the Metaexplore data repositories. However, we have not yet explored such sequences for further use, as we placed an emphasis on the further development of the aforementioned pIPO2-based approach for non-E. coli vectors.

Tandem expression vectors, substituted for long PCR for enhanced detection of inserted genes - WP6.5 was abandoned in its original aim, but the work that came in place (long PCR) was highly beneficial to the whole project. The ‘long PCR’ tasks aimed to tease out accessory genes from known insertional hot spots in IncP-1beta plasmids. On the basis of the long PCR, a suite of novel plasmid-inserted genes and gene clusters have been isolated directly from the plasmidome of the pesticide-loaded biofilters that were analyzed (Partner KUL)

WP7
Work package 7 encompassed all management, organizational, meeting, dissemination and website activities, throughout the project duration. All partners had tasks in this WP, with enhanced weights for the co-ordinator and work package leaders.

For management tasks, the project coordinator (Prof. dr. J.D van Elsas, partner 1) carried out overall project coordination tasks and secured the overall project planning and budget. He also coordinated WP1, and was further assisted by several work package leaders and co-leaders, who carried out the scientific coordination of the remaining different WP’s. The coordinator was assisted by a part-time project manager (Dr. J.E. Oppentocht). The coordinator monitored the project scientific progress and ensured cross-fertilisation between research lines and work packages. Both the coordinator and the project manager kept contact with the consortium members (in particular the work package leaders). For financial and legal issues, the project coordinator and project manager worked closely together with the financial and the legal departments of the University of Groningen, as well as the relevant EU officers.

Specific coordination and management achievements during this reporting period were the following:

• Organisation and facilitation of project meetings:
  o A 2-day progress meeting on 13–14 May 2013 in Warwick, UK (organised by partner 6 (U Warwick)). 8th progress meeting.
  o A 2-day progress meeting on 7–8 April 2014 in Helsinki, Finland (organised by partner 12 (VTT)). 9th and final progress meeting.
Moreover, the putative halogenases found are attractive for molecular-chlorine-free incorporation of halides into organic molecules, contacts with industry on the use of such enzymes for biocontrol and waste degradation and for production of chiral amino acids. Concerning both applications, there are close contacts with industry on the use of such enzymes for biocontrol and waste degradation and for production of chiral amino acids.

Compounds of increasing pharmacological importance and industrial chemicals. Concerning both applications, there are close contacts with industry on the use of such enzymes for biocontrol and waste degradation and for production of chiral amino acids.

The Metaexplore project established a Europe-based international platform on environmental metagenomics research, providing knowledge and experience on the construction of metagenomic DNA libraries, activity and genetic screening of these for interesting gene functions and the development of gene function databases which can be exploited by both scientific and industrial stakeholders. Moreover, the project has focused increasingly on direct sequence based analyses. The project is relevant to the overall objective of the Theme 2 work programme, as it contributes to the construction of a European knowledge-based bioeconomy. In that context, the project brings together multiple European scientific partners (across a range of countries) and four industrial SMEs as stakeholders. The partners exploited and exploit current research opportunities, i.e. cultivation-free metagenomics and gene isolation from within microbial communities, that address environmental and economic challenges. Regarding the economic challenges, the project has yielded a large gene function database and already identified several novel biocatalyst functions, which can be valorized by several industrial sectors (e.g. chemical, food, feed, pharmaceutical, agricultural and environmental biotechnology). In addition, the new vectors and high-throughput technology will be applicable in related research on culturable organisms and in protein engineering with applications in the same sectors. The SMEs participating in the project are among these sectors and/or strongly associated to them. Regarding environmental challenges, in environmental biotechnology often expensive and environment-unfriendly techniques are used. Also, disposal of waste is expensive. Furthermore, water treatment is high on the agenda of every country. Particular attention is currently given to the occurrence of micropollutants, i.e. xenobiotic compounds that are often present in low concentrations but are still harmful to the ecosystem. Typical examples are pesticides derived from agricultural practices, compounds like PCBs and surfactants and endocrine-disrupting compounds. The European Water Framework Directive (WFD; COM 2000/60/EC) addresses the EC member states for their responsibility in the next years regarding the improvement of water quality, in particular in respect of persistent organic pollutants. The recovery of new gene functions in pollutant degradation from uncultured organisms can give us more insight in the natural presence of such functions. In addition, they can be used in treatment of wastewater, groundwater and soil. Finally, most - if not all - of the technologies and applications where biocatalysts are involved can be considered environment and energy-friendly.

Furthermore, a great challenge for our society is to meet the growing demand for energy and to supply raw materials for industry in a sustainable manner. Alternative raw materials are needed for the future supply of energy and chemicals. Recalcitrant waste materials like lignin and chitin contain various building blocks for chemical production. This proposal developed enzyme-based tools for valorisation of chitin and lignin. As such, the results from this project have a clear impact on societal needs, the need to treat and mitigate human activities impacting the quality of the environment, of natural resources and biodiversity of natural ecosystems. We thus addressed the growing demand for sustainable use and production of renewable bio-resources. The enzymes produced are in principle available for (1) environmental/agricultural purposes (chi18H8, 53D1, G9_54, several laccases), and (2) producing novel compounds of increasing pharmacological importance and industrial chemicals. Concerning both applications, there are close contacts with industry on the use of such enzymes for biocontrol and waste degradation and for production of chiral amino acids.

Moreover, the putative halogenases found are attractive for molecular-chlorine-free incorporation of halides into organic molecules,

Potential Impact:

Description of the final results and their potential impacts and use (including socio-economic impact and the wider societal implications of the project so far)

The Metaexplore project established a Europe-based international platform on environmental metagenomics research, providing knowledge and experience on the construction of metagenomic DNA libraries, activity and genetic screening of these for interesting gene functions and the development of gene function databases which can be exploited by both scientific and industrial stakeholders. Moreover, the project has focused increasingly on direct sequence based analyses. The project is relevant to the overall objective of the Theme 2 work programme, as it contributes to the construction of a European knowledge-based bioeconomy. In that context, the project brings together multiple European scientific partners (across a range of countries) and four industrial SMEs as stakeholders. The partners exploited and exploit current research opportunities, i.e. cultivation-free metagenomics and gene isolation from within microbial communities, that address environmental and economic challenges. Regarding the economic challenges, the project has yielded a large gene function database and already identified several novel biocatalyst functions, which can be valorized by several industrial sectors (e.g. chemical, food, feed, pharmaceutical, agricultural and environmental biotechnology). In addition, the new vectors and high-throughput technology will be applicable in related research on culturable organisms and in protein engineering with applications in the same sectors. The SMEs participating in the project are among these sectors and/or strongly associated to them. Regarding environmental challenges, in environmental biotechnology often expensive and environment-unfriendly techniques are used. Also, disposal of waste is expensive. Furthermore, water treatment is high on the agenda of every country. Particular attention is currently given to the occurrence of micropollutants, i.e. xenobiotic compounds that are often present in low concentrations but are still harmful to the ecosystem. Typical examples are pesticides derived from agricultural practices, compounds like PCBs and surfactants and endocrine-disrupting compounds. The European Water Framework Directive (WFD; COM 2000/60/EC) addresses the EC member states for their responsibility in the next years regarding the improvement of water quality, in particular in respect of persistent organic pollutants. The recovery of new gene functions in pollutant degradation from uncultured organisms can give us more insight in the natural presence of such functions. In addition, they can be used in treatment of wastewater, groundwater and soil. Finally, most - if not all - of the technologies and applications where biocatalysts are involved can be considered environment and energy-friendly.

Furthermore, a great challenge for our society is to meet the growing demand for energy and to supply raw materials for industry in a sustainable manner. Alternative raw materials are needed for the future supply of energy and chemicals. Recalcitrant waste materials like lignin and chitin contain various building blocks for chemical production. This proposal developed enzyme-based tools for valorisation of chitin and lignin. As such, the results from this project have a clear impact on societal needs, the need to treat and mitigate human activities impacting the quality of the environment, of natural resources and biodiversity of natural ecosystems. We thus addressed the growing demand for sustainable use and production of renewable bio-resources. The enzymes produced are in principle available for (1) environmental/agricultural purposes (chi18H8, 53D1, G9_54, several laccases), and (2) producing novel compounds of increasing pharmacological importance and industrial chemicals. Concerning both applications, there are close contacts with industry on the use of such enzymes for biocontrol and waste degradation and for production of chiral amino acids.

Moreover, the putative halogenases found are attractive for molecular-chlorine-free incorporation of halides into organic molecules.
which serve as pharmaceutical building blocks. The number of such molecules is rapidly growing, but industrial biocatalytic halogen incorporation is still not implemented. Also, dehalogenases can be explored for the kinetic resolution of optically active haloalkanes, halogenated esters and halogenated alcohols. Dehalogenases may also be used for nucleophilic ring opening of epoxides, and are regarded as an important group of upcoming biocatalysts.

List of Websites:

http://rug.nl/research/metaexplore/
j.d.van.elsas@rug.nl