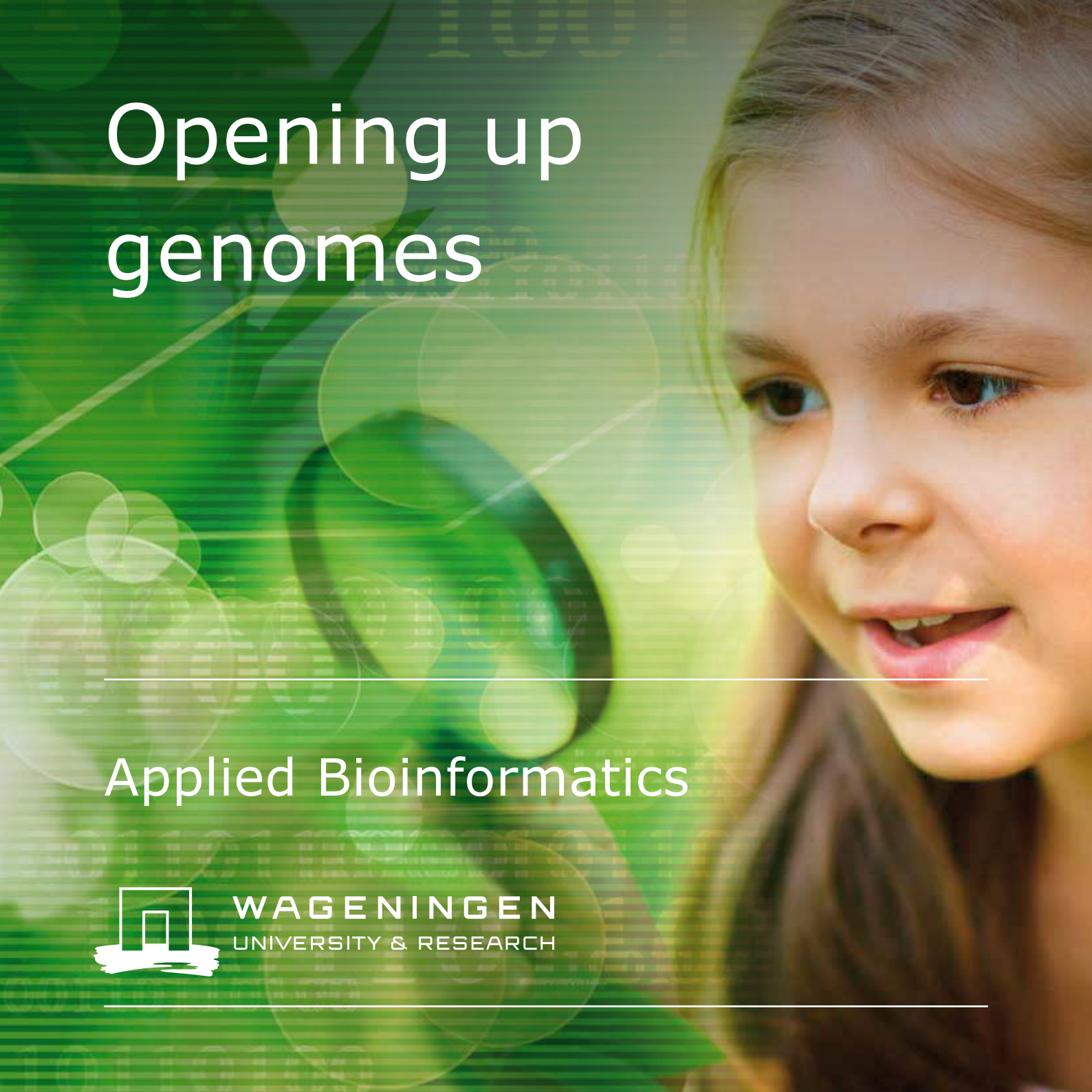
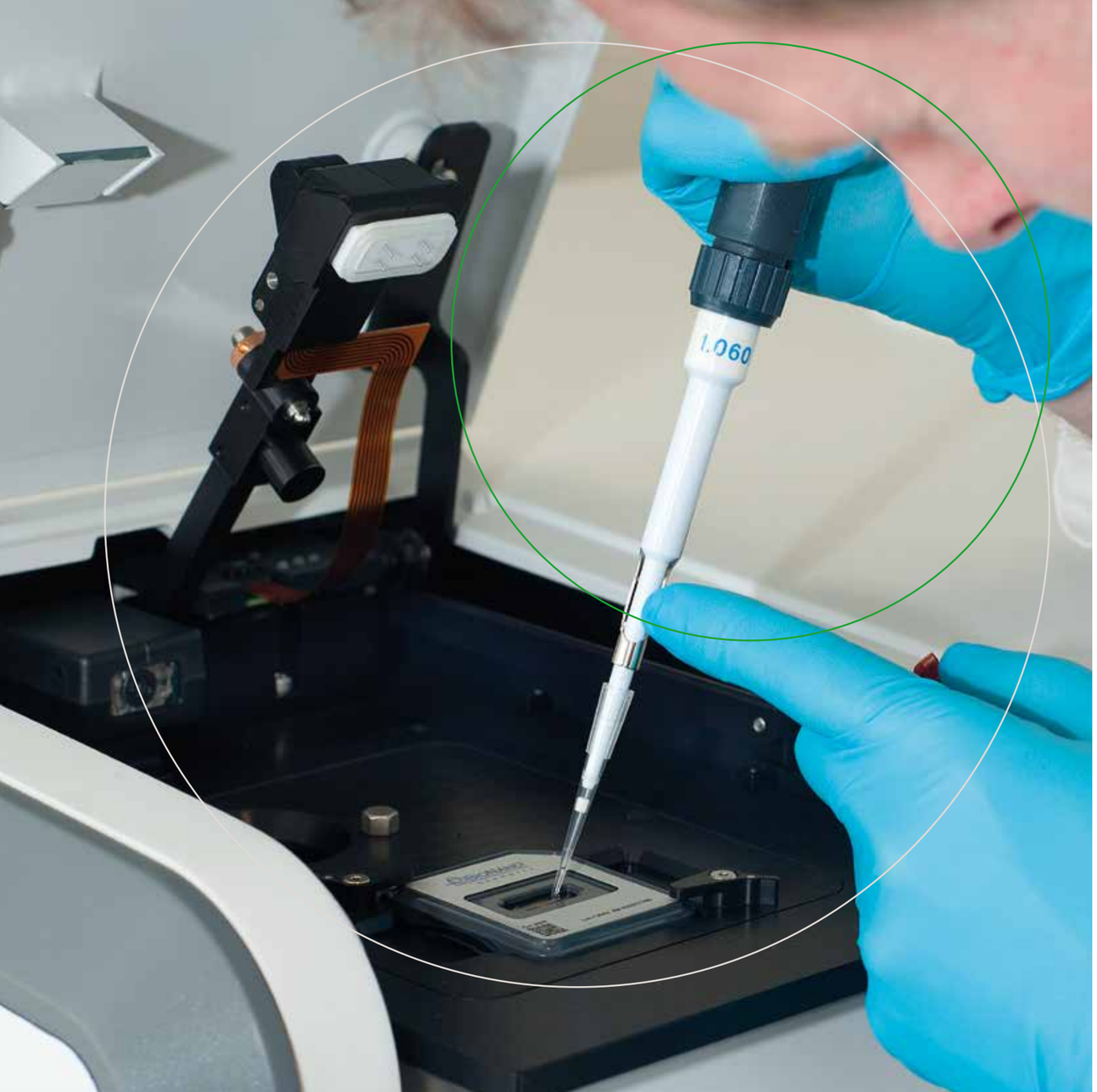
Opening up genomes



Applied Bioinformatics

WAGENINGEN
UNIVERSITY & RESEARCH



Applied bioinformatics for your genomes

We @ WUR offer you the latest high-throughput sequencing, DNA analysis technology and expertise for targeting complex genomes. This we combine with state-of-the-art computational infrastructure focussing on high-quality hybrid assembly, annotation, data integration as well as development and use of appropriate tools and technologies.

Our infrastructure and expertise boost biological innovation by using the DNA sequence or RNA complement of the genome. Advances in DNA/RNA applications require a reference genome, a pangenome or genomes of populations to drive progress.

All genomic DNA may be equal, but some genomes are more equal than others. Determining a genome sequence is not really a challenge any more, unless that genome is polyploid, highly repetitive or otherwise complex.

Effective exploitation of genomes for breeding, biotech, biomedical, or veterinary applications, is a different story. You can easily drown in data and tools available. Delivering genomes in ways that can be directly used for research and application is our core expertise. It allows us to help you extract considerable added value from large data sets. Your data. Your BIG data.

Genomes at work: preparing for petabytes

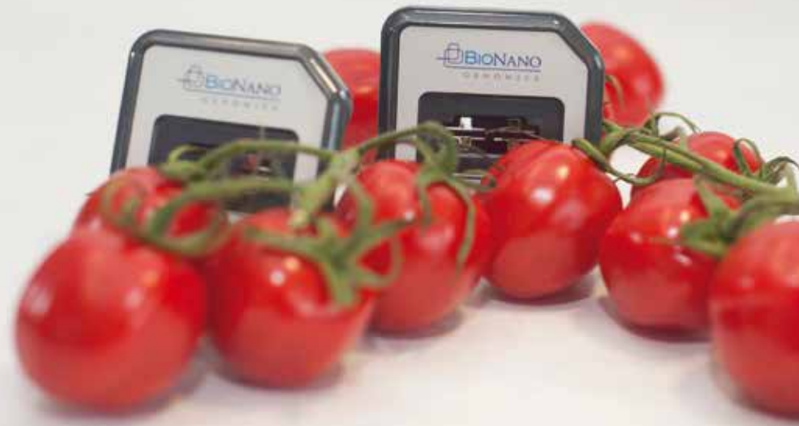
Need a genome? Want to compare genomes? Need to know about gene expression or function? Determining full DNA genomes is now attractive and affordable. Yet, every sequencing platform has its best use and requires a targeted approach to extract maximum value. We offer broad experience and expertise in sequencing platforms, both modern long-read single-molecule and short-to-medium size PCR-based sequencing. State-of-the-art technologies are available for earlier unthinkable applicatons. We specialize in complex genomes and metagenomes and are prepared for petabytes. Data are assembled in full genomes using the best software available, adjusted to the challenges at hand. Hybrid assemblies of complex genome data is our forte. Big data may surprise you, they have no surprises for us.

The longest DNA molecule we have sequenced on our PacBio is now over 60 kB long. Routinely we generate 1 Tb sequence per full HiSeq 2500 run. Hybrid assembly of complex polyploid genomes such as potato is progressing towards smooth haploblock typing, analysed and visualised with smart combinations of tools available for public, private or in-house applications.

Genome mapping for innovation

Do you doubt the quality of your assembly? Need to improve? As one of the first in Europe, we routinely perform optical mapping for complex crop genomes. Data from in-house generated high molecular weight DNA isolations are fully integrated in our sequencing and assembly pipelines. Genome assemblies of unparalleled quality are generated that bridge the gap between base pairs and chromosomes and between geneticists, breeders and DNA.

Optical mapping enables us to routinely analyse DNA restriction fragments up to 1 Mb in size and reconstruct chromosomal fragments up to 11 Mb in size. This way, we identify and correct major errors in, for example, the public tomato genome assembly. The combination of optical mapping and hybrid assembly defines the new standard for genome quality.



In silico success

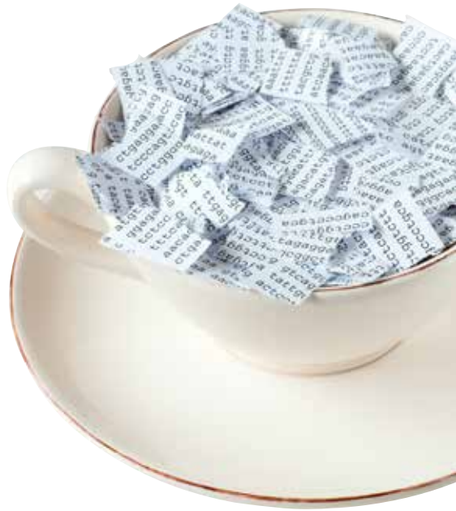
Any idea what the part of the genome you are interested in is actually doing? Where are the genes and regulatory elements? How do these relate to traits of commercial interest? How can these be used for crop improvement, better diagnostics or more resource-efficient production? A genome sequence is only useful when it contributes to your understanding of its function.
We are specialised in automated annotation of function using the latest software, incorporating available public and/or your in-house data. Moreover, we master innovative algorithms to predict gene function using networks of heterogeneous data.

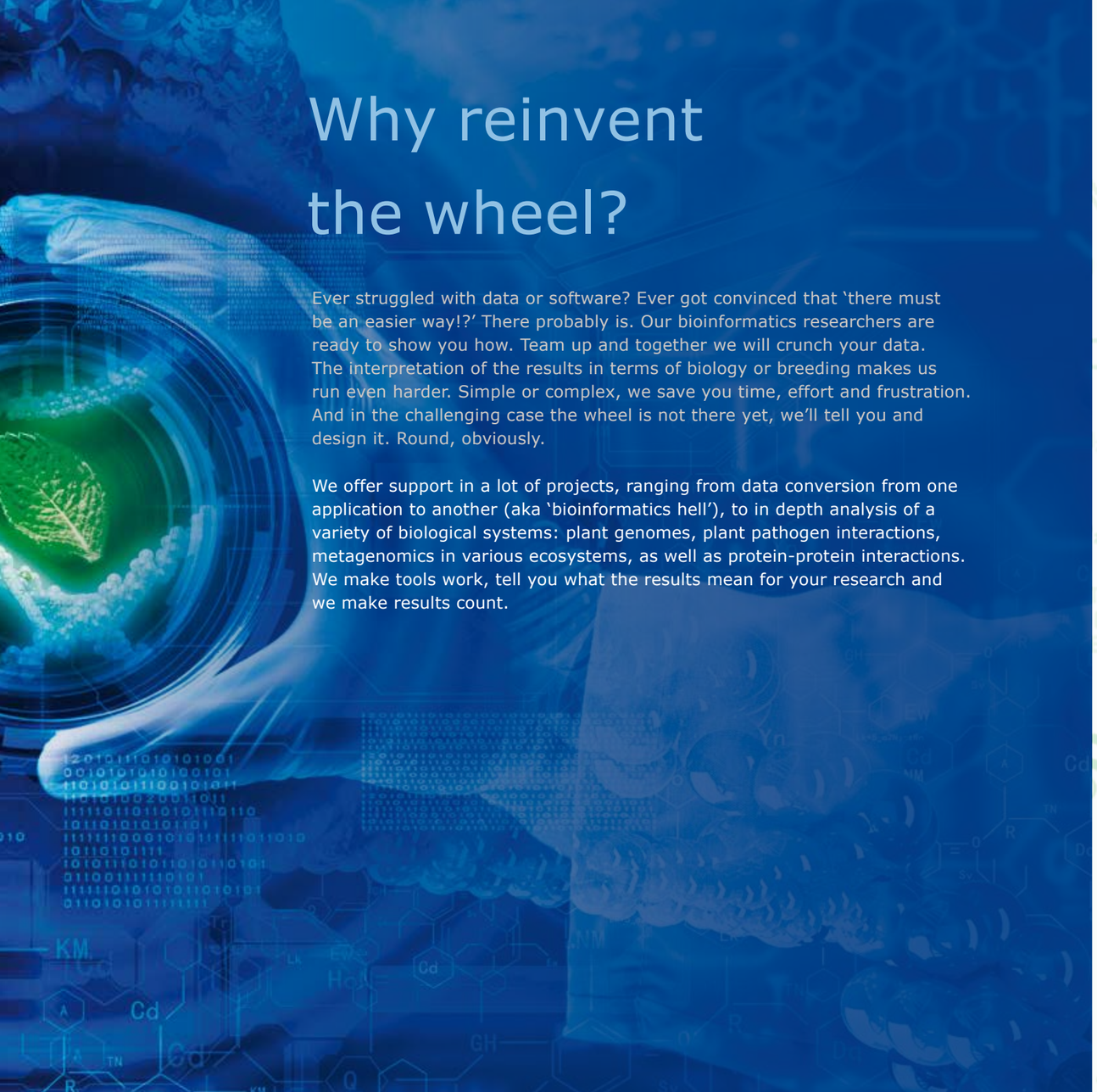
For automated annotation of genes, regulatory elements or other elements in any genome, we exploit an integrated computational pipeline based on MAKER. We excel in supervised machine learning to assign functions to genome elements based on genetic variation. To guide and speed up costly experimental validation of annotation, we are experts in the prediction of gene function (hence traits-of-interest), using network-based numerical approaches and big data integration.

Markers for value

How do you exploit genomic data for rapid breeding strategies? You will need markers and platforms for analysing plants and populations for traits-of-interest. We are at the forefront of the analysis of genomic variation. Moreover, once you have genomic regions linked to traits, we offer advanced computational methods to zoom in and help predict the causal candidate genes. This includes exploiting your latest imaging data for phenotyping. We translate introgression events into breeding strategies that promote obtaining the desired combination of genes faster.

We eat the identification and analyses of SNPs for breakfast. Using advanced computational and statistical approaches, we identify the most likely candidate genes that result from QTL mapping efforts. This way, we establish links between genotype and phenotype faster and more reliable than others, improving the cost-effectiveness of QTL approaches. Our advanced analyses and visualisation of genome variation allows identifying patterns of recombination that predict the likelihood of success for a cross. We help to accelerate breeding by proper selection of parents in breeding schemes.





Why reinvent the wheel?

Ever struggled with data or software? Ever got convinced that ‘there must be an easier way!?’ There probably is. Our bioinformatics researchers are ready to show you how. Team up and together we will crunch your data. The interpretation of the results in terms of biology or breeding makes us run even harder. Simple or complex, we save you time, effort and frustration. And in the challenging case the wheel is not there yet, we’ll tell you and design it. Round, obviously.

We offer support in a lot of projects, ranging from data conversion from one application to another (aka ‘bioinformatics hell’), to in depth analysis of a variety of biological systems: plant genomes, plant pathogen interactions, metagenomics in various ecosystems, as well as protein-protein interactions. We make tools work, tell you what the results mean for your research and we make results count.

Beyond DNA

Do you feel that all that DNA stuff that bioinformaticians are so thrilled about is interesting but rather overhyped? Is your real interest in proteins, what they do, and the pathways they are involved in? Or you are into metabolites and the integration of so-called multi-omics data? If you are struggling with the complexity of big data integration (genomics + transcriptomics + metabolomics + proteomics + phenomics), struggle no longer. We have access to state-of-the-art tools for analysis, or build these ourselves. Our expertise in machine learning allows us to detect patterns in data that steers future research and directs future business. We help you translate complexity into understanding.

We offer smart data exploration for multi-omics data and make attractive infographics to facilitate the interpretation of your data. We help you discover potentially novel features in your data with the use of metabolic pathways integrating KEGG and Galaxy workflows, or custom-made modules for the identification of metabolites. We offer you the benefits of advanced tools and expertise for data classification as well as pattern recognition, supervised or unsupervised. Our expertise in feature selection let you focus on those aspects of your data that are most relevant. We perform statistically sound data integration offered to you in a way you understand. Uncanny data complexity will transform into happy feelings about the answers we let the data give you.

Early adoption is in our genes

Already heard about the next game-changing sequencing platform or the new revolutionary bioinformatics software? Able to apply it immediately to your research and business? We know that you know that it is an almost impossible challenge: keeping up with the latest developments in sequencing technologies, bioinformatics or data mining for squeezing novel insights out of big data. We therefore routinely scout for potentially disruptive technologies to provide a competitive edge, technologies that add value without hype. We quickly adopt and fine tune new technologies to make them fit for purpose. We ensure a smooth transition from the expensive development stage to a cost-effective production phase. We cherish a position as your sparring partner for data-driven innovation. Join our unique technology watch program.

Based on market developments and needs, we continuously evaluate the potential of emerging technologies or novel bioinformatics solutions. We collaborate with CAT-AgroFood of Wageningen University & Research to acquire advanced platforms for analyses and to create new collaborative opportunities. We establish close contacts with early providers of promising technologies (PacBio, BioNano Genomics, 10x Genomics), so we are at the forefront of new applications and technological improvements. When sufficient added value and interest, we investigate how new technologies fit your needs and we go implement these with you.

Track record (selection)

Cnidaria: fast, reference-free clustering of raw and assembled genome and transcriptome NGS data.
Aflitos et al. (2015) BMC Bioinformatics 16, 352, DOI: 10.1186/s12859-015-0806-7.

Introgression browser: high-throughput whole-genome SNP visualization.
Aflitos et al. (2015) Plant Journal 82, 174-182, DOI: 10.1111/tpj.12800.

Towards recommendations for metadata and data handling in plant phenotyping.
Krajewski et al. (2015) Journal of Experimental Botany 66, 5417-5427, DOI: 10.1093/jxb/erv271.

Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing.
Aflitos et al. (2014) Plant Journal 80, 136-148, DOI: 10.1111/tpj.12616.

Biological process annotation of proteins across the plant kingdom.
Bargsten et al. (2014). Current Plant Biology 1, 73-82, DOI: 10.1016/j.cpb.2014.07.001.

Prioritization of candidate genes in QTL regions based on associations between traits and biological processes.
Bargsten et al. (2014) BMC Plant Biology 14, 330, DOI: 10.1186/s12870-014-0330-3.

The genome of the stress-tolerant wild tomato species Solanum pennellii.
Bolger et al. (2014) Nature Genetics 46, 1034-1038, DOI: 10.1038/ng.3046.

Rice cytochrome P450 MAX1 homologs catalyze distinct steps in strigolactone biosynthesis.
Zhang et al. (2014) Nature Chemical Biology 10, 1028-1033, DOI: 10.1038/nchembio.1660.

System-wide hypersensitive response-associated transcriptome and metabolome reprogramming in tomato.
Etalo et al. (2013) Plant Physiology 162, 1599-1617, DOI:10.1104/pp.113.217471.

A large-scale evaluation of computational protein function prediction.
Radivojac et al. (2013) Nature Methods 10, 221-227, DOI: 10.1038/NMETH.2340.

The tomato genome sequence provides insights into fleshy fruit evolution.
Sato et al. (2012) Nature 485, 635-641, DOI: 10.1038/nature11119.

Need help or need help desperately?

Problems in getting your (complex) genome of interest? Unsure about the quality of your genome assembly? Difficulties in comparing genomes? Urgently need the function of a genes or gene regulation? Is the function of a protein bothering you? Curious about the nature and potential use of a metabolite? Desire to save money in QTL analyses? Looking for new strategies to speed up breeding? Haunted by data analyses, bioinformatics or applied statistics?

Come to us!

The ambition to assist you when you need help is in our genes. We generate data and/or we analyse these. For you. With you. Our unique combination of state-of-the-art platforms for data generation in sequencing (DNA, RNA), optical mapping, as well as protein and metabolite analysis, is combined with vast biological, biochemical and breeding experience and expertise. In any academic and/or commercial context, we will help you to recognize the real biological relevance and application in discouragingly large piles of data. Your data. Your BIG data.

Platforms available

For data generation

- *PacBio Sequel*
- *Illumina Miseq*
- *Illumina HiSeq 2500*
- *BioNanoGenomics Irys*
- *10x Genomics Chromium*
- *Oxford Nanopore Technologies MinION*

For data analyses

- *Cluster with 100's CPU's integrated with GPU's*
- *CLCBio*
- *State-of-the-art tools for bioinformatics and biostatistics*
- *In-house tools for advanced analyses and visualization (BMRF, MeTOT, iBrowser and more)*
- *Easy access to HPC and supercomputer centres (SURFsara/ HPC-WUR)*

Contact
Wageningen Plant Research
Dr. Gabino F. Sanchez Perez
Phone: +31 317 48 17 42
E-mail: gabino.sanchezperez@wur.nl

www.wur.nl/appliedbioinformatics
www.wur.nl/catagrofood
Droevendaalsesteeg 1
6708 PD Wageningen
The Netherlands

WAGENINGEN
UNIVERSITY & RESEARCH