

3. Inventory of Scientific Views; The role of bioinformatics

3.1. Introduction

Bioinformatics involves the development and application of computational resources in biological research. The discipline started to take shape in the early eighties in response to a need for computational tools with which the growing amount of bio-molecular sequence data could be better managed and analyzed. Among the early achievements of what would nowadays be called 'bioinformatics' were, for instance, the establishment of the first molecular sequence database in 1982, the EMBL Nucleotide Sequence Data Library, and the development of tools for the alignment and database comparison of DNA and protein sequences, such as the Smith-Waterman algorithm (Smith and Waterman, 1981) and the FASTA package (Pearson, WR, Lipman, 1988), respectively. In the past ten years, bioinformatics has undergone a spectacular growth and has evolved, from a somewhat arcane intersection of molecular biology, computer science, mathematics and statistics, into a mature scientific discipline. An illustration of this growth is given by the number of PubMed abstracts containing the keyword bioinformatics (Figure 1). Without bioinformatics, the practice of modern, large-scale biology has become unthinkable.

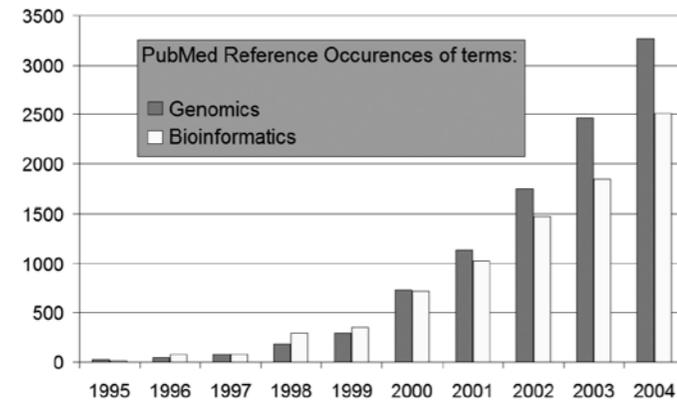


Figure 1. Growth in the number of PubMed records in the past decade, containing the keywords "bioinformatics" or "genomics".

Two technological developments have been instrumental for the advancement of the field. Firstly, the advent of high-throughput technologies in the late nineties changed the face of experimental biological research. Within a short time-span, beginning with the introduction of capillary DNA sequencers and DNA microarray technology, laboratories all over the world acquired the means to rapidly generate large amounts of bio-molecular and genetic data. The management, processing and analysis of these datasets soon called for novel, dedicated and user-friendly software, enabling bench-top biologists to fully exploit the results of their experiments. Secondly, the early internet, that had hitherto functioned as an information and communication network mainly operated by expert users and computer scientists, saw the introduction of its first, user-friendly interface in the early nineties. Since then, the internet has grown explosively and unleashed its enormous potential as an information resource and application environment to the life sciences community. With current hardware technologies, including high-speed internet connections, the immense amount of data and information stored in databases worldwide is at everyone's fingertips and its use has become every-day practice in biological research.

To date, a diverse array of powerful, high-throughput technologies has made its way into almost every area of molecular and genetic research. It ushered in a shift in research methodology, from hypothesis-driven, mostly gene-centric investigation to data-driven, exploratory investigation of biological systems. In the latter, the network of interactions among cellular components has become the focus of attention. In practical terms, this meant that, in contrast to a mere ten years ago, when genes, transcripts, proteins, and metabolites were analyzed basically on a one-by-one basis, an experimentalist today typically will study a large or complete complement of these biomolecules in response to environmental or genetic alterations to the system under investigation. The terms genomics, transcriptomics, proteomics, and metabolomics are commonly used to refer to these system-wide explorations of molecular components.

Bioinformatics facilitates many stages of this new way of experimentation, from the initial experimental design to the final steps of analysis and integration of results with information available from public databases via the internet. In the previous section we have reviewed modern molecular technologies used in research for breeding, crop improvement and development of sustainable product chains. Among these were several of the aforementioned high-throughput technologies. In this section we will give an overview of the bioinformatics resources that are being used in concert with such technologies.

3.2. Bioinformatics resources for plant breeding research

In many areas of biology, we want to come as efficiently as possible from the phenotype to the genotype. In breeding and breeding research the challenge lies in translating key-agricultural traits of crops (relating to yield, quality, resistance to disease and sustainability) in molecular

terms and so making these amenable to genomic analysis and faster breeding.

There are several areas of development within bioinformatics that are of major relevance to current plant breeding research:

1. Genome annotation
2. Statistical genetics
3. Comparative genomics
4. Metabolite identification and metabolic pathway elucidation
5. Biomolecular data management, integration, and mining

We will discuss these hereafter with reference to the technologies discussed in the previous section.

3.2.1. Genome annotation

Once the genome sequence of an organism has been determined, using the strategies outlined in 1.A.2., a process of analysis starts that aims to extract biological meaning from the raw DNA sequence. Results from these analyses are commonly stored in a relational database that provides useful connections between the bits of information inside the database as well as outlinks to information in other databases around the world. Most genome databases are made accessible to the research community through the internet and use rich interfaces and genome browsers that enable interactive, graphical exploration of the data. This entire process of extraction, storage, and representation of information is referred to as genome annotation.

The initial steps of annotation are “structural annotations” and start with the identification or prediction of physical attributes of the sequence and anchors these on to it. The attributes include protein- and RNA encoding genes, regulatory signals, ESTs, repetitive elements and segmental duplications, physical and genetical markers, SNPs, etc. The key activity in structural annotation is gene-finding. In eukaryotes, the coding content of the genome is usually very low (<10%) and identifying the genes and correctly predicting their intron-exon structure is not a trivial task. A wide array of gene prediction algorithms and tools has been devised in the past ten years. Many of these work by analyzing motifs and statistical properties of putative genes and often require a considerably sized set (>200) of reliable gene structures as models to train the software in preparation of the prediction step. Such models can be built by alignment of EST and cDNA datasets to the genomic sequence. This is also one of the many reasons why the availability of large expressed sequence datasets for a species is so important.

In a second phase, functional information is added to the primary sequence attributes (functional annotation), most importantly to the predicted genes. Although experimentation using biochemical assays or reverse genetics still is the golden standard for establishing gene function, computational prediction of function, the holy grail of bioinformatics, is gaining power and accuracy. This is largely due to the modern practice of massively storing results from both laboratory research and in silico analyses in databases and making these accessible to the wider

research community.

The evidence coming from this body of genomics and functional genomics information can be mined with automated computational procedures for gene function prediction, by exploring the diverse kinds of associations among genes in these datasets. There are three levels at which such associations can be found. At an operational level, a gene may be associated in terms of its expression with other genes of known function or with specific biochemical, cellular, physiological or developmental processes. At a structural level, a gene can be linked with other genes through genomic neighborhood or through sharing of similar regulatory signals, while the gene product may be associated with protein structures of known function, with protein complexes of known function, or with particular subcellular localizations. At an evolutionary level, a gene is related in terms of its nucleotide sequence to other genes within its own genome or within the genome of other species. Evolutionary associations form the basis of most classical function assignments in computational gene annotation. This method is commonly known as homology-based function prediction. It relies on the assumption that homologous proteins in different species perform similar functions and that experimentally gained knowledge obtained in one species can therefore be transferred to its homologue in another species.

When available, each of the above types of association will contribute to the global probability that a particular gene is involved in a particular biological process and can thus be assigned a particular function.

Computationally, homology-based function prediction starts with subjecting a gene from a newly sequenced genome to a database homology search with a program such as BLAST (Altschul et al., 1990). One can choose from a large collection of databases available for this purpose, among which the manually curated SWISS-PROT database stands out for the reliability of its annotations. BLAST will return a ranked list of genes from other species held in the database to which the query sequence is most similar. Annotations from these “database hits” can then, in a fully automated way, be parsed from the BLAST output, attached as a functional annotation to the query gene, and be stored in the annotation database of the new genome.

Although annotation transfer by homology is powerful, it is also fraught with several difficulties. Obviously, the gene may be short of clear homologues in the sequence databases or it may be homologous to other genes of unknown function. In addition, homologues not necessarily need to perform the same function, in particular when gene duplication is involved in their evolutionary origin. Such genes, known as paralogues, therefore have to be distinguished from true orthologues, i.e. homologous genes derived from a single ancestral gene through speciation only. Accurate orthology prediction is difficult, particularly when one is dealing with genes from large gene families, and may often require a considerable computational effort.

In addition to BLAST, more sensitive analyses and sequence comparisons can be performed to assign putative functions, in particular when plain sequence divergence obscures homology relationships. One approach is to compare the protein sequence of a newly sequenced gene against the various databases of protein functional domains and motifs, including PFAM,

PRINTS, PROSITE, ProDom, SMART, BLOCKS. An extremely useful addition to these databases has been the recent development of InterPro (Apweiler et al., 2001), a system that unifies all of these overlapping resources and allows them to be searched in a single run, returning a single integrated output.

Further functional information that can be attached to predicted genes include alignments of available EST and cDNA sequences, gene expression data, protein structure, interaction and localization data, classification data, e.g. using the gene-ontology scheme, and cross-references to databases that hold further information on homologous sequences such as metabolic and regulatory pathway databases, structural databases, etc. Another level of information with which genome annotations can be enriched involves structural similarities to other genomes. Such similarities are found through gene order comparisons and whole genome alignments and will be discussed hereafter in the context of comparative genomics methodology.

Genome annotation is a complex and multi-step process in which many databases are used and a broad range of computational tools is employed in both parallel and sequential fashion. To this end, the in- and outputs from different analysis have to be interconnected, jobs have to be computationally managed in a robust way and results have to be integrated for representation. Not only is the automation of this entire process important in order to match the increasing pace with which the raw sequence data are being produced. It is also critical for the organization and visualization of the extracted information with which we are to build our understanding of the primary sequence. These requirements have made the operation, accuracy, and flexibility of automated annotation systems a prime challenge in the field of genomics and, with respect to genomes of agricultural crops and their model species, one of utmost importance to breeding research. The relevance of having high-quality genome annotations of crop and model species available and accessible in an easy and seamless way becomes apparent, for example, when through breeding research a QTL or candidate gene region has been localized on the genome. Subsequently, the researcher will want to explore the genome annotation to find out what genes or regulatory elements map to this specific region. Examining the annotated gene functions may then provide clues as to which genes are most likely to be involved in the trait under investigation.

3.2.2. Statistical genetics

Plant breeding relies on the selection of favorable genetic variation that corresponds to phenotypic variation of agricultural traits of interest. The goal of statistical genetics is to localize regions in the genome that influence these traits and to dissect their underlying molecular and environmental components. The field traditionally has had close ties to forward genetics technology, outlined in the previous section, and to the practice of marker-assisted breeding and crop improvement.

Statistical genetics develops methodology and computational tools for genetical linkage map-

ping and the analysis of quantitative trait loci. The identification of single-gene traits with qualitative variation through the construction of high-density genetic maps has been extremely successful, both in plant and animal species. In contrast, unraveling the molecular basis of more complex, quantitative traits by QTL analysis has proved to be far more difficult. Yet, the availability of complete genome sequences and their rich and accurate annotation has opened up faster routes for the identification of candidate genes in QTL. An example of the dissection of a major fruit weight QTL in tomato (Alpert KB, Tanksley SD, 1996) may illustrate the application and value of bioinformatical analysis of QTL and candidate gene regions. Using high-resolution physical and genetic mapping, the localization of the fruit weight QTL could be narrowed down to a region of less than 150 kb. The QTL was deduced to be a single gene with a large effect (30% of the variation) and could be further isolated by positional cloning. Had the complete genome sequence of tomato been available at that time, however, the investigators could have moved on directly to an analysis of the annotation of the respective region in an attempt to identify or delineate the number of candidate genes or regulatory sequences and their variants responsible for the trait.

One step further is that genome annotation would include an analysis aimed at predicting what sequence variation in genes, proteins and regulatory elements might have functional effect. It will be extremely difficult to achieve acceptable sensitivity and specificity with such computational predictions in the near future. However, the ongoing accumulation of genome sequence data and the insights we are gaining from their comparative analysis in terms of conservation of functional and regulatory sites, can be expected to improve the results of such predictions of functional effect.

From a rather different angle, another important proposal to bring statistical genetics and genomics together has recently been put forward. Jansen and Nap (2001) suggested both fields to be merged in an approach of sampling variation in gene, protein or metabolite expression and molecular markers in parallel in a segregating population. They dubbed the field "genetical genomics". The innovative idea was to treat expression as a quantitative phenotypic trait, which, in conjunction with molecular marker analysis, makes these data amenable to QTL analysis. This provides an entirely novel route to the dissection and identification of factors that influence the expression of a gene. Genetical genomics is starting to fulfill its promises, witness a recent flow of publications that apply the approach to gene expression studies (see Editorial in Nature Genetics, March 2005).

3.2.3. Comparative genomics

The same evolutionary principle that underlies the method of homology-based gene function prediction outlined above, underlies the field of comparative genomics as a whole. Genomes are compared to each other in light of their common ancestry and the inferences and predictions made from such analyses are rooted in evolutionary reasoning. The basic premise is that

conserved segments between different genomes are constrained by selection and are therefore most likely to be functional. Without selection, random mutations will wear away similarities over time. Discovering all the functional parts and the details of their structure in a newly sequenced genome can thus be greatly enhanced through comparative genomics.

Powerful alignment algorithms have been developed to perform cross-species comparisons of genomic sequences, including AVID, BLASTZ, MUMmer, MLAGAN, WABA, and many others (Ureta-Vidal et al., 2003), and comparative genomics repositories are being established that hold precomputed alignments (e.g. www.tigr.org) which can be explored by researchers for regions of their specific interest. The global alignments between two or more species can be used directly to delineate conserved regions and provide important information on: 1) the extent of colinearity or synteny among the genomes, that is, regions of the genome in which the gene order is similar; 2) colinearity of physical and genetic maps between closely related species; 3) the location of genes, which are more conserved sequence regions than non-functional ones; 4) the detailed intron-exon structure of genes, and; 5) the location of conserved non-coding sequences, mostly regulatory elements. Alignment strategies that focus on the latter type of information are also known as phylogenetic footprinting and phylogenetic shadowing. In phylogenetic footprinting, regulatory elements are identified as stringently conserved sequences by comparing species over large evolutionary distances. In phylogenetic shadowing, known relationships among more closely related species are used and in the context of a model of sequence divergence for the same purpose. Again, these methods are extremely valuable for exploring candidate gene regions and QTL to narrow down the segments with the highest probability of being involved in expression of the trait under investigation.

Comparative genomics further provides a very direct way to identify species-specific genes and other genetic elements among closely related species that differentiate them phenotypically. A gripping example is the comparison of human and chimpanzee genomes. Their genomes are extremely close, showing over 98% similarity at the nucleotide sequence level, yet at the same time they represent two species that show a number of profound phenotypic differences. Comparison of their gene complements and regulatory mechanisms of gene expression will thus enable faster delineation of candidate genes and regions involved in these differences.

In the context of breeding research and crop improvement, comparative genomics provides the conceptual framework and the necessary tools for the transfer of genomic information from model species to related crop species.

3.2.4. Metabolites and metabolic pathway elucidation

Plants contain a bewildering array of metabolites. Many of these compounds stem from secondary pathways that uniquely evolved in plants. They include constituents essential for a healthy human diet, human dietary supplements with health-promoting properties, important nutrients for farm animals, compounds that determine food flavor and taste, and, not in the

least, a large number of compounds used either directly or indirectly as pharmaceuticals. Knowledge of the biochemical pathways that underlie the synthesis of these metabolites by plants is essential for their engineering and optimization in existing crops and possible transfer to novel crops.

Metabolomics is a relatively recent development in plant functional genomics as well as within the 'omics' family of high-throughput technologies. The aim of the field is the discovery and identification of the complement of metabolites present in plants and the elucidation of the pathway structure and dynamics that lead to their synthesis. Bioinformatics plays a key role in the development of metabolomics. At present, the mass spectrometry technologies that drive the field evolve faster than the computational means with which the data can be processed, interpreted and integrated with other types of molecular data. The most significant limitation is our current inability to automatically translate mass spectral data into molecular data on compounds and thus biochemically identify the metabolites. For some technologies, commercially-available reference databases can be used to this end, but for others, in particular those commonly employed for the analysis of plant secondary metabolites, no such standard reference systems are available. The latter identification strategies are therefore based on manually combining accurate mass information with additional, complementary information from other sources. This creates a serious bottleneck that hampers the further development of metabolomics as a tool to comprehensively study pathways, not only qualitatively but also quantitatively. Novel bioinformatics strategies that allow for faster and more accurate metabolite identification are therefore urgently needed.

Breeding research is expected to benefit in particular from improved metabolite and pathway identification in combination with the genetical genomics approach of quantitative genetic analysis of metabolite expression levels.

3.2.5. Biomolecular data management, integration, and mining

Critical for opening up genomic information for breeding research is the integration of the heterogeneous molecular, genetical and phenotypic data repositories. It enables researchers to direct their questions simultaneously at multiple levels of organization and derive answers that reveal associations among them. There are many factors that hamper this system-level type of data integration. Apart from sociological and commercial reasons, which may force laboratories not to (fully) disclose their data, most factors are of a technical nature. To overcome these, not only computational solutions are needed, but also development and broad acceptance in the research community of standardization of data models and data exchange formats.

Different approaches are being followed to solve the data integration problem, the most successful among which include 1) link integration, in which data sources are linked to each other through hypertext links; 2) data warehousing, in which basically all data relevant for the pur-

pose of a specific biological database are modeled coherently and brought together under a single roof, and; 3) web services (Stein, 2003). In the last approach, similar to link integration, data are left residing in their original repository, with their specific domains of biological information remaining intact, but a special software layer is developed for standardized access and data exchange among databases. This allows the user to automatically retrieve all relevant information on a biological question using just a single query. Furthermore, the technology allows not only the integration of datasets but also of bioinformatics analysis tools. The emerging web service technologies now appear very promising in overcoming important hurdles in data integration and are currently being developed in the BioMoby and MyGrid projects.

A different aspect to the data integration problem is a semantic one and involves the need to use common, standardized vocabularies to describe information within a certain biological knowledge domain. Such vocabularies are called ontologies, which, more specifically, provide a formal system for describing objects, concepts, and other entities in a knowledge domain, as well as the hierarchical relationships that exist among them. A number of international initiatives have been taken in the past five years to develop biological ontologies. The most successful among these is the Gene Ontology, which describes the gene product attributes of function and cellular location. Particularly relevant for the integration of plant genomic and phenotypic information is the Plant Ontology initiative, which has started to develop ontologies describing anatomy, morphology and development for a number of plant model species. This ontology will greatly improve our means to compare phenotypes among species and associate these with the various types of genomic information.

References

- Alpert KB, Tanksley SD. (1996) *Proc Natl Acad Sci U S A*. 93:15503-15507.
Altschul, SF, et al. (1990) *J Mol Biol* 215: 403–410.
Apweiler, R. et al. (2001) *Nucleic Acids Res* 29:37–40.
Pearson, WR, Lipman, DJ. (1988) *Proc Natl Acad Sci U S A*. 85:2444-2448.
Smith, TF, Waterman, MS. (1981) *J Mol Biol* 147:195-197.
Stein, LD. (2003) *Nat Rev Genet* 4:337-345.
Ureta-Vidal, A, et al. (2003) *Nat Rev Genet* 4:251-262.

4. Inventory of Scientific Views; Plant genomics and consumers in a sustainable food chain

4.1. Introduction

All the activities in the area of plant genomics research may lead to the development of interesting future products and the potential implications are likely to be wide and may affect many consumers. Various consumer issues may occur in the areas of People, Planet, and Profit which are important in the context of an appropriate societal embedding of genomics as a scientific field. Somehow these consumer issues need to be described, addressed and managed appropriately to realize the potential of plant genomics. This is rather difficult because consumers are still unaware of plant genomics, and also because at present, no plant genomics enabled or enhanced products can be found in the market.

4.2. Consumer issues in the food chain

Improved disease resistance qualifies best for application of genomics, also abiotic stress and health promoting characters are appealing applications. Plant genomics could focus on the 'difficult' traits, but most important is that products appeal to markets and consumers. Markers have been the first product of potato genomics, in addition to increased knowledge. Such markers may be used for diagnostics. However, the real product of plant genomics will be a new cultivar. That may take another 4 to 15 years. In the process of cultivar making, genomics (markers) will facilitate the selection process, but not necessarily speed it up. This production process requires the coordinated activities of the various branches of the food chain in order to provide convenient, wholesome, tasty, safe and affordable foods. Also, the size and complexity of the food sector ensures that no single player can control a single process from seed production, through farming and processing to a final product marketed in a retail outlet. Furthermore, the scientific advances in genome research and their exploitation via biotechnology is leading to a technology driven revolution that will have advantages for the consumer and food industry alike.

Currently the most general issue is that society will not accept agricultural genomics because it will consider 'genomics' synonymous with 'genetic engineering'. Given public fears with respect to genetic engineering, this would imply a rejection of genomics. Apart from the benefits missed, an issue to consider as well is that such a rejection would also deny experts to do the type of work they know and like best. The 'genomics equals genetic engineering scenario' is far from hypothetical. Both genomics and genetic engineering deal with genes, genetic material and improvements of plants. The distinction is a subtle one at most, which is easily forgotten.

The term 'genomics' may turn out to be a poor one.

Agricultural genomics may already have image problems in relation to genetic engineering. Worldwide, most prominent proponents of agricultural genomics were previously proponents of genetic engineering. This is also the case in the Netherlands. Moreover, the promises of agricultural genomics are not very different from the earlier promises of plant genetic engineering, nor are the research topics.

Another issue may be certain aspects of (over?)regulation. The level of regulatory scrutiny currently imposed on a genetically engineered crop is high and unprecedented for any product of plant breeding. Development and costs of such regulatory requirements may have significantly negative impacts on agricultural genomics.

In the development of agricultural biotechnology, private funds have slowly taken over public funds in size and impact. This development is continuing in the development of agricultural genomics. Agricultural genomics is a costly enterprise. High throughput technologies may make a single data point relatively cheap; total costs are considerable, due to the sheer numbers involved. Private investments far outweigh public funds, supported by mergers and continuous scale-up of life science companies. This implies that a few companies may decide over genomics research targets and applications. The corporate control of biotechnology in general, and future agricultural genomics in particular, are likely to generate considerable social concern. Despite all the promises, agenda setting in genomics seems still focused on short-term goals. These relate to conventional, high-yield industrial agriculture aimed at profit. Publicly funded programs such as TransForum may help to rebalance the control and agenda setting power of the private companies.

Consumer issues about corporate control are immediately related to issues of ownership and intellectual property (IP). Different from plant breeding, ownership in agricultural genomics is based on patenting and patent protection. Genomics research is seen as economic investment that requires return. The patenting frenzy of agricultural biotechnology will continue in plant genomics. This will raise questions about the equity, accessibility and desirability of plant genomics and its applications. Also in this case a program such as TransForum may enhance the possibility for the public to gain access to genomics information. When plant genomics is seen as a high-input, high-cost, high-protected enterprise, a related legitimate issue is about the benefits of such genomics for the developing world. The developing world will face most serious problems with food supply. This issue is known as 'the genomics divide'. In the current settings of genomics, it is unclear whether these problems will truly affect the genomics research agenda. Not everybody in society is likely to consider plant genomics into the quality of potato for baking french fries a priority target for research. Plant genomics could be seen as again a technological solution for problems that may have at least as many negative impacts as it presents solutions. It can be perceived to contribute to a further industrialization, economization and mechanization of agricultural production that is seen as undesired. In probably small parts of society, this will be related to difficult concepts as 'the integrity of life'. Particular

schools in plant breeding are still debating the acceptability of markers and similar discussions can be expected on the large-scale approaches of agricultural genomics. Moreover, any impact of plant genomics on a further industrialization of food production may trigger uncertainties about the safety of genomics products.

Overall, there seems place and need for substantial research into the mechanisms of agenda setting in plant genomics research, the extent of corporate control and the diversity of society's evaluation of such issues. Policy makers and genomics researchers seem aware of the need to include society and the need to deserve a 'license to produce' and 'a license to sell', rather than to exclude society and go on. How this awareness works out in day-to-day management decisions is yet unclear and requires more analyses and possibly new approaches.

4.3. Sustainability

Literally, sustainability means 'the ability to sustain'. This implies either to maintain something or keep it going for a period of time, or something that is supported by giving help, encouragement or support. Over the last decades the best-known, widely adopted, and now almost 'traditional', definition of sustainable development was given in the Brundtland report (1989). Sustainable development is here defined as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs". However, there are many different interpretations, modifications and reformulation of what is sustainability. A World-bank inventory listed no less than 190 different attempts. Sustainability has become linked to a large variety of human activities, concepts and issues, such as biodiversity preservation, and corporate social responsibility. Meanwhile, sustainability has become a priority for the world's policy makers. The concept is now at the core of laws in many countries around the world and has become a societal norm. Generally, people will do what they believe is 'right', i.e., whatever is consistent with their personal beliefs and/or the norms of the society they live in.

In the view of many, somewhere along the course of time, mainstream agriculture has lost its roots of sustainability. The development of increasingly intensive and specialized forms of agriculture has culminated in what some refer to as 'industrial agriculture' or 'chemical farming systems'. The farm is viewed as a factory with inputs and outputs, aiming at increased yields with reduced costs, often by exploiting the economy of scale. It is accompanied by use of and reliance on agrochemicals, large-scale mono-cropping, mechanization and energy consumption. Such agriculture sees as its primary objective to produce as much food and fiber as possible for the least cost. According to Lyson (2001), current conventional agriculture is anchored to the scientific paradigm of reductionist experimental biology, in combination with the reductionism of neoclassical economics, driven by a continued (desire for) industrialization.

Although successful in the past, virtually every aspect of this conventional agriculture has become problematic. The problems center on both the environmental aspects and on the social

and community aspects of agriculture. There are problems with the supply and quality of water, erosion, soils that become unfertile, contamination by pesticides, use of manure, and various levels of pollution of groundwater, soils and atmosphere. This makes agriculture a major cause of damage to the earth's ecosystem at large.

Combining sustainability and plant genomics is the combination of a difficult, multidimensional concept with a complex technology. To do so, plant genomics researchers have to translate their research activities into sustainability. Such a translation will mainly focus on the applications or applied aspects of genomics research. Having shown above that sustainability in general is seen as very positive in current society, genomics could try to become associated with this norm. Such an association may be used to translate scientific approval in societal approval, or to counterbalance societal concerns.

Sustainability in its most broad connotation has to do with (natural) resources (see above). Genomics can for example help replacing petroleum with biomass derived mostly from plants. Such a biomass is or could be a renewable natural resource, hence more sustainable. Actually, many of the goals of agricultural genomics as summarized above seem to reflect the need and desire for increased sustainability in the sense of renewable resource use and a significantly reduced environmental burden of agrochemicals. In this context, the issue is whether sustainability is used to legitimize agricultural genomics, or, vice versa, whether this genomics research and results will effect or change the norm sustainability of society. Moreover, if so, how that may happen. This relationship between legitimization, research and sustainability in relation to genomics is an issue that merits more analysis.

Depending on its research agenda and achievements, agricultural genomics may change its perspective with sustainability and ecological agriculture in society. For the future of genomics, its relationship with sustainability should therefore become better defined and clearer in society. Initially, this could be attempted on a case-by-case basis and may be used for generalizations in the future. Unless the public comes to agree that the benefits of agricultural genomics are equivalent to a sustainable, desirable and acceptable agricultural practice, society is likely to fail to recognize and realize much of the potential benefits of agricultural genomics.

4.4. Conclusion

Plant genomics, especially when presented as a 'new' technology, will have to face all the suspicions any new technology is usually confronted with. Plant genomics can be considered a next step in the process of agricultural biotechnology that started with the discovery of the Laws of Genetics by Mendel and the double helix structure of DNA by Watson and Crick. Plant genomics will have to face all the ethical, moral, social and technical issues associated with agricultural biotechnology in general. These are issues about corporate control, ownership, distribution of profit and benefits, overall approach as well as safety.

Also, the distinguishing characteristics between plant genomics, plant breeding and plant bio-

technology could be clarified better. A truly hot potato is the relationship between genomics and genetic engineering, in terms of both scientific activities and image of genomics. It would seem a valuable exercise to analyze in detail which societal groups are using what terminology of 'genomics' in what context and with what agenda. Whereas the stakeholders tend to stress the differences between genomics as a research activity and genetic engineering as a potential (but not automatic) application of that research, the public at large seems already in the process of confusing the two. Policy makers often attribute this reaction of the consumers to a lack of knowledge and to ethical and emotional concerns. (e.g. Marris et al., 2002). However, these rational policies are likely to enhance the ethical and emotional concerns of consumer and leave little room for the every day experiences of consumers. Communication and information about risk did not take into account the actual issues of consumers, but instead the knowledge of experts about risk assessments was communicated. Hilgartner (1990) has described this as the "deficit model". Particularly the government and groups of experts assumed that the public is in some way deficient in their understanding of risk. Also, not all risks are evenly important. According to Miles and Frewer (2001) people perceive risks as higher when they experience a lack of control over their exposure to potential hazards. So, trust in risk assessment and risk management is likely to be a particularly important determinant of public confidence in food safety in cases where there is a lack of trust. Miles and Frewer (2001) consider genetically modified foods, where consumer concern did not focus primarily on risk per se, but rather on the lack of personal control on the part of the consumer over consumption, as an example.

Plant genomics research and applications may effect or change the societal norm sustainability, or sustainability may be used to legitimize plant genomics. Understanding the relationships between legitimization, sustainability and plant genomics merits more analyses. From the perspective of sustainability, the approaches and assumptions of genomics can be seen as undesired and irreconcilable with its aims and approaches. This can be translated in the different perspective of holism versus reductionism. Whereas sustainability is seen as a holistic approach, plant genomics would be seen as the culmination of reductionism. More analysis is required to decide whether ecological agriculture and plant genomics are representatives of two essentially incompatible approaches, or can be combined. Can plant genomics fit in the approaches of ecological agriculture and if so, how? Vice versa, can any sustainable agricultural production scheme afford to deny the potential benefits of plant genomics? This creates a potential dilemma of sustainability that will be very interesting to analyze in a case-by-case approach. For the future of genomics, its applications and their relationship with sustainability should become better defined and clearer in society. Further analysis of the societal link between ecological agriculture and plant genomics may result in predictive capacities for the way plant genomics will be approached by society, or will have to be presented to society.

Plant genomics could be perceived as either a new way to change agriculture into a more natural, eco-friendly, low-tech production system or a next step in the further industrialization,

economization and mechanization of agricultural production, and may change the agro-food industry into a high-tech/high-value business. Which of these two options is a desire from and in the interest of society requires research attention, as does the question how the two options relate to sustainability. Research into the mechanisms of agenda setting in plant genomics research, the extent of corporate control and the diversity of society's evaluation of such issues is worth attention.

5. Implementation in the production chain

The advances in knowledge on genome organization, gene sequences, gene functions and gene expression are applied in marker-assisted breeding, transgenic crops and expression profiling. The bulk of the ongoing efforts are taking place in universities and research institutes, mainly funded by governments, i.e. public money. Because of the high costs of molecular genetic research, applications implemented by commercial companies can be found primarily in larger breeding companies involved in producing varieties in crops with a high economic impact. Application of marker-assisted breeding is most advanced in arable crops, such as maize and rice (large cultivation areas) and in vegetables such as tomato (high seed prices).

Transgenic crops cultivated globally encompass 8 varieties in 6 species, i.e. cotton, maize, canola, soybean, papaya and squash melon. Especially, the first four species represent bulk crops with a high economic value because of their large cultivation areas. In addition to development costs, costs for approving and adopting GM crops for market introduction and consumer acceptance are high and hamper large scale implementation in 'smaller' crops by smaller companies.

Expression profiling is based on linking specific quality traits to expression of specific genes, either during cultivation (monitoring crop health, disease infection status) or after harvesting (in storage, transport or at the retailers). The arrays used for this allow diagnostic sampling and monitoring of crop quality through the entire production chain. This facilitates crop cultivation management systems aimed at durable growth reducing the usage of agrochemicals and it enables producers, retailers and consumers to critically monitor product quality at all respective stages. Several specialized biotechnology companies are presently developing such expression arrays.

6. Potential Areas for concrete Points of Action

Regarding the state-of-the-art in technology, bioinformatics and societal aspects involved in genomics research as presented in the previous chapter several main points of action can be identified that will contribute to the desired transition into a sustainable agriculture in The Netherlands. These main points are resistance to pathogens, drought or salt tolerance, bio-based economy, consumer demands and chain management. Moreover, a clear and transparent, two-way communication about the issues involved and how to tackle them is a prerequisite. The establishment of a Technological Top Institute (TTI) on 'Green Genetics' in The Netherlands will stimulate and facilitate initiatives embedded within the BSIK program and co-ordinated by TransForum. Together with projects funded by IOP-Genomics and by other programs of the Dutch Ministries of Economic Affairs and of Agriculture, Nature and Food Quality the efforts on extending and exploiting plant genomics in The Netherlands can be significant and have a recognizable impact.

6.1 Resistance to pathogens.

In the year 2000 (most recent data available, CBS Voorburg/Heerlen) a total of 5.346.278 kg of chemical crop protectants (6.6 kg/ha) was used in the Netherlands. In 2001 in a joint initiative, representatives of the government and producers (farmers/growers) agreed to work towards a reduction in the use of chemicals in agriculture and horticulture. The formulated goal is to have a reduction in 2010 with 95% compared to the level of 1998. The obvious benefits of such a reduction would be on health and economic aspects for the growers themselves and for consumers (less exposure, less residues, less production costs, lower prizes), and on environmental aspects (less exposure of non-target organisms, less water contamination by spilling or washing out). The majority of the chemicals is used in the production of potatoes, bulbs and fruit. In these particular crops it proved to be extremely difficult if not impossible to achieve to goals set because of the open cultivation systems and the long or difficult conventional breeding processes. Sources of resistance are usually scarce and can be broken by the pathogens with relative ease. Stacking of multiple resistance genes or using non-host systems can contribute enormously to the durability of the trait and are explicit features that can be achieved using the full potential of plant genomics, either by molecular marker assisted breeding or by genetic modification. Integrated resistance management in cultivation or agro-systems can be greatly facilitated by the use of several types of diagnostics. For instance, using sampling and monitoring systems based on the known DNA sequences of important pathogens it can be easily established in cultivation at what point in time which pathogen is entering the field. Expression arrays can be used to monitor the onset of expression of genes involved in the plant's defense against patho-

gens, hence of infestation, prior to the moment at which infection becomes visible. This will make it possible to reduce the amount of chemicals necessary for control, enhancing durability in the primary production. Preventive spraying will not be required anymore. The main (not the only!) parties within the Dutch production chain with an interest in this trait are primary producers and specific consumer groups.

6.2 Drought or salt tolerance.

Although coping with adverse environmental conditions such as drought or salt stress is primarily a global problem at the moment, there are at least three arguments for tackling this problem also in The Netherlands. Changing climate conditions, such as the global warming up, could have consequences for The Netherlands in increasing the periods of time in which insufficient rainfall will take place or in increasing the surface of particular soil types, i.e. sandy soils, that are troubled by insufficient rainfall or by the reduced possibilities for artificial raining by pumping up and spraying surface water because of water shortages for this purpose. Another effect could be the rise of the sea level that could result in deeper penetration of salty ground water in coastal areas and thus creating more salt growth conditions for crops in those areas. Breeding companies based in The Netherlands are nowadays part of international conglomerates. This means that their skills and efforts should also be focused on global themes and problems. When they are to keep their leading position in plant breeding they should also work on developing crops tolerant to environmental stress conditions. For developing countries with large marginal areas, arid and salt, this is of great importance. The main parties within the Dutch production chain with an interest in this trait are breeding companies.

6.3 Biobased Economy.

A group of renowned experts and personalities prepared a vision document and presented it to the European Commission. It is called 'Plants for the Future – a European vision for plant genomics and biotechnology'. A few quotes from this vision to start this chapter: "investing more in researching plant genomics will pay potentially huge dividends", "there are enormous opportunities to use plants much more effectively to meet the challenges and demands facing European society in an environmentally friendly and sustainable way", "the future competitiveness of Europe's agricultural and food processing industries will depend on plant genomics, biotechnology and their smart application". Some of the approaches to the goals in exploiting as full as possible the potential of plant genomics that are suggested are compliant with the goals of TransForum, e.g. improve the genetic diversity of crop plants, develop renewable materials and develop more efficient biofuels.

The crucial starting point in the developing perception on the importance of a biobased economy is that today's society based on mineral oil and its derivatives can not last for ever. The

stocks of mineral oil are limited and will be exhausted in the near future. With the expanding economies of countries such as China and India and the need for their share of the oil stocks for maintenance and growth of their welfare and prosperity, in addition to the needs of the Western nations and other emerging economies, it is clear that we need to look for alternatives. Plants as efficient converters of solar energy into biological material can provide raw materials, specialty compounds and energy through biofuels. The main party within the Dutch production chain with an interest in this trait is the processing industry.

6.4 Plant Metabolites for human health

It has long been known that the metabolites in our food have a profound influence on our health in both positive and negative terms. One can immediately think here, e.g. of plant-based vitamins and micronutrients that are essential, or non-nutritional components in our daily diet that are required for maintenance of a healthy physiology. On the negative side, plants are also major sources of health-threatening metabolites such as toxins and poisons which, at very low ingestion levels can even prove fatal. Such metabolites have, contrastingly, frequently proven ideal lead compounds for the development of new drugs and pharmaceutical products with application in e.g. the treatment of cardiovascular disease (digoxin) and cancer (taxol, vincristin). As such, knowledge of the occurrence and biosynthesis of plant metabolites is important both to our understanding of the fundamental biochemical nature of our food as well as to the future analysis of the hugely under-exploited natural chemical biodiversity which is available within the plant kingdom. Currently, new tools are being developed for the rapid and broad analysis of the biochemical composition of plant extracts. Metabolomics has been reported to represent a major advance in our capacity to define biochemically biological materials which generates unique information with a broad application in the fields of e.g. food science, nutrition, plant breeding, processing technology etc.

6.5 Data integration for plant comparative genomics

In the wake of the whole genome sequencing projects, monolithic databases have been developed to accommodate data for a number of model organisms, the so-called Model Organism Databases (MODs). For plants, the first and most comprehensive one has been the Arabidopsis Information Resource (TAIR). With the accumulation of ever more complete genome sequences and associated datasets of ESTs, gene expression, and metabolites, a trend has set in to integrate these resources at higher taxonomical levels. The data warehouse solution EnsMart of the Ensembl MODs on vertebrate species is a prime example of this trend. Similar developments are underway for crops, e.g. the SOL Genomics Network website for Solanaceae. The power and successes of comparative genomic approaches underlie these developments: more can be learned about, e.g. Brassica through its comparison with Arabidopsis than by studying Brassica

alone. The taxonomic integration, however, still is much confined to genome sequence data, partly due to the unavailability of other datasets and partly because most computational tools for comparisons have historically been developed for DNA and protein analysis. An important direction for taxonomic data integration and software development will therefore be aimed at tools for cross-species comparative analysis of e.g. gene and metabolite expression datasets.

6.6 Implementation of the 'Omics' technologies in the production chain

Within the chain of agro-food production, physiological processes affect the quality of the produce. Factors that influence potential shifts in quality are manifold, including crop variety, developmental and physiological state of the product at the time of harvest, and post harvest and processing procedures and conditions. The development of dedicated arrays and diagnostic biosensors based on studying the genome (genomics), gene expression (transcriptome), gene protein products (proteomics), and the resulting chemical composition (metabolomics) will enable high-throughput monitoring and prediction of different aspects of plant or plant product quality at all stages of the production chain, information that is valuable for decision support in chain management. Genomics provides a diverse and powerful toolbox for the development of bio-marker based diagnostics, including e.g. microarray chips for monitoring shifts in gene expression, and metabolomics to discover key metabolites associated with physiological changes. The same genomics technologies are also instrumental to uncover the underlying metabolic and genetic basis of these processes, which in turn will deliver leads for breeding for product chain quality traits.

It will allow tracing and tracking of components (e.g. health-promoting substances), general product quality (off-flavors, decay), stress, senescence etc. as the result of storage or logistics or shelf-life. Based on these instruments all parties within the chain can decide whether or not to buy or sell the product.

6.7 Consumer demands

In the past, consumers have not always been the main determining factor in the development of new varieties. As a result of the discussions on GM crops and more generally as a result of more available time, financial means and opportunity through the internet for this, consumers have become more conscious, aware and interested in aspects of the production of food, feed and flowers. More and more they claim transparency and a role in determining traits that they would like to see in new crops or varieties. Traits such as increased shelf-life, good taste and healthiness are considered important. Also, the retailers support the consumers in formulating their desires or suggesting their needs. Although not always possible in all aspects of a crop, producers should communicate more frequently and more in depth with retailers and consumers in a two-way fashion on the aspects that determine a modern variety. Instruments for this type of communication need to be developed, if not available yet. The main parties within the Dutch production chain with an interest in this aspect are retailers and consumers.

6.8 Chain management

As stated in the previous paragraph it is important to involve all parties from the entire production chain, including end-users, in communicating and in determining the characteristics of new crops and varieties. Directing this process, making sure that all partners contribute and are taken seriously, that all partners comply with the goals that have been set jointly and take the necessary measurements for this, requires an appointed central coordinator, managing the entire chain. The tasks and responsibilities of such a chain manager need to be defined and accepted by all parties. The chain manager should increase the efficacy of implementing through the entire chain the results of the genomics research. All parties are stakeholders in this aspect.

7. Research Agenda and Societal Opportunities

TransForum brings together private companies, governmental bodies and knowledge institutes in projects aimed at the transition of agriculture into a sustainable enterprise. Scientific knowledge should be combined with applied knowledge and societal innovations. To this end, TransForum has two instruments, Innovative Practice Projects (IPs) and Scientific Projects (WPs). In the IPs the emphasis is on true integration of both scientific and applied knowledge present and available with all participants, who actively contribute to reach the project goals. Scientific methodologies, transition management and strategies for increased durability are aimed at testing in practice the strategies for innovation. In WPs some particular questions arising from the IPs can be addressed in more detail and in depth.

Resistance to pathogens has been the main target in several research initiatives, e.g. in Rosaceae such as rose, strawberry and apple, but also in Solanaceae such as potato and tomato (CBSSG). The Innovative Practice Project, VC-010, Healthy pip fruit chain, is an example of a project, using genomics, aimed at making disease resistant fruit varieties, preparing a communication plan to inform the public and determining and monitoring decisive factors in formation and changing of public opinions. Implementation is guaranteed because of involvement of the appropriate representatives of fruit growers.

Industrial crops have been a subject for research for quite some time, firstly in identifying potentially interesting species and looking at problems or potentials in breeding such crops. Calendula and Crambe were taken further as crops with interesting seed-oil composition; hemp as a fibre crop and Miscanthus as an energy crop. Crambe is explored as an ice-breaker crop for GM applications in preparing a non-food, waxy esters (lubrication) producing crucifer in a collaborative efforts with scientists in Sweden. For Calendula, an international production chain has been formed in an IP, IN-008, Innovative Practice project Calendula, in which several aspects of using calendula oil as solvent in paints are tested in several facets of the chain.

Crambe, being a crucifer just like the fully sequenced model crop *Arabidopsis thaliana*, can be used in exploring the potential of comparative genomics alongside parallel initiatives in Brassica, another cruciferous crop. An IOP-Genomics project proposal on Brassica Vegetable Nutrigenomics has recently been granted funding. Here, information available from *Arabidopsis* will be extrapolated directly to Brassica rapa. The first target compounds are glucosinolates, that are also a defined research target in Crambe. Through plant metabolites we are at the crossroads between bio-based economy items and health components defined by consumer demands. This is an example of divergent targets but convergence in approaches. The information generated in one project can clearly help progress in another project.

With the topics mentioned in chapter 2 and the examples and crop families as focal points mentioned here, the stage is set for the formulation of new IPs and WPs.