Accession An accession is a group of related plant material from a single species which is collected at one time from a specific location.

Agouti gene is a gene that controls the distribution of the natural pigment, melanin, in the hair of mammals and helps determine their coat color patterns.

Alignment In bioinformatics **sequence alignment** is a way of arranging the sequences of DNA, RNA or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences.

Allogamy Pollination with pollen from one plant on the style from another plant (cross pollination, cross fertilization)**.**

Allohexaploid Having six complete sets of chromosomes derived from different species.

Allopolyploid (amphiploid) An individual whose chromosomes are composed of more than two genomes each of which has been derived more or less complete but possibly modified from one of two or more species.

Amplicon In molecular biology, an amplicon is a piece of DNA or RNA that is the source and/or product of amplification or replication events. It can be formed artificially, using various methods including polymerase chain reactions (PCR).

Annotation DNA annotation or genome annotation is the process of identifying the locations of genes and all of the coding regions in a genome and determining what those genes do.

Assembly DNA sequence assembly is a process through which short DNA sequence fragments (called reads or samples) are merged into a longer DNA sequence in the attempt to reconstruct the original DNA sequence.

Association mapping Association mapping (genetics), also known as "linkage disequilibrium mapping", is a method of mapping quantitative trait loci (QTLs) that takes advantage of historic linkage disequilibrium to link phenotypes (observable characteristics) to genotypes (the genetic constitution of organisms), uncovering genetic associations.

Autogamy Pollination with pollen from the same plant (self-pollination, self-fertilization).

Backcrossed hybrids are sometimes described with acronym "BC", for example, an F1 hybrid crossed with one of its parents (or a genetically similar individual) can be termed a BC1 hybrid, and a further cross of the BC1 hybrid to the same parent (or a genetically similar individual) produces a BC2 hybrid.

BACs A bacterial artificial chromosome is a DNA construct, based on a functional fertility plasmid (or F-plasmid), used for transforming and cloning in bacteria, usually *E. coli*. BACs were often used to sequence the genome of organisms in genome projects, for example the Human Genome Project. A short piece of the organism's DNA is inserted in BACs, and then sequenced.

BC1 Backcrossing is a crossing of a hybrid with one of its parents or an individual genetically similar to its parent, in order to achieve offspring with a genetic identity which is closer to that of

the parent. It is used in horticulture, animal breeding and in production of gene knockout organisms.

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

Botrytis cinerea is a necrotrophic fungus that affects many plant species, although its most notable hosts may be wine grapes.

Breeders Rights (Plant Variety Rights) Plant breeders' rights (PBR), also known as plant variety rights (PVR), are rights granted to the breeder of a new variety of plant that give the breeder exclusive control over the propagating material (including seed, cuttings, divisions, tissue culture) and harvested material (cut flowers, fruit, foliage) of a new variety for a number of years.

Cas9 (CRISPR associated protein 9) is an RNA-guided DNA endonuclease enzyme associated with the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) adaptive immunity system in *Streptococcus pyogenes*, among other bacteria

Centimorgan is a map unit used to express the distance between two gene loci on a chromosome. A spacing of one centimorgan indicates a one per cent chance that two genes will be separated by crossing over.

Chromosome a threadlike structure of nucleic acids and protein found in the nucleus of most living cells, carrying genetic information in the form of genes.

Cisgenesis A product designation for a category of genetically engineered plants. A variety of classification schemes have been proposed that order genetically modified organisms based on the nature of introduced genotypic changes, rather than the process of genetic engineering.

Clavibacter michiganensis* ssp *michiganensis *A*n aerobic non-sporulating gram-positive plant pathogenic actinomycete that currently constitutes the only species within the genus *Clavibacter*.

Clonal multiplication reproduces clones, which contain all the genetic information of the parent tree. The term clone is used to mean a genetically uniform plant material derived from a single individual and propagated exclusively by vegetative means.

Comparative genomics A field of biological research in which the genomic features of different organisms are compared.

Complementary DNA (cDNA) In genetics, **cDNA** is DNA synthesized from a single-stranded RNA (e.g., messenger RNA (mRNA) or microRNA) template in a reaction catalyzed by the enzyme reverse transcriptase.

Complexity reduction of genome an efficient method that can be used to simultaneously amplify a set of genetic loci across a genome with high reliability can provide a valuable tool for large-scale SNP genotyping studies.

Contig a set of gel readings that are related to one another by overlap of their sequences. All gel readings belong to one and only one contig, and each contig contains at least one gel reading. The gel readings in a contig can be summed to form a contiguous consensus sequence and the

length of this sequence is the length of the contig. Contig is also used to refer to a contiguous sequence in a genome assembly.

CRISPR (**clustered regularly interspaced short palindromic repeats**) is a family of DNA sequences found within the genomes of prokaryotic organisms such as bacteria and archaea.

De novo sequencing is the initial generation of the primary genetic sequence of a particular organism.

Dihaploid plants are derived from tetraploid crop plants may be important for breeding programs that involve diploid wild relatives of the crops.

Dioecious a plant or invertebrate animal having the male and female reproductive organs in separate individuals.

Diploid a cell or nucleus containing two complete sets of chromosomes, one from each parent.

Directed Breeding Selective breeding (also called artificial selection) is the process by which humans use animal breeding and plant breeding to selectively develop particular phenotypic traits (characteristics) by choosing which typically animal or plant males and females will sexually reproduce and have offspring together.

DNA methylation is a process by which methyl groups are added to the DNA molecule. Methylation can change the activity of a DNA segment without changing the sequence.

DNA-sequence is the process of determining the nucleic acid sequence – the order of nucleotides in DNA. It includes any method or technology that is used to determine the order of the four bases: adenine, guanine, cytosine, and thymine.

Dominant allele results in the same phenotype whether its paired allele is identical or different.

Doubled haploids A doubled haploid (DH) is a genotype formed when haploid cells undergo chromosome doubling. Artificial production of doubled haploids is important in plant breeding.

Emasculation (removal of anthers) is generally used to promote cross pollination in plants and avoid self-pollination. This is done to achieve the beneficial variation which are not established due to inbreeding by self-pollination.

Endophyte(s) are endosymbionts, often a bacterium or fungus, who lives within a plant for at least part of its life cycle without causing apparent disease.

Epialleles A specific DNA methylation pattern of a genetic locus.

Epigenetics the study of changes in organisms caused by modification of gene expression rather than alteration of the genetic code itself.

Exon is any part of a gene that will encode a part of the final mature RNA produced by that gene after introns have been removed by RNA splicing. The term *exon* refers to both the DNA sequence within a gene and to the corresponding sequence in RNA transcripts.

F1 Hybrid (also known as filial 1 hybrid) is the first filial generation of offspring of distinctly different parental types. F1 hybrids are used in genetics. The term is sometimes written with a subscript, as F_1 hybrid. Subsequent generations are called **F_2**, **F_3**, etc.

fitTetra is an R package for genotype calling of tetraploid samples based on bi-allelic marker assays such as Illumina GoldenGate, Infinium and KASPar.

Functional studies are often used broadly to refer to the many technical approaches to study an organism's genes and proteins, including the "biochemical, cellular, and/or physiological properties of each and every gene product"

Gametes are mature haploid male or female germ cells that are able to unite with another of the opposite sex in sexual reproduction to form a zygote.

Gametophytic self-incompatibility is a general name for several genetic mechanisms in angiosperms, which prevent self-fertilization and thus encourage outcross and allogamy. It should not be confused with genetically controlled physical or temporal mechanisms that prevent self-pollination

GC-MS is an analytical method that combines the features of gas-chromatography and mass spectrometry to identify different substances within a test sample

Gene is a sequence of nucleotides in DNA or RNA that codes for a molecule that has a function.

Genetic distance is a measure of the genetic divergence between species or between populations within a species, whether the distance measures time from common ancestor or degree of differentiation

Genetic gain is the amount of increase in performance that is achieved through artificial genetic improvement programs.

Genetic map distance (cM) Recombination between linked genes can be used to map their distance apart on the chromosome. The unit of mapping (1 cM) is defined as a recombinant frequency of 1 percent.

Genome editing is the manipulation of the genetic material of a living organism by deleting, replacing, or inserting a DNA sequence, typically with the aim of improving a crop or farmed animal, or correcting a genetic disorder.

Genomic Breeding is the development and application of scientific methods, procedures, and technologies that permit direct manipulation of genetic material in order to alter the hereditary traits of a cell, organism, or population.

Genomic Estimated Breeding Value, GEBV In animal breeding value is its genetic merit, half of which will be passed on to its progeny. These estimates are called Estimated Breeding Values (EBVs).

Genomic selection (GS) is a method to predict the genetic value of selection candidates based on the genomic estimated breeding value (GEBV) predicted from high-density markers positioned throughout the genome.

Genomics is the branch of molecular biology concerned with the structure, function, evolution, and mapping of genomes.

Genotyping by sequencing also called GBS, is a method to discover single nucleotide polymorphisms (SNP) in order to perform genotyping studies, such as genome-wide association studies (GWAS). GBS uses restriction enzymes to reduce genome complexity and genotype multiple DNA samples. After digestion, PCR is performed to increase fragments pool and then GBS libraries are sequenced using next generation sequencing technologies, usually resulting in about 100bp single-end reads.

Genotyping is the process of determining differences in the genetic make-up (genotype) of an individual by examining the individual's DNA sequence using biological assays and comparing it to another individual's sequence or a reference sequence. It reveals the alleles an individual has inherited from their parents.

GFP = Green Fluorescent Protein is a protein composed of 238 amino acid residues (26.9 kDa) that exhibits bright green fluorescence when exposed to light in the blue to ultraviolet range.

Heterosis hybrid vigor, or outbreeding enhancement, is the improved or increased function of any biological quality in a hybrid offspring.

Heterozygosity is the possession of two different alleles of a particular gene or genes by an individual.

HiSeq (Illumina) is a powerful high-throughput sequencing system that enables large-scale genomics for a broad range of applications and study sizes.

Hot spot of recombination Recombination hotspots are regions in a genome that exhibit elevated rates of recombination relative to a neutral expectation. The recombination rate within hotspots can be hundreds of times that of the surrounding region. Recombination hotspots result from higher DNA break formation in these regions.

Hybrid potato seeds. Most potato varieties are cultivated from seed potatoes. This means each tuber is a clone of another tuber, which makes them genetically identical. A small portion is available as seed, which is known as a true potato seed (TPS) variety.

Hybrid purity Hybrid seeds are developed by the hybridization or crossing of two parent lines that are 'pure lines' produced through inbreeding. By crossing pure lines, a uniform population of F1 hybrid seed can be produced with predictable characteristics. F1 hybrid seeds can be tested on the percentage of unwanted self-pollination or unwanted cross-pollination.

Hydroponics the growing of plants in nutrient solutions with or without an inert medium (such as soil) to provide mechanical support.

Intragenesis was developed as alternative to transgenesis. Intragenesis implies that plants must only be changed with genetic material derived from the species itself.

Introgression Lines (IL) in genetics is the movement of a gene (gene flow) from one species into the gene pool of another by the repeated backcrossing of an interspecific hybrid with one of its parent species. All introgression lines together should represent the whole genome of the donor species.

Intron is any nucleotide sequence within a gene that is removed by RNA splicing during maturation of the final RNA product. The word *intron* is derived from the term *intragenic region*, i.e. a region inside a gene.

Junk DNA DNA that does not code for a protein, usually occurs in repetitive sequences of nucleotides, and does not seem to serve any useful purpose.

KASP technology, Kompetitive Allele Specific Polymorphism is a homogenous, fluorescence-based genotyping variant of polymerase chain reaction. It is based on allele-specific oligo extension and fluorescence resonance energy transfer for signal generation.

Kruskal Wallis analysis or one-way ANOVA on ranks is a non-parametric method for testing whether samples originate from the same distribution. It is used for comparing two or more independent samples of equal or different sample sizes.

Linkage analysis is the coupling of two genes' patterns of inheritance because they are located on the same chromosome.

Linkage disequilibrium is the non-random association of alleles at different loci in a given population. Loci are said to be in linkage disequilibrium when the frequency of association of their different alleles is higher or lower than what would be expected if the loci were independent and associated randomly.

Linkage drag Genetic linkage is the tendency of DNA sequences that are close together on a chromosome to be inherited together during the meiosis phase of sexual reproduction. Two genetic markers that are physically near to each other are unlikely to be separated onto different chromatids during chromosomal crossover, and are therefore said to be more *linked* than markers that are far apart. In other words, the nearer two genes are on a chromosome, the lower the chance of recombination between them, and the more likely they are to be inherited together. Markers on different chromosomes are perfectly *unlinked*.

Linkage map A linear map of the relative positions of genes along a chromosome. Distances are established by linkage analysis, which determines the frequency at which two gene loci become separated during chromosomal recombination.

Locus location(s) on the chromosome (region).

LOD value LOD stands for "logarithm of the odds." In genetics, the LOD score is a statistical estimate of whether two genes, or a gene and a disease gene, are likely to be located near each other on a chromosome and are therefore likely to be inherited. A LOD score of 3 or higher is generally understood to mean that two genes are located close to each other on the chromosome. In terms of significance, a LOD score of 3 means the odds are a thousand to one that the two genes are linked, and therefore inherited together.

Markers A genetic marker is a gene or DNA sequence with a known location on a chromosome that can be used to identify individuals or species. It can be described as a variation (which may arise due to mutation or alteration in the genomic loci) that can be observed.

MAS refers to the use of DNA markers that are tightly-linked to target loci as a substitute for or to assist phenotypic screening.

Meiosis a type of cell division that results in four daughter cells each with half the number of chromosomes of the parent cell, as in the production of gametes and plant spores.

Metabolite a substance formed in or necessary for metabolism.

Metabolomics the scientific study of the set of metabolites present within an organism, cell, or tissue.

Methylation (DNA methylation) is the main way gene activity is adjusted during life, especially during early development. It is a process by which methyl groups are added to DNA. This suppresses gene transcription. Two of DNA's four nucleotides, cytosine and adenine, can be methylated.

Monoecious A monoecious plant is a plant where both the male and female reproductive systems exist on the same plant.

Multiple allelism Three or more alternative forms of a gene (alleles) that can occupy the same locus. However, only two of the alleles can be present in a single diploid organism.

Mutation breeding sometimes referred to as "variation breeding", is the process of exposing seeds to chemicals or radiation in order to generate mutants with desirable traits to be bred with other cultivars. Plants created using mutagenesis are sometimes called mutagenic plants or mutagenic seeds.

Nanopore sequencing is a third generation approach used in the sequencing of biopolymers-specifically, polynucleotides in the form of DNA or RNA. Using nanopore sequencing, a single molecule of DNA or RNA can be sequenced without the need for PCR amplification or chemical labeling of the sample.

Next generation **sequencing** refers to non-Sanger-based high-throughput DNA sequencing technologies. Millions or billions of DNA strands can be sequenced in parallel, yielding substantially more throughput and minimizing the need for the fragment-cloning methods that are often used in Sanger sequencing of genomes.

Nitrogen use efficiency (NUE): the efficiency with which soil nitrate-N is converted into grain N. The nitrate-N comes from fertiliser, crop residues, manures, and soil organic matter, but it is the efficiency of conversion of fertiliser into grain that is generally of greatest concern to growers.

PacBio Sequencing with long-read sequencing, enabled by Single Molecule, Real-Time (SMRT) Sequencing.

Parthenocarpy is the development of the ovary into a seedless fruit.

PCR Polymerase chain reaction is a method widely used in molecular biology to make many copies of a specific DNA segment. Using PCR, a single copy (or more) of a DNA sequence is exponentially amplified to generate thousands to millions of more copies of that particular DNA segment.

Phenomics is an area of biology concerned with the measurement of phenomes (a phenome is the set of physical and biochemical traits belonging to a given organism) as they change in response to genetic mutation and environmental influences.

Phenotyping The observable physical or biochemical characteristics of an organism, as determined by both genetic makeup and environmental influences.

Physical distance A process called recombination or 'crossing over'. Physical mapping looks at the physical distance between known DNA sequences (including genes) by working out the number of base pairs (A-T, C-G) between them.

Polygenic A trait that is controlled by a group of nonallelic genes (called polygene) Supplement. Polygenic traits are controlled by two or more than two genes (usually by many different genes) at different loci on different chromosomes. These genes are described as polygenes.

Polymorphic A gene is said to be polymorphic if more than one allele occupies that gene's locus within a population. A polymorphic variant of a gene can lead to the abnormal expression or to the production of an abnormal form of the protein; this abnormality may cause or be associated with disease.

Polyploids A cell or an organism having a genome with multiple (more than two) sets of homologous chromosomes. Having multiple (more than two) sets of homologous chromosomes that make up the genome of a cell or an organism.

Positional sterility means that pollen are produced but not released.

Protandry. The condition of flowers whose male parts mature before the female ones.

Proteomics branch of biotechnology concerned with applying the techniques of molecular biology, biochemistry, and genetics to analyzing the structure, function, and interactions of the proteins produced by the genes of a particular cell, tissue, or organism, with organizing the information in databases.

Protogyny The condition of flowers whose female parts mature before the male ones.

Quantitative Trait Locus (Loci) is a region of DNA which is associated with a particular phenotypic trait, which varies in degree and which can be attributed to polygenic effects, i.e., the product of two or more genes, and their environment. ... Moreover, a single phenotypic trait is usually determined by many genes.

Recombinant Inbred Lines (RILs) recombinant inbred line (RIL) population is developed using single seed descent from the F_2 generation. The result is a set of homogeneous, homozygous lines for which large amounts of seed can be produced for replicated trials. This type of population is often useful for mapping QTLs.

Recurrent parent Backcross breeding enables breeders to transfer a desired trait such as a transgene from one variety (donor parent, DP) into the favored genetic background of another (recurrent parent, RP). If the trait of interest is produced by a dominant gene, this process involves four rounds of backcrossing within seven seasons.

Resequencing The sequencing of part of an individual's genome in order to detect sequence differences between the individual and the reference genome of the species.

Restriction sites, or restriction recognition sites, are locations on a DNA molecule containing specific (4-8 base pairs in length) sequences of nucleotides, which are recognized by restriction enzymes.

Resistance genes (*R* genes) are genes in plant genomes that convey plant disease resistance against pathogens by producing R proteins.

RIL A recombinant inbred line population is developed using single seed descent from the F_2 generation. The result is a set of homogeneous, homozygous lines for which large amounts of seed can be produced for replicated trials.

RNA-Seq (**RNA sequencing**), also called whole transcriptome shotgun sequencing, uses next-generation sequencing (NGS) to reveal the presence and quantity of RNA in a biological sample at a given moment.

S. arcanum *wild tomato species*
S. chacoense *wild potato species*
S. pimpinellifolium *wild tomato species*

Scaffolds are portions of the genome sequence reconstructed from end-sequenced whole-genome shotgun clones. Scaffolds are composed of contigs and gaps. A contig is a contiguous length of genomic sequence in which the order of bases is known to a high confidence level.

Self-pollination the pollination of a flower by pollen from the same flower or from another flower on the same plant.

Self-incompatibility (**SI**) is a general name for several genetic mechanisms in angiosperms, which prevent self-fertilization and thus encourage outcross and allogamy. It should not be confused with genetically controlled physical or temporal mechanisms that prevent self-pollination, such as heterostyly and sequential hermaphroditism (dichogamy).

SeqSNP is a targeted genotyping by sequencing service.

Sequence reads give the length of a certain fragment.

Sequence-specific nuclease (SSN) A nuclease (also archaically known as nucleodepolymerase or polynucleotidase) is an enzyme capable of cleaving the phosphodiester bonds between monomers of nucleic acids. Nucleases variously effect single and double stranded breaks in their target molecules.

SgRNA Short-Guide RNAs (sgRNA) are designed to combine crRNA (CRISPR RNA) and tracrRNA (trans-activating CRISPR RNA) into a longer RNA strand.

Simple Sequence Repeat (SSR) A microsatellite or SSR is a tract of repetitive DNA in which certain DNA motifs (ranging in length from 1–6 or more base pairs) are repeated, typically 5–50 times. Microsatellites occur at thousands of locations within an organism's genome. They have a higher mutation rate than other areas of DNA leading to high genetic diversity. Microsatellites are often referred to as short tandem repeats (STRs) by forensic geneticists and in genetic genealogy, or as simple sequence repeats (SSRs) by plant geneticists.

SNP genotyping is the measurement of genetic variations of single nucleotide polymorphisms (SNPs) between members of a species.

SOLiD sequencing, (Sequencing by Oligonucleotide Ligation and Detection) is a next-generation DNA sequencing technology developed by Life Technologies and has been commercially available since 2006. This next generation technology generates hundreds of millions to billions of small sequence reads at one time.

RNAses Ribonuclease (commonly abbreviated RNase) is a type of nuclease that catalyzes the degradation of RNA into smaller components.

Susceptibility genes are host genes needed for functioning of the pathogene. Mutations in susceptibility genes influence the resistance level of the plant.

Synthetic biology is an interdisciplinary branch of biology and engineering. The subject combines disciplines from within these domains, such as biotechnology, genetic engineering, molecular biology, molecular engineering, systems biology, membrane science, biophysics, chemical and biological engineering, electrical and computer engineering, control engineering and evolutionary biology. Synthetic biology applies these disciplines to build artificial biological systems for research, engineering and medical applications.

Systems biology is the computational and mathematical modeling of complex biological systems. It is a biology-based interdisciplinary field of study that focuses on complex interactions within biological systems, using a holistic approach (holism instead of the more traditional reductionism) to biological research.

TAqMan probes are hydrolysis probes that are designed to increase the specificity of quantitative PCR.

Taxonomic studies is the science of defining and naming groups of biological organisms on the basis of shared characteristics.

Tillering A **tiller** is a stem produced by grass plants, and refers to all shoots that grow after the initial parent shoot grows from a seed. Tillers are segmented, each segment possessing its own two-part leaf. They are involved in vegetative propagation and, in some cases, also seed production.

Tobacco Rattle Virus (TRV) The virus causes the plant disease tobacco rattle in many plants, including many ornamental flowers including *Narcissus*. It causes the disease corky ringspot in potatoes. The disease manifests in various ways, and signs can include brown rings and arcs on the surface of a potato, and discolored spots on the interior.
Training populations are formed that is composed of plant lines covering all of the important material in the breeding program in question. The training population is then genotyped and phenotyped for all traits of interest to the breeding program.

Transcriptomics Study of all RNA molecules in one cell or a population of cells.

Turgor is the state of turgidity and resulting rigidity of cells or tissues, typically due to the absorption of fluid.

Volatile compounds (**VOCs**) are organic chemicals that have a high vapor pressure at ordinary room temperature. Their high vapor pressure results from a low boiling point, which causes large numbers of molecules to evaporate or sublimate from the liquid or solid form of the compound and enter the surrounding air, a trait known as volatility.