All living things use the thread-like molecule DNA as their genetic material. One complete copy of the DNA of an organism is called its ‘genome’. Sequencing genomes has revolutionised our understanding of evolution and how heredity works. This trail will explore the genomes of some familiar plants and reveal some of their evolutionary secrets.
DNA in the Garden

DNA (short for deoxyribonucleic acid) is the material that stores the instructions for making the components of all living things and enables this information to be passed to future generations. It acts like an instruction manual for the organism.

DNA is composed of four basic building blocks called ‘bases’, which are nitrogen-containing biological compounds. They form base pairs and stack on top of each other, leading directly to the formation of the double-helix DNA structure.

‘Genes’ are sections of DNA that contain the code for a particular trait or characteristic like eye colour. There are a number of different varieties of the same gene (known as ‘alleles’); so a gene coding for eye colour can code for blue, green, brown etc. Humans have up to 23,000 genes. Genes are packaged together into spaghetti-like structures called chromosomes, which normally occur in pairs – half from one parent and the other half from the other parent. Humans have 23 pairs of chromosomes. Many plant species, however, have more than two sets of each chromosome; this is known as their ‘ploidy level’.

One complete copy of the DNA of an organism, including bits that don’t code for anything, is called its ‘genome’. Genomes are measured in millions of base pairs (Mb = Mbp) = 1,000,000 bp.

See an online version of this trail at www.botanic.cam.ac.uk/dnatrail
The amount of DNA in a genome differs dramatically between species. Unwinding the DNA in a cell of the plant with the largest genome, the fritillary, would stretch a staggering 84 metres, compared with just over two metres in our own genome.

Having a large amount of DNA carries an inherently high biological cost. The more DNA in an organism, the longer it takes to copy it. Plants with large genomes thus grow slower than their slim-genome relatives. This can restrict the type of life strategy that a plant adopts, such as whether it is an annual or perennial.

DNA sequencing involves determining the order of the DNA base pairs. Knowing the sequence for the entire genome of an organism, and understanding what each gene codes for, allows us to understand the evolutionary history of a species and how organisms respond to their environment. The code can then be linked to traits, such as flower colour and shape, which are the manifestation of the genetic code in the living organism.

Ensembl Plants is a genomes database produced by EMBL’s European Bioinformatics Institute (EMBL-EBI) and holds the genomic data of many plant species, some of which have been highlighted in this trail.

For each plant species featured in this trail you will learn something about the size of their genome, their chromosome number and how many sets of chromosomes they have (you can learn more about this on the back page of this trail leaflet). In addition you will find out how sequencing the genome of these species has revealed their evolutionary secrets.
Poplar (*Populus trichocarpa*)

485Mb, 19 chromosomes, diploid

Poplar is a deciduous, broadleaf tree native to western North America. It is also an important source of timber. Poplar is fast growing, reaching reproductive maturity in as few as 4-6 years. Moreover, its genome is relatively small and for that reason it was the first tree species to have its genome sequenced. It has become an important model tree species for genetic, taxonomic and evolutionary studies due to its small genome size, fast growth, ability to reproduce vegetatively and relatively long lifespan. Poplar species are predominantly dioecious, meaning that their male and female reproductive structures are located in different individuals, preventing self pollination. This, along with being wind pollinated, results in high levels of gene flow between populations and hence different populations are relatively similar to each other.

Potato (*Solanum tuberosum*)

840Mb, 12 chromosomes, tetraploid

Potato is the world’s fifth most important food crop after rice, wheat, maize and sugarcane. The potato was domesticated in the Andes about 7,000-10,000 years ago but was only introduced elsewhere about four centuries ago. Although thousands of cultivars were bred in the Andes, inbreeding of the few potato cultivars that made their way to the rest of the world has resulted in reduced variation, mainly because the potato is grown from ‘seed potatoes’, rather than seed. One consequence of this reduced variation is that they are more susceptible to pests and diseases. This was one of the causes of the Irish potato famine.

Its genome consists of 12 chromosomes, just like tomato and pepper, and it is tetraploid (having four sets of chromosomes). Genomic tools are routinely used in institutions such as the International Potato Centre in Peru to breed new varieties.
Tomato (*Solanum lycopersicum*)
828Mb, 12 chromosomes, diploid

Tomato is a member of the nightshade family, Solanaceae, which includes a variety of crop plants such as chilli pepper and potato. Like the potato, the tomato originated in the Andean region of South America, was grown by Aztecs in Mesoamerica and spread to Europe by early Spanish explorers. Today, hundreds of varieties are grown throughout the world, with the largest producers being China and the United States. The reference genome for tomato is that of cultivar Heinz 1706. The genome sequences of tomato and related species, such as *Solanum pimpinellifolium*, are essential resources to understand the domestication history of the tomato and for breeding new cultivars resistant to diseases such as blight. In addition to its value as a food, the tomato has served as an important model system for the study of fruit ripening, plant-pathogen interactions and molecular genetic mapping.

Grape (*Vitis vinifera*)
487Mb, 19 chromosomes, diploid

Many different grape varieties have been bred around the world and have played an important role in the cultural heritage of humanity since the Neolithic period, mostly in the production of wine, which uses over 70% of the world’s grapes. A key step in the domestication of the grapevine was the transition from separate sexes (dioecy) in wild *Vitis vinifera* ssp. *sylvestris* (*V. sylvestris*) to hermaphroditism in cultivated *Vitis vinifera* ssp. *sativa* (*V. vinifera*). This happened in the domestication of papaya and strawberry too. The domesticated grape was only the second woody species to have its genome sequenced, and the first fruit crop. The first grape genome published was that of cultivar Pinot Noir PN40024. Sequencing the genome is now rapidly leading to insights into wine aroma and flavour development, and this will hopefully lead to new and exciting wines for us to enjoy!
DNA in the Garden

This trail has been produced in collaboration with:

EMBL-EBI  wellcome  connecting  science
1. *Populus trichocarpa*
2. *Solanum tuberosum*
3. *Vitis vinifera*
4. *Solanum lycopersicum*
5. *Capsicum annuum*
6. *Hordeum vulgare*
7. *Zea mays*
8. *Arabidopsis thaliana*
9. *Brassica oleracea*
10. *Theobroma cacao*
11. *Musa acuminata*
12. *Oryza sativa*

*Please note: many plants on this trail are not visible all year round.*
Barley (*Hordeum vulgare*)

5300Mb, 7 chromosomes, diploid

Barley is the world’s fourth most important cereal crop and it is cultivated in all temperate regions. It was one of the first cereals to be domesticated over 10,000 years ago in the Fertile Crescent. About two-thirds of the barley in the world is used for animal feed. The remaining goes to the malting, brewing, and distilling industries, producing mostly malt whiskey and many types of beer. It has one of the largest plant genomes sequenced to date. Thanks to genomics we now have an unprecedented ability to identify important genes for agronomical traits, such as yield or malting quality, and also genes that help us understand evolutionary history. For example, gene HORVU2Hr1G092290 controls whether a two-row or six-row flower spike is produced, whilst the Brittle genes determine whether mature grains are easily removed, which is a key domestication trait.

Chilli Pepper (*Capsicum annuum*)

3000Mb, 12 chromosomes, diploid

One of the oldest domesticated crops in the Americas, the chilli pepper is also the most widely grown spice crop in the world. It is also a close relative of potato, tomato and tobacco (*Solanaceae*). Chilli peppers are hot because of a substance called capsaicin, which gives them their distinctive hot, peppery taste. This molecule stimulates areas of the skin and tongue that normally sense heat and pain, fooling the brain into thinking they are burning. Capsaicin has been found to repel or poison mammals but not birds. Peppers come in a variety of colours. Chlorophyll, produced in large amounts as the plant grows, is responsible for the unripe, green colour. Reduction of this molecule during late growth gives way to carotene and anthocyanin molecules, which gradually produce yellow, red or purple colours. Sequencing its genome will serve as a platform for improving the nutritional value of *Capsicum* species.
Maize (Zea mays)
2400Mb, 10 chromosomes, diploid

Maize has the highest world-wide production of all grain crops. Whilst in many regions of the world it is a food staple, the majority of maize grown is used for animal feed and ethanol fuel. Maize was domesticated in Central America from wild teosinte. Teosinte had tiny ears, with kernels protected within a hard casing making them very difficult to eat. Selection by farmers transformed the tiny ears of teosinte into large corn cobs. It has a large genome, with 85% of the genes found to be special ‘jumping genes’ (transposons), which can hop out of their original spot in the genome and then wedge themselves in another, random place. When they land, they may disrupt the activities of nearby genes. Maize proved to be the perfect model organism for the study of these ‘jumping genes’, as they can lead to variably pigmented kernels in the cob. The discovery of these genes earned Barbara McClintock a Nobel Prize.

Thale Cress (Arabidopsis thaliana)
135Mb, 5 chromosomes, diploid

Arabidopsis thaliana is a small annual plant native to Eurasia that most gardeners are likely to dismiss as a “weed”. Compared to other relatives of the mustard family (Brassicaceae), such as cabbage or rapeseed, it has little agronomic significance. It is, however, an important model organism for studying plant development due to its short generation time, small size, large number of small seeds and small genome. In 2000, it was the first plant ever to have its genome sequenced. Researchers around the world have used this tiny plant as a Rosetta Stone for plant biology. Many discoveries about flowering were made first in Arabidopsis and then transferred to other species and crops. Today many other plants are being used as models for specific topics. For instance, Brachypodium distachyon is a better model for studying monocotyledonous plants, which include all cereal crops.
**Wild Cabbage (Brassica oleracea)**

488 MB, 9 chromosomes, diploid

Cabbage, broccoli, cauliflower, kohlrabi, kale and Brussels sprouts are all common vegetables, which are simply different cultivated varieties of the same species, *Brassica oleracea*. In its uncultivated form it is called wild cabbage, and is native to coastal Southern and Western Europe. Wild cabbage is a hardy plant with a high tolerance for salt and lime and low tolerance to competition from other plants. It has become established as an important human food crop plant, because it is rich in essential nutrients including vitamin C, which are stored over the winter in its leaves. The genome of *B. oleracea* is extremely useful in helping to understand the timing and nature of the evolutionary events that have led to the diversification of the *Brassica* genus and serves as an important resource for *Brassica* vegetable and oilseed crop breeding.

---

**Cacao (Theobroma cacao)**

445Mb, 11 chromosomes, diploid

The plant from which chocolate is produced is native to the Amazon rainforest. Its flowers are pollinated by midges. Inadequate pest control practices (insecticide spraying) decrease pollinator populations and ultimately yield. Cacao is cultivated in over 50 countries. A member of the Malvaceae family, its beans are harvested from pods for use as the food chocolate, in confections and cosmetics. Cacao is a species with a relatively small genome. The study of the genomes of cacao cultivars is helping researchers identify the genes that control the levels of alkaloids and terpenoids that affect its flavour and melting temperature. Genetic tools are also being used to find the parts of the genome that confer resistance to infection by a common disease, water mould *Phytophthora*.
Banana (Musa acuminata)
472Mb, 11 chromosomes, triploid

The banana is one of the world's favourite fruits and vital for food security in many tropical and subtropical countries. Even though there is some variability in bananas, half of the current global production relies on clones (copies) of one cultivar called ‘Cavendish’. This is a problem because its resistance to pests and diseases is very low, resulting in the loss of harvests. In 1950, an outbreak of Panama disease almost wiped out the commercial Gros Michel banana production, the dominant cultivar at the time. There is an urgent need for banana improvement to avoid further outbreaks. However, bananas are difficult to breed: they are triploid (three sets of chromosomes) and cannot produce viable seeds. The study of their genomes is expected to speed up breeding by identifying genes responsible for fruit quality and pest resistance.

Rice (Oryza sativa)
375 Mb, 12 chromosomes, diploid

Oryza sativa subsp. japonica was the first crop to have its genome sequenced. The Rice Annotation Project has been describing the function of rice genes since 2004. Rice is the main source of nutrition for half of the global human population. It is also a model plant used to study other cereals, which together provide about 90% of our calories. Rice can be cultivated under a wide range of environments, from arid highlands to flooded lowlands. For this reason the study of rice genomes is expected to be critical when facing problems associated with climate change and food production. Breeders and researchers are producing varieties with traits such as salt tolerance, pest resistance or higher yield.
What is haploid or diploid?
Genes are grouped together into spaghetti-like structures called chromosomes in the nucleus of the cell. Different organisms have differing numbers of chromosomes. Chromosomes normally occur in pairs - half from one parent and the other from the other parent. Sometimes an organism can have more than two sets of chromosomes and the number of sets of chromosomes is called ploidy. A cell can be haploid (one set of chromosomes - usually the case for sex cells such as sperm and eggs), diploid (two sets - most common situation in animals), triploid (three sets), tetraploid (four sets), and so forth. Plants are quite commonly polyploid meaning they often have more than two sets of chromosomes. This potentially brings advantages to these species such as an improved ability to respond to environmental change.

EMBL’s European Bioinformatics Institute (EMBL-EBI)
In addition to hosting a genomes database, EMBL-EBI which is based in Cambridge helps scientists exploit other complex information to make discoveries that benefit humankind.

Why does mint feel cold, or chilli feel hot? How can a delicious tomato be a relative of deadly nightshade? Plants make complex chemicals to defend themselves, colour themselves and even communicate. ChEMBL is a database that stores a lot of data on plant chemicals and brings this together with genomic data to help with the development of new drugs.

See an online version of this trail at www.botanic.cam.ac.uk/dnatrail