

Link

[plantbreeding.wzw.tum.de](http://plantbreeding.wzw.tum.de)

**Rather than analyzing individual genes, the team considers entire DNA profiles.** They compare numerous lines of maize and search their profiles for patterns to which beneficial characteristics can be assigned.



# Intelligent plant improvement: Searching for the needle in the haystack

The genes responsible for different characteristics of plants make up a complicated puzzle that, to date, researchers have had little success solving. Prof. Chris-Carolin Schön's interdisciplinary team at the TUM School of Life Sciences in Weihenstephan is using machine learning methods to analyze millions of combinations. This is an effort to find the best genes for particularly efficient, productive maize plants.

**A**re you looking for a stylish ceiling light for your living room? Perhaps you're interested in e-bikes, smartphones or exciting audiobooks? For every search request a user makes, their web browser lists a wide selection of suitable products, often accompanied by the note: "Customers who viewed this item also bought ...". The recommended items are usually exceptionally well suited to the user's preferences. This is enabled by collecting millions and millions of search requests and comparing them with Internet users' personal details – from their shoe size and place of residence to their last holiday destination. Mathematical algorithms search in this gigantic data pool for prominent associations and make use of recurring patterns. This makes it possible to research and influence customers' buying behavior.

"We work with similar mathematical models to forecast the characteristics of future plant varieties," explains Chris-Carolin Schön. The Professor of Plant Breeding conducts research on key crops, such as maize and sunflower, at the TUM School of Life Sciences in Weihenstephan. "We want them to thrive and produce high yields, also in an era of climate change," emphasizes Schön. "We need robust varieties adapted to the respective climate to secure food supplies for a growing global population and, in doing so, to keep the use of water, energy and agrochemicals as low as possible."

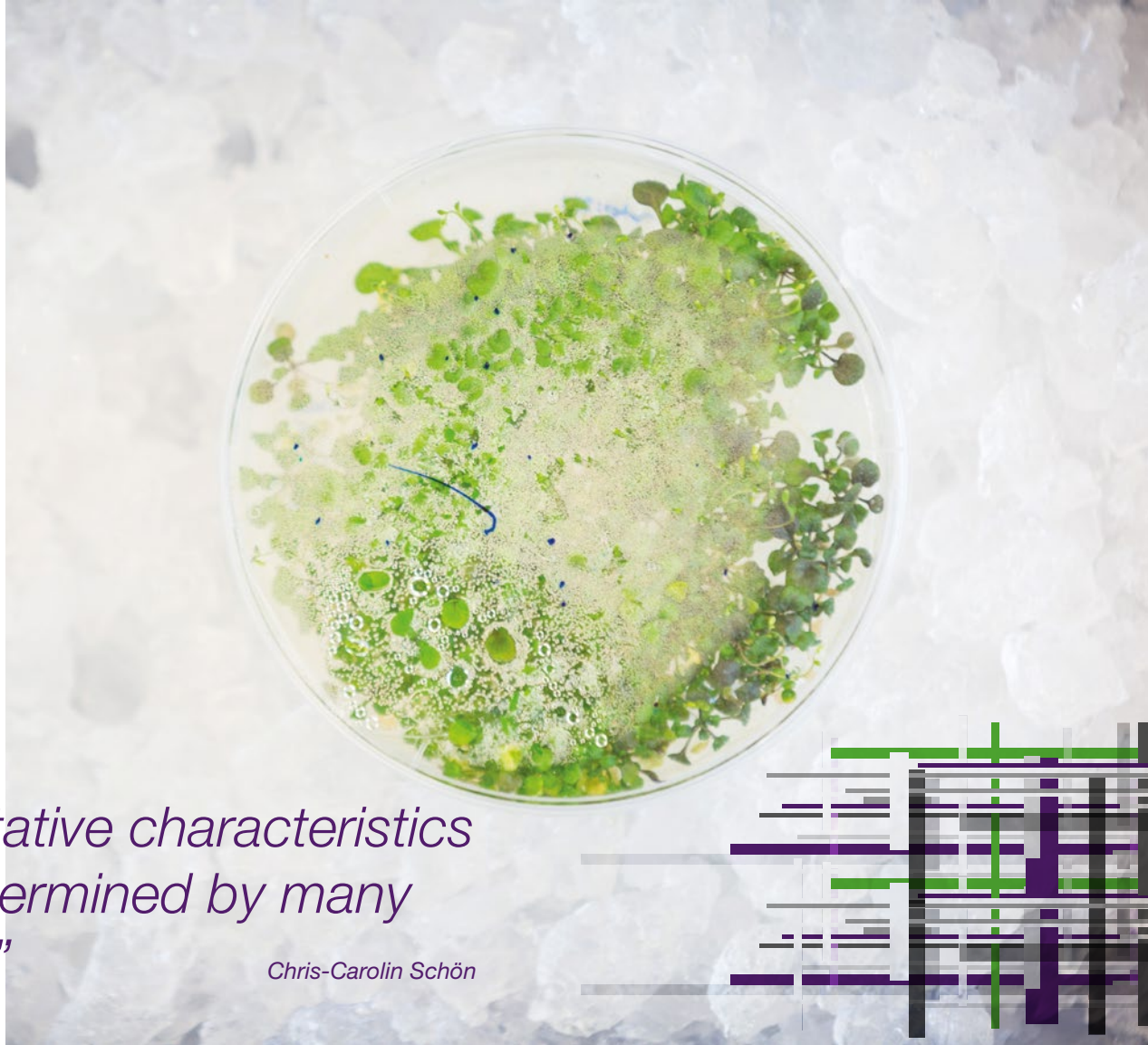
Maize is one of the world's most important crops for the production of food, animal feed and fuel. Cultivating maize in Europe entails certain challenges.



**Starting from regionally adapted maize landraces, the researchers have produced almost a thousand lines for genetic analysis.**

**When analyzing young plants, the team measures cold tolerance and early development of each maize line – important characteristics for environmentally friendly cultivation.**





*“Quantitative characteristics are determined by many genes.”*

*Chris-Carolin Schön*

**Chris-Carolin Schön and her team** not only analyze maize but also other important crops, such as rye and sunflower.

Picture credit: Tom Freudenberg/TUM; Graphics: edlundsepp

Maize is planted late in spring and develops slowly due to high temperature requirements, bearing the risk of soil erosion, nutrient depletion and the need for herbicide treatment. In addition, young maize plants can easily become overrun by wild herbs, which necessitates the use of herbicides. “We can address these challenges with new, cold-tolerant varieties that develop more quickly at the beginning of the vegetation period and allow farmers to plant earlier,” explains Chris-Carolin Schön.

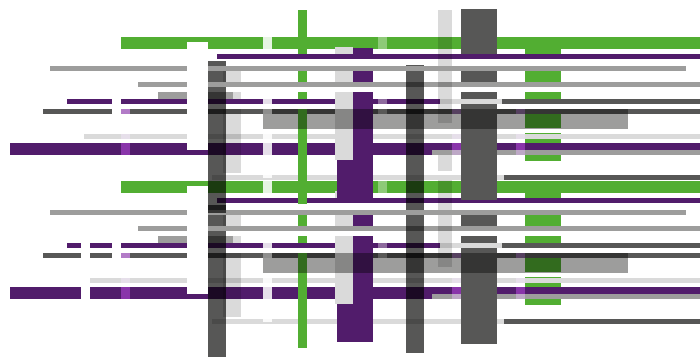
The geneticist points out that this work also relates to changing the flowering time: “To avoid drought stress over the summer months, the plants should flower early – but not too early, because that would have serious impacts on yield. Finding the optimum is complicated.” ▶



The development of young plants, their robustness, their flowering time and their yield are typical quantitative characteristics. This means the trait values follow a continuous distribution and it is difficult to distinguish between slow and fast development, robust or fragile plants, early or late flowering, high or lower yield. However, it remains unclear which genetic factors control the manifestation of these characteristics. “We are increasingly coming to realize that quantitative characteristics are determined by a very high number of genes in complex interactions. This is why, instead of analyzing individual genes, we analyze entire DNA profiles. We’re comparing the profiles of as many different maize lines as possible, looking for patterns to which we can assign beneficial characteristics,” explains Chris-Carolin Schön.

When it comes to characterizing plant traits, the same principle applies as in forecasting customer preferences: the larger the dataset, the more precise the predictions. In the course of their work, the scientists started with three maize landraces to produce almost a thousand lines. Leaf samples from each of these lines were sent to a specialized German laboratory that produced extensive DNA profiles. “We now know the genetic makeup at 600,000 positions along the genome of each and every line,” explains Armin Hölker. The doctoral student became familiar with plant breeding methods during his Master’s thesis. He now works on genetic data in Chris-Carolin Schön’s laboratory to identify the optimal model for predicting the genetic values of maize lines.

“The task is to predict what contribution each individual genetic component can make to the crop’s yield or to other relevant characteristics,” outlines Armin Hölker. To make this prediction, he not only needs DNA profiles but also measurements on the phenotype of as many maize lines as possible. For this reason, a total of over 80,000 plants has been planted from all these genetically characterized maize lines at TUM’s research station in Roggenstein.



Similar trials have been implemented at ten other sites from northern Germany to northern Spain. Once the seeds germinated, Armin Hölker and his colleagues had their hands full, as he explains: “We measured the cold tolerance and early development of young plants as well as further agronomic characteristics such as flowering time, maturity, yield and maximum plant height.”

This comprehensive set of measurement data, combined with the associated DNA profiles for every maize line, formed the data matrix for Hölker’s statistical model. He uses statistical methods to identify which DNA profiles are associated with high yields, faster development of young plants, and other quantitative characteristics. The doctoral student refers to this process as “training”. This is because the model “learns” from comparing genetic data with the phenotypic data obtained from field experiments and matches them as precisely as possible. In the next step, the model has to examine the DNA profiles of maize lines with unknown phenotypes. This is precisely the objective of the task: The intelligent prediction model should make the need for trait measurements, an expensive and time-consuming task, obsolete. “This means that, in future, we will be able to examine thousands of unknown maize lines using only their genetic data to see whether they could lead to new maize varieties better suited to their environment. This will save time and money,” explains Armin Hölker. ▶

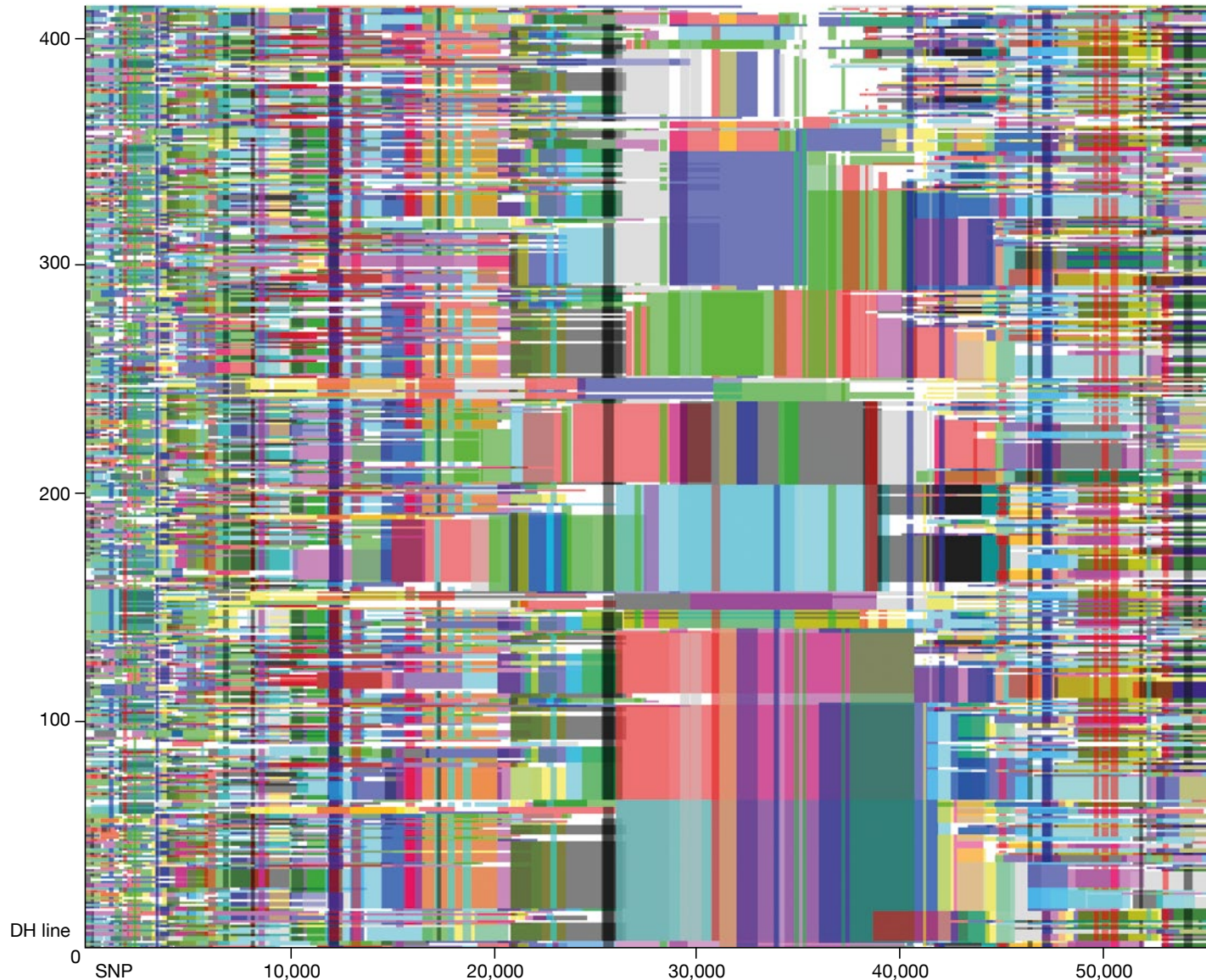
◀ **At TUM’s Roggenstein research station,** more than 80,000 plants were planted to check the appearance of each of almost 1,000 genetically characterized maize lines.

### Genomic selection: Overall picture at a glance

The maize genome consists of 2.3 billion DNA compounds known as nucleotides distributed over ten chromosomes. Only a fraction of these nucleotides form the coding units we call genes. The vast majority do not have any coding function, at least from what is known to date.

Nevertheless, these nucleotides can also be used to predict specific plant characteristics. Their overall combination and sequence forms a specific pattern in every maize line that separates it from all others. The image below depicts more than 70,000 genomic markers (horizontal axis) from a maize landrace represented by 409 individual lines (vertical axis) and shows chromosome 1 (in total, maize has 10 chromosomes).

If the nucleotide pattern of a particularly productive or drought-tolerant maize line has been identified, it is highly probable that plants with a similar pattern will be similarly productive or drought-tolerant. Consequently, these patterns alone make it possible to detect lines that possess the desired characteristics when selecting among a high number of plants with unknown phenotypes. The benefits of this process – known as genomic selection – are clear: It allows us to differentiate between productive and less productive plants and to select the more useful strains without the need to assess their appearance in field experiments.



The extensive training dataset of DNA profiles and phenotypic measurements is not only suited to predicting the genetic value of plants. “The great thing about it is that the same data can also help us to shed light on the genetic mechanisms underlying specific characteristics,” says Chris-Carolin Schön. “We regard maize as a model plant and work on fundamental questions that also arise with other crops: How do plants cope with drought stress? Do drought-resistant plants also grow well under optimal conditions and produce high yields?”

The insight we gain from this is of interest to all plant breeders – whether they’re on the search for new varieties of sugar beet, wheat or sunflower.” Of course, the TUM researcher adds, new plant varieties will not be able to stop climate change. Nevertheless, she believes it is immensely important to develop new varieties for sustainable production: “If we can reduce the input of nutrients, agricultural chemicals, energy and water, we’ll have made a big step forward!”

■ *Monika Offenberger*



---

#### **Prof. Chris-Carolin Schön**

---

Schön studied agricultural sciences at the University of Hohenheim and was awarded her Master’s degree by Oregon State University, USA, in 1990. In 1993, she received her doctorate from the University of Hohenheim and subsequently served as Coordinator for New Technologies at KWS SAAT SE. From 1996 to 2007, Schön was active as Director of the State Plant Breeding Institute at the University of Hohenheim, where she also completed her habilitation in plant breeding. She has also worked as a visiting researcher at the Molecular Plant Breeding CRC in Adelaide, Australia. Chris-Carolin Schön has held a full professorship in plant breeding at TUM since 2007; she was offered a chair at the University for Natural Resources and Life Sciences (BOKU), Vienna in 2009 and at Heinrich Heine University (HHU), Düsseldorf in 2015.

---