Strategic paper “From Genes to Organisms”

Version taskforce December 2019
Strategic paper NWO research community
“From Genes to Organisms”

Table of contents
1. Scope of the life science research community “From Genes to Organisms” ................................................. 2
2. Current scientific strengths/themes/challenges and (inter-)national position..................................................... 2
3. Focus areas for the next ten years .................................................................................................................... 4
4. Impact on society and contribution to topsectors (and related missions), national science agenda-routes and SDG’s .................................................................................................................. 5
5. Infrastructure in national & international perspective ..................................................................................... 6
6. Involvement of Dutch organisations in your research community ................................................................. 7
7. Specific challenges for the community........................................................................................................... 7
1. Scope of the life science research community “From Genes to Organisms”

- DNA and beyond: From understanding gene function and genome evolution to genetic engineering.
- “System Organism”: Elucidating how physiological interactions between multiple layers of regulation control development and functioning of tissues, organs and organisms.
- Decoding the interactions between genetic and environmental factors in determining the functioning, development, and evolution of organisms.

2. Current scientific strengths/themes/challenges and (inter-)national position

Themes and strengths: Genes encode the primary genetic information of any organism. Understanding the genomic landscape and how it determines the development and physiology of organisms is at the heart of any Life Science question. The community covers the broad range of all kingdoms of life, ranging from microorganisms to plants and animals, including humans. These are studied from genes to whole organisms. A major strength, amongst many, is that all disciplines show a wonderful integration of basic science at knowledge institutes (predominantly universities), translational research (often as collaborations between universities, institutes [KNAW, UMCs] and companies) and applied studies. As a consequence, research in the Netherlands within this domain has managed to be at the international forefront of basic science, and of great relevance in societal applications. Examples of leading contributions from the Dutch research community within this domain are highlighted in green below.

Genetics is the study of how heritable traits, encoded in DNA, get passed on to the next generation. Genetic mutation and recombination is crucial to evolution and therefore affects all domains of biology, from the molecular level of how proteins behave to the interactions between organisms within large ecosystems. In the previous decade, the various “omics” approaches, including high-throughput sequencing (HTS), and novel tools for genetic modification have revolutionized genetics and genomics research.

Rapid development of HTS machines has brought the cost of for example a human genome sequence down from $2 billion to less than $1000. This has created a number of opportunities: the creation of reference genomes for thousands of species across the tree of life, population sequencing to identify rare genetic variants and routine genome sequencing for diagnostics purposes for instance in oncology and microbiology. HTS is an enabling technology that expands the analysis of DNA and RNA to unprecedented detail and magnitude. The Netherlands is at the forefront of genomics research, in part due to an early investment through the Netherlands Genomics Initiative (NGI). These developments have provided a detailed description of the genetic make-up and existing genetic variation of organisms. Genome-wide Association Studies (GWAS) have helped identify causal genes and markers for particular phenotypes, such as organism traits or disease. With the pace of genome sequencing, large scale phenotyping and bioinformatics tool development, it is expected that we can elucidate the function of thousands of genes in the functioning of model species, including humans, in the near future. Concomitantly, this knowledge is translated to non-model species, to improve crops, livestock and microorganisms or perform diagnostics in patients.

Despite the importance of understanding the function of a particular (causal) gene we are still far from understanding the actual process and how a trait or organismal function is established. In other words, even though we are starting to unravel the beginning and endpoint of a process, we miss the connecting physiological mechanisms, which will be important for modifying or improving a trait and developing novel therapies, specifically in complex genetic traits that involve multiple loci that quantitatively affect traits, such as blood pressure or plant yield.

Essential for dissecting the molecular mechanisms that affect phenotypes is gene modification. Before the development of genome editing tools, genetic modification was largely limited to a number of model organisms and was laborious in certain complex organisms. The development of CRISPR-Cas9, helped by influential work in the Netherlands (e.g. Brouns et al., Science, 2008) has greatly simplified genetic modification of nearly each cell type in every organism. This has opened up possibilities for rapid loss of function analysis of endogenous genes, pooled genetic screens and endogenous knock-in of fluorescent tags for measuring organism-wide protein expression levels. This is enabling research directions in all kingdoms of life, as well their societal applications such as human disease, crop improvement or biofabrication.
The (epi)genetic code, upon interpretation, instructs development of the whole organism. Although the organism has to function as a whole, multicellular organisms such as plants and animals are built from tightly integrated modules: the heart follows its own developmental trajectory, different from the nervous system and both are indispensable for functioning of the organism as a whole. Or, leaves follow a different developmental trajectory than do the roots and although both can develop rather independently, precise co-adjustment and cooperation is necessary for whole-plant functioning. Understanding coordination between organs/modules is key to understanding an organism. And even though in microorganisms this complexity does not exist, interactions between individual cells (organisms), such as quorum sensing, metabolite exchange, signal exchange between cells and biofilms are of great importance to understanding microorganism function in any given condition.

Although the genotype instructs the organismal phenotype, what gene functions and gene regulatory networks connect the two is intensively studied. During development, cells acquire a specific identity. Subsequently, cells need to integrate their genetic information with signals from their environment to establish specialized functions. Cells thus integrate genetic information with external signals, which can provide information from other body parts or even from the external environment. But even inside the organism there are microenvironments for specific local developments, and these can sometimes be exploited or created by for example tumours or pathogens. Understanding the details of and interactions with these microenvironments has greatly helped understanding stem cells and their niches, and allowed the development of organoids (Examples: Kretzschmark & Clevers, 2016, Dev. Cell). Similarly, studies of interactions with the microenvironment helps understanding tissue and organ patterning in plants (Example: Mähönen et al., 2014, Nature), tumour development in animals (Example: e.g.de Visser et al., 2006, Nature Reviews cancer), or to study microbiome communities and their recruitment for plant and animal health (Examples: Carrion et al., 2019, Science).

Micro-environmental differences may explain how different cells in an organism develop in distinct ways and perform different biochemical functions. Nevertheless, even in rather similar microenvironments, there can be substantial differences in cell types performing different functions, and receiving different instructions from the universal organismal genetic codes. Studies are ongoing to identify how these variations can exist and how they function at the organ and organism scale (Examples: Pantazopoulou et al., 2017, PNAS, another on animals).

Major breakthroughs have been made using established model organisms in life sciences, such as Caenorhabditis elegans, zebrafish, Drosophila melanogaster and mice for animal biology, Arabidopsis thaliana in plant science and human cells and tissues in organoid, disease and neuro biology. Some of these include the use of natural genetic variation and quantitative genetics to identify genetic loci for regulation of development. Others employ breakthroughs in genomics to analyze RNA expression and DNA sequences at the single cell level enabling an unprecedented resolution of tissue heterogeneity. Through the course of evolution tremendous variation in species has emerged, possessing different traits and trait combinations to sometimes thrive in very harsh environments. Research on such extremophiles has allowed the elucidation of novel biochemical pathways (examples: Strous et al., 1999, Nature), which helps understand evolution and natural selection at a mechanistic level, and is instrumental in biotechnological applications.

**Challenges:** A major challenge within this Life Science domain is the adequate management and usage of the vast amount of data generated already and anticipated in the years to come, and bioinformatics research and application is gaining importance by the day. This has been true for any “omics” research approach, but is rapidly expanding beyond this terrain. For example, high-throughput phenotyping in plant and crop science is rapidly expanding, also in the Netherlands through the large research infrastructure investment in NPEC, and will generate massive datasets that should be available for re-use. Likewise, imaging centres and genetic and “omics” screens create vast amounts of data, and curating these up to a level that allows, for example, meta-analysis presents a substantial challenge. Although genomics and transcriptomics global data repositories exist and are intensively used, new infrastructures will have to become available for other domains as just mentioned. In order to understand gene function in controlling physiology in various biological systems, rigorous molecular-genetic-physiology studies are needed in model organisms. Such urgently needed studies require substantial funding of basic science in model organisms. Other challenges include the development of methods for live imaging and real-time analysis in vivo.
Applications:

- Crop improvement for enhanced stress resilience, yield increase, sustainable weed and pest suppression.
- Human disease treatment and resistance
- Bioproduction of compounds / Sustainability
- Healthy aging
- Therapeutic application of genomic, epigenomic and genome-editing tools, with start-up and biotechnology companies.

3. Focus areas for the next ten years

To further understand biological processes and how genotype controls phenotype we need to focus on i) basic processes that underlie these connections and ii) the organism’s physiology that translates genotype to phenotype. This will involve basic molecular and cellular processes including cell-cell communication, gene-gene and gene-molecule interactions, protein translation, and physiological processes including cell growth, metabolism and signalling (inter-and intra-cellular). Furthermore, signalling of and responses to the environment are important for the functioning of an organism and we need to study these to understand development of organisms. These studies are time consuming, require a fundamental approach and are predominantly predestined for model species. However, a comparative approach in which model species are being compared with relatives is also very informative and gives insights into conservation and deviations in these fundamental processes and how species have evolved.

Moving forward we envisage the cost of sequencing to drop even further, this means that the amount of genomics data available to researchers will increase even more. Quantitative genome-wide measurements for chromatin modifications, transcriptome, proteome and metabolome, combined with detailed experimental analyses of the organism’s physiology will enable an integrated view on how information is processed from DNA to whole organism function (from genotype to phenotype). A current limitation is the number of skilled computational biologists that can tackle this question, a challenge that will only become greater as more data and more data types will be generated. Basic bioinformatics skills will be crucial for future researchers to explore large-scale datasets. To further make use of these data, sophisticated quantitative models will have to be developed, likely in cooperation with physicists and mathematicians. Another limitation is that physiology research and training has lost terrain over “omics” approaches. Right now, a modern integrative physiology has to emerge, combining tools from molecular-genetics, omics and “classic” physiology to truly understand the genotype-to-phenotype map.

When we think of genetics, we think of genes and when we think of genes, we think of protein coding sequences. Continued efforts are needed to assign function to a large part of the proteome, which remains a formidable challenge. However, it is important to note that in most complex organisms coding sequences make up only a minority of the genome sequence. The majority of the genome is non-coding, but is crucial in the proper spatiotemporal regulation of genes. It is also where we find the majority of genetic variation between individuals and (closely related) species. With the rise of inter-individual genome information it will become crucial to also understand the role of the non-coding parts of the genome.

Understanding, the effects of coding and non-coding genetic information will be greatly facilitated by the rise of CRISPR-Cas9 gene editing. Creating resistance marker-free genetically modified crops can greatly speed up breeding processes. The therapeutic potential of correcting disease-causing mutations represents an additional important future direction. Key to these developments is a proper ethical assessment and engagement of the public in these manners.

Throughout the Life Science sub-disciplines there is an increasing appreciation of the fact that there is strong heterogeneity between cells and cell types. Understanding how these differences are coordinated at the organism scale, and how this coordination determines whole-organism fitness is a challenging focus for the coming years, that requires increased focus on molecular physiology research. It is clear that such research requires usage and further development of the most advanced, internationally established model organisms available: Caenorhabditis elegans, Arabidopsis thaliana, Saccharomyces cerevisiae, Drosophila melanogaster, Danio rerio, Mus musculus. At the same time, not all life can be understood from these model systems only and it is key for a viable Life Science community to develop knowledge, tools and resources in other organisms, such
as humans for medical applications, crops for food production, insects as disease and pollen vectors, and evolutionary outliers for trait and pathway discovery.

Although organism development is obviously genetically encoded, it is now also appreciated that the environment has a major impact on this developmental trajectory. An important example of future focus is the interaction between organisms and their microbiome: both in medical and food crop research the microbiome is now considered a major driver of organism development and fitness, however very little is understood about the physiological mechanisms and functionalities of these interactions and of the composition, growth and development of this microbial community. Also, the development and productivity of all of our global food crops is very strongly determined by the abiotic environment they grow in, an environment that is very rapidly changing in the current era; it is a major challenge for the coming years to decode the exact interactions between environmental signalling and (crop) plant performance to feed our planet. In oncology the interaction between tumour cells and the immune microenvironment plays important oncogenic or tumour-suppressive roles, depending on the context.

4. Impact on society and contribution to topsectors (and related missions), national science agenda-routes and SDG’s

Genes encode the primary genetic information of any organism. Understanding the genomic landscape and how it determines the development and physiology of organisms is at the heart of any Life Science question. And while curiosity-driven, basic science has been at the cornerstone of this successful and broadly relevant theme, applications can be found throughout society. Research within this discipline in the Netherlands has contributed significantly to understanding disease, developing medicines, guiding and developing crop improvement or developing production of biomaterials and biochemicals. Application has been facilitated by a long tradition of collaboration between scientists from knowledge institutes and private and societal partners (the Triple Helix). Accordingly, research within the Genes to Organism connects to a broad variety of topsectors and contributes to several of the UN’s SDGs, as well as many of the National Science Agenda (NWA) routes.

Connection with NWA routes:

- Origin of life – on earth and in the universe.
- Circular economy and resource efficiency: sustainable circular impact.
- Sustainable production of safe and healthy food.
- Blue route: water as a way to innovative and sustainable growth.
- Health care research, sickness prevention and treatment.
- Personalised medicine: the individual at the centre.
- Regenerative medicine: game changer on the way to broad application.
- NeuroLabNL: the ultimate living lab for brain, cognition and behavioural research.
- Measuring and detecting: anything, anytime, anywhere.
- Creating value through responsible access to and use of big data.
- Sustainable development goals for inclusive global development.

Connection with Sustainable Development Goals:

- Zero hunger.
- Good Health and Well-being.
- Responsible consumption and production.
Connection with societal themes from the mission-driven innovation policy:

- Agriculture, water and food.
  - Circular agriculture.
  - Climate neutral agriculture and food production.
  - Valued, healthy and safe food.
- Health and healthcare.
  - Lifestyle and living environment (personalized healthcare).
  - People with chronic diseases (regenerative and personalized healthcare, AI, Big Data etc.).
  - Dementia (understanding the cause and developing cures).

Connection to topsectors:

- Agri-Food: Duurzame plantaardige-, dierlijke- en aquatische productie; Gezond en veilig.
- Horticulture and starting materials.
- Biobased economy.
- Life Science and Health.
- Water (Aquaculture and recovery of resources from waste water).
- ICT and HTSM innovations required to support advances in biological sciences.

Connection to Key technologies:

- Advanced materials (Biomaterials and soft materials), (Bio)nanotechnology, (Bio)procestechnology, Life science technologies (Industrial biotechnology (white), Gene editing/precise genetic engineering, Organ on a chip, Nanomedicine, Stem cell technology, Synthetic cell technology, Biochips and biosensors, Biocatalysis, X-omics (Gen/prote/metabol/glyc) and Biofabrication

5. Infrastructure in national & international perspective

Universities and the associated medical centres in the Netherlands that operate within the domain from genes to organism typically have local access to standard molecular biology and biochemistry labs, local imaging centres for microscopy (confocal, superresolution, electron) and genomics cores for DNA/RNA high-throughput sequencing. In addition, there are a number of national facilities that transcend local facilities, such as the Netherlands Proteomics Center (Utrecht) and several facilities funded from the Roadmap Large Scale Scientific Infrastructure (NWO). These include the Netherlands Plant Eco-Phenotyping Centre (NPEC), High Containment Research Facility Biosafety Level 3 (BSL3), ELIXIR-NL for data infrastructure in the Life Sciences, Mouse Clinic for Cancer and Ageing (MCCA), Netherlands Electron Microscopy Infrastructure (NEMI), Netherlands Bioimaging Advanced Microscopy (NL-BioImaging AM), NL-Openscreen (drug discovery), Unlock, X-omics. These facilities are world-leading, sometimes unique and sometimes comparable to facilities abroad, but always key to sustain Dutch Life Science at the international forefront.
6. Involvement of Dutch organisations in your research community

- All Dutch Universities.
- NWO and KNAW institutes: NIOO, NIN, Hubrecht Institute, Westerdijk, AMOLF, NIOZ.
- Medical Centres: AMC, UMCU, Erasmus MC, RadboudUMC, UMCG, LUMC, Maastricht UMC, VUmc, NKI.
- TO2: Wageningen Research, TNO (Healthy Life).
- Others (e.g. industry, non-profit organizations): RIVM, Crop improvement companies under the umbrella of Plantum (companies include, but not limited to, Rijk Zwaan, Enza zaden, Royal van Zanten, Deliflor, Bejo zaden, Nunhems-BASF, de Ruiter-Bayer). Animal breeding companies, including CRV, Hendrix genetics. Biotechnology companies, for instance DSM, DOW Chemical and many others. Companies in the fields of Pharma and Medical technology.

7. Specific challenges for the community

- Curiosity-driven, basic science research is the cornerstone of this research community, and is under pressure of tendencies in current funding schemes to request applications, utilization, valorization.
- Current European legislation on GMO applications, and its recent calling on CRISPR/Cas9 technology, is a major obstruction for research and knowledge utilization. For example, improved crops achieved through such technologies cannot be implemented, even if they are more sustainable and better yielding than conventional counterparts.
- Physiology is the research domain that translates genetically encoded information into the organism's phenotype, performance and behaviour. Physiology research has become underrepresented in the current high-throughput “omics” Life Science research landscape. Nevertheless, a modern molecular-physiology research domain is pivotal to achieving major goals in biotechnology, crop improvement, human disease and biofabrication.
- Like in any other field within the Life Sciences, the gender ratio at group leader positions, especially full professors, is still skewed towards men. This may send a highly undesirable, negative signal to upcoming scientists.
- Data infrastructures are lagging behind current technological developments.