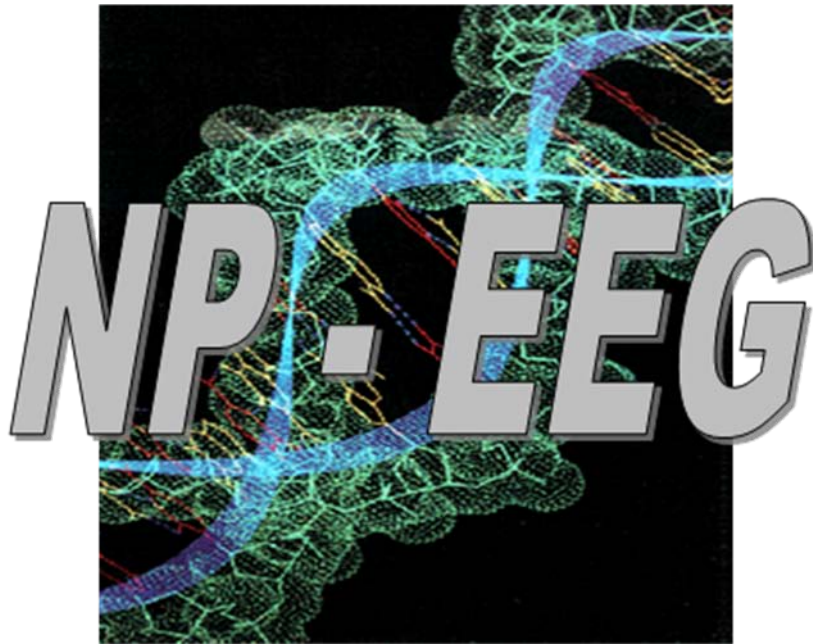


**National Program Ecological and Evolutionary Genomics**

**NP-EEG**

**Pre-proposal for NWO and NCI**

**NERO (Netherlands Ecogenomics Research Organisation)**



## Introduction

Ecogenomics is the emerging scientific discipline that studies the structure and function of a genome with the aim to understand the relationship between the organism and its biotic and abiotic environment (Van Straalen & Roelofs 2006). Ecogenomics involves several lines of research such as: unraveling patterns of genome evolution using (complete) DNA sequences; studying the impact of environmental factors on species at a functional genomics level by DNA microarray-based technology; investigating and discovering uncultivable microorganisms by metagenomics approaches; studying the role of (epi)genetic variation in stress responses and its impact on evolutionary change; and studying gene-environment-function relationships among different species using comparative genomics. The main drivers from a scientific point of view are to use new opportunities provided by the 'omics'-technology revolution and apply these tools to better understand gene-function relations, to learn more about species interactions and the mechanisms of evolutionary change, to discover novel micro-life species and functions, and to better monitor and predict the potential impact of environment, landscape and climate changes on populations and ecosystems.

## Ecogenomics in the Netherlands

The current state of Dutch ecogenomics research was assessed and discussed in two national workshops (Soeterbeeck 2006; Reehorst Ede, 2008). A comprehensive overview was laid down in the report "Ecogenomics in the Netherlands: exploration of opportunities and necessities", which was presented at the National Ecogenomics Day 2008 at the Reehorst Ede on 29 February 2008.

The report clearly demonstrated that:

1. Ecogenomics research is becoming a major part of ecological research at all Dutch Universities, and many Dutch ecological research institutes.
2. Ecogenomics in the Netherlands covers a wide range of ecological and evolutionary questions, focusing on plants, animals and microorganisms and their biotic and abiotic interactions, at organizational levels from the (sub)individual level all the way up to ecosystems.
3. Ecogenomics research in the Netherlands is characterized by an unique organizational model, where all Dutch groups are participating in a Netherlands Ecogenomics Research Organisation (NERO)
4. Many countries have already recognized the importance of ecogenomics for the future, as indicated by dedicated research programs set up by national and international organisations. In total, approximately xx Meuro has been allocated to such ecogenomic programs (NERC, Sweden, USA, ESF, etc.)

## Netherlands Ecogenomics Research Organization (NERO)

NERO provides the national platform for fundamental and applied ecogenomics. NERO aims at bringing together knowledge of various (sub)disciplines and at presenting a clear profile of (Dutch) ecogenomics. NERO was first presented at the National Ecogenomics Day 2008 and is the direct follow-up of the platform Ecological and Evolutionary Genomics, with participation of the Bsik-NGI-consortium Ecogenomics and the Genes4water initiative. NERO will be presented at [www.ecogenomics.eu](http://www.ecogenomics.eu).

NERO sees its **mission** as follows:

*NERO will provide a platform function for ecogenomics, will act as a co-ordinating organisation, facilitating communication between the research field, financing agencies and end-users, will facilitate knowledge transfer in the form of workshops, thematic presentation days, and advanced international courses.*

## **National Program Ecological and Evolutionary Genomics (NP-EEG)**

NERO has developed a **strategy** that brings together fundamental ecogenomics research, focussed on development of knowledge, and applied ecogenomics research, focussed on valorisation of this knowledge.

Core in this strategy is the development of two intertwined research programs. The first program, focussing primarily on further development of basic and fundamental scientific knowledge, is described in this document as the **National Program Ecological and Evolutionary Genomics (NP-EEG)**, see below).

Closely connected is the second program, **ECOLINC**, which primarily focuses on the societal and economic valorisation of existing and newly developed knowledge. This program has been submitted for financing from FES money, while being integrated in in the FES-cluster proposal "Bio-based sustainable industrial chemistry" for the period 2010-2015.

While both proposals have strong connections, they differ in the emphasis: **NP-EEG** places emphasis on extending fundamental ecogenomics knowledge, as a requirement for developing applications, while **ECOLINC** places emphasis on the development of ecogenomics applications for biotechnology, while exploiting existing fundamental knowledge.

The NP-EEG program, of which an outline is given in this policy document, is designed along three themes:

1. Metagenomics
2. Adaptation and Environmental Change
3. Comparative Genomics and Evolution

These themes represent international key issues in ecogenomics, and at the same time represent, emphasize and reinforce Dutch strongholds in ecogenomics, as became evident in the "Ecogenomics in the Netherlands" report.

Fundaments of the NP-EEG are:

- the focus will lie on a functional approach to the relationship between biotic and abiotic environment on the one hand and genes and gene functioning on the other
- Core theme is biological variation, both at the phenotypic and the genomic level, in natural environments and ecological relevant settings
- Point of departure is the study of processes, not the study of model organisms

## **THEME 1 Metagenomics**

Metagenomics is the study of genetic material recovered directly from environmental samples. The original definition by Handelsman (1999) referenced the idea that a collection of genes sequenced from the environment could be analyzed in way analogous to the study of a single genome. The exploding interest in environmental genetics has resulted in the broader use of the term metagenomics to describe any sequencing of genetic material from environmental (i.e. uncultured) samples, even to describe work that focuses on one organism or gene. Here we refer to the original definition in order to differentiate between traditional molecular environmental approaches and ecogenomics based approaches.

According to a recent report from the US National Research Council ("The new science of metagenomics: revealing the secrets of our microbial planet", 2007), the emerging field of metagenomics presents the greatest opportunity to revolutionize understanding of the microbial world. It sheds light on thousands of new microorganisms leading to new biological concepts as well as practical applications for human health, agriculture, and environmental stewardship. The main scientific merits of the metagenomics approaches

are i) to provide tools to study entire communities of microbes and how they interact and ii) to discover new products and functions in the largely unknown world of the unculturable microbes.

The first applications of metagenomics approaches were mainly directed towards describing the genomes of microbial communities. Target communities ranged from simple ones consisting of a few species, such as the acid mine drainage community (Tyson et al 2004) to specific phyla, such as Acidobacteria (Kielak et al 2008) and entire communities (Venter et al 2004, Rubin et al...).

Two basic experimental approaches are applied, the sequence based approach and the function based approach. In sequence-based metagenomics, the focus is on unraveling the entire *genetic sequence* found in a sample. Pyrosequencing facilities and similar modern sequencing methodologies will be tremendously instrumental to the realization of complete analyses of microbial communities of unprecedented size and complexity in the near future. Function-based metagenomics explores the products that microbes can produce. In function-based metagenomics, metagenomic libraries are screened for various functions, such as antibiotic production.

Whereas metagenomics is expanding our knowledge of the gene content, functional significance, and genetic variability in natural microbial communities, limited information exists concerning the regulation and dynamics of genes in the environment. Recently, first information was published on the analysis of gene expression in a naturally occurring microbial community ( Lopez et al 2008). RNA amplification technologies were adapted to produce large amounts of cDNA from small quantities of total microbial community RNA.

The main objectives of metagenomic studies within this program should be related to

- the structure and functioning of key microbial communities in samples of specific environments such as the rhizosphere, marine hot spots, active sludge samples, polluted sites.
- the interactions among microbes and of microbes with higher organisms in order to find out how these communities perform such functions as balancing the atmosphere's composition, fighting disease, and supporting plant growth.
- the search for new products ("Nature Mining"), in particular new antibiotics, vitamins and enzymes

The afore mentioned NSF report calls for a new Global Metagenomics Initiative to drive advances in the field in the same way that the Human Genome Project advanced the mapping of our genetic code. The proposed activities within this program fit well within such Initiative

## **THEME 2**

### **Adaptation and Environmental Change**

Climate change, and environmental change in general, affects the distribution and abundance of organisms. This is mainly due to the different sensitivities and responses of species and ecotypes upon various abiotic and biotic stresses (e.g. temperature, micro-organisms, herbivores, contaminants). The responses not only depend on the ambient environment but also on the amount of genetic variation present and the constraints among different fitness traits brought about by trade-offs. Under field conditions multiple stresses often occur simultaneously affecting a suit of fitness traits. The responses of organisms to these blends of stresses are difficult to predict from the responses to isolated stresses (Mittler, 2006) or from single trait responses. Furthermore, the existence of one stress can affect the performance upon a second stress. A good example is the lower herbivory or pathogen resistance of plants that respond to cues from neighboring competitors compared to plants grown under optimal conditions (Izaguirre et al, 2006; Genoud et al, 2007). Such interactions can occur at the level of signaling events (e.g. light-mediated regulation of defense genes) or at the level of resource partitioning (Cipollini, 2004). Stresses not only occur simultaneously but often they vary

strongly in duration and intensity. These variations select for different suites of adaptive traits that are set in motion by different signaling and transcriptional networks (Voeselek and Pierik, 2008).

In order to understand how fitness is affected in changing environments it is important to identify signal-transduction networks, its genetic components (genes) and the genetic and epigenetic processes that induce adaptive changes in morphology, anatomy and physiology (phenotype). The omics technologies will aid to elucidate the molecular and genetic basis for ecological responses to the environment. More particular we need to understand which suites of genes are involved in stress responses that are related to fitness. Recent work of Dinneny et al. (2008) demonstrated that such analyses must be performed with high levels of cell specificity and developmental resolution. Intact plant roots exposed to for example nutrient deficiency stress revealed only a fraction of the genes that were differentially regulated upon the stress than did cell specific analyses. Such exercises indicate a tremendous dilution of information when analyses are not performed with the required detail and precision (Voeselek and Pierik, 2008).

A modern day combination of ecology and genomics therefore calls for functional genomics approaches that transcend the levels of accuracy and resolution that have been adopted so far. In addition to this strong claim on experimental and technical detail, it is also crucial to study different environmental adversities, both separately but certainly also in ecologically relevant combinations.

Model species serve as a guide for identifying relevant gene networks and help us understand complex processes in non-models. Ultimately however, the knowledge of certain (model) species has to be scaled up to the population and community level in ecologically relevant settings. As a first step the generality of certain suites of traits and their genomic regulation related to specific environmental changes should be tested via a comparative genomics approach.

A number of relevant aims can be formulated to improve our understanding of adaptations to changing environmental conditions with the aid of omics technologies:

- (i) Investigate the genetic mechanism underlying life-history changes using both model and non-model species
- (ii) Identify relevant genes, proteins and metabolites, construct networks and identify epigenetic control mechanisms that are functionally important during single and multiple environmental changes with omics technologies.
- (iii) Scale up these results to the population and community level. In this respect the first aim should be to generalize gene expression patterns in various species tolerant or intolerant to certain environmental changes or combinations of stresses.

### **Theme 3**

#### **Comparative genomics and evolution**

Comparative genomics strives to improve our understanding of function by studying conserved and diverged features in genomic data from multiple organisms (Hardison 2003). Genetic and genomic changes form the basis of species radiations as well as the evolutionary origins of novel traits. Comparative genomics now allows us to study biological diversity with unprecedented detail, not only diversity in numbers (e.g. species richness), but also in functional traits. For instance, adaptation of species to different environments may involve co-opting existing functional pathways for other purposes, or using other cues to induce a change in their developmental program.

Comparative genomics is currently being used in three major research areas. First, modern comparative analyses are used to test phylogenetic hypotheses on the history of life on earth. In turn, genomics data can also be used to construct more robust phylogenies (Rokas, Williams et al. 2003). This new intersection of genomics and phylogenetics has been termed "phylogenomics" (Philippe, Delsuc et al. 2005). This field is also identified as a major theme for the research in the 'Netherlands Centre for

Biodiversity' (NCB), the planned merger of Naturalis, the Zoological Museum Amsterdam and the National Herbarium.

Second, comparative genome sequencing is being used to unravel the dynamics of entire genomes, including studies of genomic rearrangements (Richards, Liu et al. 2005), segmental duplications (Bailey and Eichler 2006), polyploidisation and gene loss (Dujon, Sherman et al. 2004). A surprising finding has been that copy number variants (CNVs), of even very recent origin, can have profound phenotypic consequences (Estivill and Armengol 2007). The alignment of multiple genomes can also be used to reconstruct an ancestral genome (Rocchi, Archidiacono et al. 2006). Last but not least comparative genomics will be essential for annotating the rapidly increasing number of genomic and EST sequences of non-model species.

Third, comparative genomics data have been used to identify genes and regulatory elements under selection. This allows us to identify the traits that matter in evolution: genes and traits that are under diversifying selection, and genes and functions that are conserved over long evolutionary timeframes. Examples for the use of comparative genomics in studying the evolution of (novel) traits vary from the origin of anaerobic glucose fermentation in yeast (Ihmels, Bergmann et al. 2005) to senescence in animals and humans (Passtoors, Beekman et. al 2008) and the genes that are unique to human brain development (Sikela 2006). Having multiple genomes of varying phylogenetic depths has also proven particularly useful for detecting conserved non-coding sequences (Kellis, Patterson et al. 2003; Hughes, Cheng et al. 2005)

These developments together hold great promises to improve our fundamental understanding of how genomes evolve in response to internal as well as ecological drivers, and how this affects the biodiversity on earth. In addition, the relevance for society is very high, for instance in the detection of genes, traits and substances of great economic importance (e.g., for health, crop improvement, livestock breeding), or for developing governance strategies and protection of biodiversity.

There are several key-issues relevant to the future of comparative genomics, issues that are directly related to the four research areas listed above:

1. Taxa should be selected and studied that cover a phylogenetic range appropriate for specific adaptive differences or traits of economic importance (e.g. relatives of crop species, or pathogens, or key taxa for ecosystem functioning).
2. The Netherlands needs to invest in facilities and expertise needed for investigating changes in genome organization and their impact on the creation of adaptive variation and diversity.
3. Knowing that we cannot (always) rely on gene presence or expression in fully sequenced model species, efficient tools need to be developed for functional screens of large numbers of organisms and in real ecological settings. Current evidence suggests that emphasis should be given to the role of regulatory variation for the functioning in the wild.

## **Ecogenomics for society and economy**

The "omics" technologies and approaches (including proteomics and metabolomics), that form the core of NP-EEG, have and will continue to revolutionize the science of ecology. These technologies and approaches allow to study the genetic basis of traits that are crucial for adaptation and evolution, at an unprecedented level of detail. They allow mining nature, the in-depth exploration of the diversity of processes and functions that lies hidden in nature. For full and optimal exploitation of this potential of ecogenomics, a solid and detailed scientific basis is of the utmost importance. The NP-EEG, which focuses

primarily on the development, of basic and fundamental ecogenomic knowledge, will therefore make an essential contribution to exploiting the great potential of ecogenomics for society and economy.

The relevance of ecogenomics for societal issues has a large number of important aspects. For instance, ecogenomics is very relevant for maintaining sufficient high quality water supply, for sustainable land development and planning and for sustaining biodiversity, for designing criteria for the sustainability of bio-based applications in industrial biotechnology, for the development of efficient tools for the monitoring of ecosystem and human health, and for the design of efficient and early implemented mitigation measures of the effects of climate change on human safety and ecosystem service functions.

Consequently the valorization potential of ecogenomics is high, and has implications for many industrial and societal stakeholders. Results of ecogenomics research will feed into various biotechnology applications, including the bio-production of fuels, chemicals, vitamins and antibiotics, (i.e. white biotechnology) and the optimisation of existing and design of new fermentation processes for water purification and the cleaning of contaminated sediments (i.e. blue biotechnology). Ecogenomics will be essential for the development of improved sensors/bioassays that can be used as monitoring tools for investigating the impact of contaminants on environment and human health, and for monitoring the health and quality of soil and water.

Nature mining and metagenomics will unleash the unknown microbial diversity, that drives the planet's energy dynamics, the nutrient cycles, and that maintains the health of the inhabitants. Nature mining will result in the discovery of new bioactive compounds, biocatalysts and bio-construction enhancers, and will enhance the exploitation of nature's capacity for bioremediation, heat and carbon storage capacity, and disease suppression.

Comparative genomics will help in extrapolating the results from model lab organisms, like mice, zebrafish, Drosophila, and C. elegans, to higher organisms including humans. It will thereby provide the scientific basis for using model organisms to predict the effect of toxic compounds and newly developed drugs in human individuals. In addition, as invertebrate models are not considered to be laboratory test animals according to EU legislation, comparative genomics will make an important contribution to the reduction of the use of laboratory animals.

-----

## Proposal:

In the light of the previous it is clear that a fundamental research program, designed along the three described themes, is essential for further development of the field of ecogenomics, and for laying a fundamental basis under any attempt to apply this knowledge for societal and industrial valorisation.

Therefore **NERO** proposes the following:

1. support of the further development of fundamental ecogenomics research in the Netherlands by establishing a **National Program – Ecological and Evolutionary Genomics (NP-EEG)**, financed in a joint effort by NWO-ALW and NGI
2. development of a **technology support model** for the NP-EEG, taking into account the high cost-level of ecogenomics research, and the need to create “focus and mass” in research, taking into account its strong multidisciplinary nature
3. development of a **financing model** for the NP-EEG, that is primarily and foremost based on scientific merits, but in which the potential for direct valorisation is also incorporated

**We ask NWO and NGI to jointly invite NERO to work out this initiative to a full-scale program**, refining the thematic priorities, financing conditions and management structure.

The Netherlands is a respected player in the international ecogenomics community, with many individual Dutch researchers being recognised as belonging to the international top. On the other hand, Dutch research groups might be less attractive for international consortia due to low visibility, and lack of structural financing. Especially in ecogenomics, the Netherlands could and should profit from its current leading role by investing in this field, thereby securing its international position.

The start-up of a National Program Ecological and Evolutionary Genomics would be a key contribution to facilitate achieving this goal.

Januari 2009

Joop Ouborg (RUN)  
Peter van Tienderen (UVA)  
Rens Voesenek (RUU)  
Hans van Veen (NIOO)