

of the data by participating in temporary, large genome sequencing networks. These would complement the principal dedicated sequencing centres, which will have to undertake the vast bulk of the work involved in the large eukaryotic genomes. The effective application of the network approach in developing, as well as developed, nations shows that the model is widely (and perhaps universally) applicable, therefore significantly extending the opportunity to be involved in practical genomics. The available data indicate that the use of sequencing networks is accelerating; this in itself could represent an important alteration not only in the content of molecular biology, but also in the way that it is executed.

Andrew J. G. Simpson is at the Laboratory of Cancer Genetics, Ludwig Institute for Cancer Research, Rua Professor Antônio Prudente, 109, 4th floor, 01509-010, São Paulo, SP, Brazil. e-mail: asimpson@ludwig.org.br

1. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
2. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546–563–546, 567 (1996).
3. Favello, A., Hillier, L. & Wilson, R. K. Genomic DNA sequencing methods. *Methods Cell Biol.* **48**, 551–569 (1995).
4. Fraser, C. M. & Fleischmann, R. D. Strategies for whole microbial genome sequencing and analysis. *Electrophoresis* **18**, 1207–1216 (1997).
5. Stein, L. Genome annotation: from sequence to biology. *Nature Rev. Genet.* **2**, 493–503 (2001).
6. Ridgeway, S. & Reddy, M. Residue levels of several organochlorines in tursiops-truncatus milk collected at varied stages of lactation. *Mar. Pollut. Bull.* **30**, 609–614 (1995).
7. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).
8. Chambaud, I. *et al.* The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*. *Nucleic Acids Res.* **29**, 2145–2153 (2001).
9. Andersson, S. G. *et al.* The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140 (1998).
10. Kunst, F. *et al.* The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–256 (1997).
11. Simpson, A. J. *et al.* The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature* **406**, 151–157 (2000).
12. Dias, N. E. *et al.* Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc. Natl Acad. Sci. USA* **97**, 3491–3496 (2000).
13. Bonalume Neto, R. Brazilian scientists team up for cancer genome project. *Nature* **398**, 450 (1999).
14. Camargo, A. A. *et al.* The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome. *Proc. Natl Acad. Sci. USA* **98**, 12103–12108 (2001).
15. Simpson, A. J. & Perez, J. F. ONSA, the São Paulo Virtual Genomics Institute. Organization for Nucleotide Sequencing and Analysis. *Nature Biotechnol.* **16**, 795–796 (1998).
16. Shirai, M. *et al.* Comparison of whole genome sequences of *Chlamydia pneumoniae* J138 from Japan and CWL029 from USA. *Nucleic Acids Res.* **28**, 2311–2314 (2000).
17. Hayashi, T. *et al.* Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* **8**, 11–22 (2001).
18. Ng, W. V. *et al.* Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl Acad. Sci. USA* **97**, 12176–12181 (2000).
19. Kuroda, M. *et al.* Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet* **357**, 1225–1240 (2001).

20. She, Q. *et al.* The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl Acad. Sci. USA* **98**, 7835–7840 (2001).

Acknowledgements

I thank the hundreds of scientists, students and associates who have made genomics such a telling reality in Brazil. In addition, I thank F. Perez (Scientific Director of FAPESP), L. Old (Scientific Director of the Ludwig Institute) and R. Brentani (Director of the São Paulo Branch of the Ludwig Institute) for their continuing support and encouragement, and without whom network-based genome sequencing would not have been established in the state of São Paulo. I am also grateful to O. Caballero and A. Vettore for their critical reading of this manuscript. The sequencing of *X.fastidiosa*, *X.citri*, *X.campestris* and *L.xyli* has been funded principally by FAPESP, and the national genome project to sequence *C.violaceum* has been coordinated and funded exclusively by the Ministério de Ciência e Tecnologia (MCT) in conjunction with Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

OPINION

Improving plant breeding with exotic genetic libraries

Dani Zamir

Naturally occurring variation among wild relatives of cultivated crops is an under-exploited resource in plant breeding. Here, I argue that exotic libraries, which consist of marker-defined genomic regions taken from wild species and introgressed onto the background of elite crop lines, provide plant breeders with an important opportunity to improve the agricultural performance of modern crop varieties. These libraries can also act as reagents for the discovery and characterization of genes that underlie traits of agricultural value.

Every plant now under cultivation was once wild. Plant evolution under domestication has led to increased productivity, but, at the same time, domestication has narrowed the genetic basis of crop species. The challenges that face modern plant breeders are to develop higher yielding, nutritious and environmentally friendly varieties that improve our quality of life without harnessing additional natural habitats to agricultural production. This article asserts that it is time to return to the wild ancestors of crop plants to use them as a source of the genetic variation that has been lost during domestication. Exotic germplasm resources, which include wild species and LANDRACES, often carry many agriculturally undesirable alleles. However, genetic studies can identify the agriculturally valuable traits of wild species, and INTROGRESSION BREEDING can transfer these traits to commercial varieties.

Online links

FURTHER INFORMATION

TIGR Comprehensive Microbial Resource:
<http://www.tigr.org/tigr-scripts/CMR2/CMRHomePage.spl>

Dedicated sequencing centres

Genoscope: <http://www.genoscope.cns.fr/>

Kasuzo DNA Research Institute:

<http://www.kazusa.or.jp/en/>

TIGR: <http://www.tigr.org/>

The Sanger Centre: <http://www.sanger.ac.uk/>

Sequencing consortia

Brazilian genome virtual institute of genomic research:

<http://www.brgene.incc.br/>

ONSA: <http://www.watson.fapesp.br/genoma3.htm>

Access to this interactive links box is free online.

To enhance the rate of progress of introgression breeding, I believe that we should invest in the development of a genetic infrastructure of 'exotic libraries'. Each line in this library would contain a marker-defined chromosomal segment from an agriculturally unadapted source crossed onto a background of an ELITE VARIETY; each exotic genome would be represented in its entirety in a set of introgression lines. Once such permanent exotic resources, in the form of seed banks, have been generated from a diverse selection of ACCESSIONS, they can be screened for multiple phenotypes to identify alleles of economic importance. In this article, I present the rationale for developing exotic libraries from the perspective of their practical application to plant breeding.

Art and science of plant breeding

Plant breeding is the art and science of the genetic improvement of crops to produce new varieties that have increased productivity and quality (FIG. 1). Genetic variation is the engine that propels breeding to meet future challenges. We select, from the available pool of genetic variation, individual plants that contain desirable traits, which are then crossed to create a new pool of variation that is enriched for the selected phenotypes. The variation at our disposal exists in modern varieties, in exotic germplasm and in genetically modified (GM) plants. The success of a breeder depends on his or her ability to select a few

PERSPECTIVES

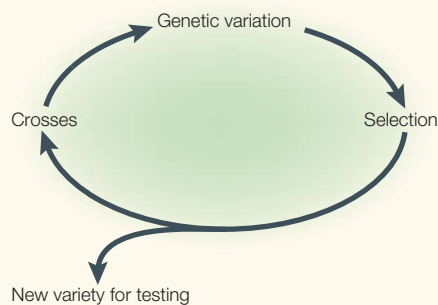


Figure 1 | The central dogma of plant breeding. The crop improvement cycle is repeated for several generations until the breeder is satisfied with the product and releases a new variety for testing. In modern agriculture, hundreds of varieties are released each year and very few of these end up in the market place.

individual lines that have a superior phenotype and to discard others in a way that allows large-scale testing efforts to be centred on a significantly improved gene pool. It is important to note that although scientific technologies can greatly assist in the development of new varieties, the key components in plant breeding are new ideas combined with a strong element of intuition. For example, in the 1960s, plant breeders had the idea and the germplasm to develop high-yielding dwarf varieties of wheat in an agricultural advance that is now known as the ‘Green Revolution’¹. Four decades later, plant research has shown that the mutant phenotypes behind the Green Revolution were caused by alterations in genes that encode plant hormone response modulators². So, plant breeders use biological phenomena, the scientific basis of which is often unknown, such as HETEROISIS, which is used to create hybrid plants of improved quality.

Crop domestication

The breeders who made the first marked advances in crop productivity worked 10,000 years ago in the Fertile Crescent of the Middle East, and domesticated several cereal and pulse crops³. Early breeders selected out the rare mutations of the native flora — such as recessive alleles for non-brittle spike in barley — that prevented seed dispersal at maturity, so facilitating the harvest. As a result, crop populations became dependent on human intervention for their reproduction⁴. On the basis of archaeological and genetic evidence, it seems that domestication of most crop plants occurred in specific ‘centres of origin’ throughout the world, and has generally involved only a few founding genotypes. The ‘founder effect’ principle in crop

evolution is responsible for the fact that many crop plants contain only a small fraction of the genetic variation that is present in their wild relatives. Watermelon domestication is a good example of this. The wild watermelon species harbour dominant genes that are responsible for an extremely bitter taste and white fruit flesh (FIG. 2). Early farmers probably selected rare recessive mutants with non-bitter attributes, as well as red flesh, and all the watermelons we eat today originate from this early selection. This is supported by genetic evidence showing that very little polymorphism exists at molecular markers in the genome of the cultivated watermelon, compared with the amount of variation that is present in the entire genus⁵.

Over the centuries, farmers have knowingly selected improved plant types in their fields, which have arisen through recombination, naturally occurring mutations and spontaneous outcrossing events with wild relatives⁶. This slow breeding process led to the development of landraces that are adapted to local field conditions. With the formulation of the Mendelian principles of heredity, plant breeders made much more rapid progress by making directed crosses and by selecting for the desired combinations of parental characteristics. The Green Revolution developed high-yielding, short, inbred varieties of wheat and rice for growing in tropical regions, which provided LODGING tolerance under high fertilization regimes, so increasing yield¹. The development of hybrids also had an important impact on yield increases of maize and revealed heterosis as a force in plant breeding. These breeding efforts, although aimed at fulfilling the demands of intensive agriculture, resulted in further erosion of the genetic diversity of elite gene pools. Furthermore, many landraces were lost as farmers throughout the world shifted to growing high-yielding varieties.

“Plant breeding is the art and science of the genetic improvement of crops to produce new varieties ... Genetic variation is the engine that propels breeding to meet future challenges.”

Wild species in breeding

The potential of wild species as a source of genetic variation to bring about crop improvement was recognized early in the twentieth century⁷. However, the use of these exotic genetic resources in breeding programmes was a time-consuming and laborious process that often ended in failure. The transfer of traits from poorly adapted germplasm that carries many undesirable genes into elite lines required many backcrosses, an efficient selection procedure and much luck. The problems that stood, and still stand, in the way of progress when crossing wild and domesticated species include: cross incompatibility between the wild species and the cultivated crop; F_1 hybrid sterility; infertility of the segregating generations; reduced recombination between the chromosomes of the two species; and genes of negative effect being tightly linked to the trait of interest (so-called ‘linkage drag’). Despite these difficulties, when we evaluate the contribution that wild introgression breeding has made to the development of modern-day varieties, it is evident that this approach has had an important effect on the development of some crop species, particularly with respect to developing new varieties with improved disease resistance. Furthermore, although exotic germplasm has been exploited as a source of monogenic traits, little has been done with respect to traits that are influenced by QUANTITATIVE TRAIT LOCI (QTL). Traits such as yield, quality and stress response show complex inheritance patterns that result from the segregation of numerous interacting QTL, the expression of which is modified by the environment.

In the following paragraphs, I highlight some representative examples of the use of seed bank collections of wild germplasm in the breeding of agriculturally important crops. This is not an exhaustive review of all such cases, but is meant to provide some insight into how the evolution, the history of domestication, and the reproductive patterns of certain crop species can inform us as to how useful it might or might not be to exploit exotic germplasm to improve the agricultural performance of crop species.

Wheat. The history of wheat cultivation and that of human civilization are closely allied. Wheat breeders knew how to benefit from phenotypes identified in wild introgressions for disease resistance (they discovered close to 30 independent genes⁸), and have produced a few of the rare examples of the agricultural implementation of QTL, which were found in

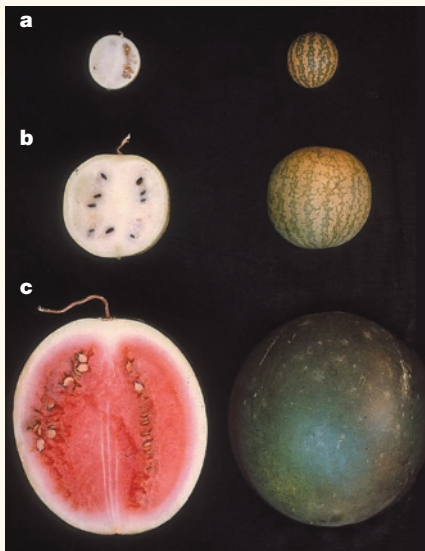


Figure 2 | Wild, hybrid and cultivated watermelons. **a** | The bitter fruits of the wild watermelon species *Citrullus colocynthis*, and **b** | the F_1 hybrid progeny from crosses between *C. colocynthis* and the cultivated watermelon (*Citrullus lanatus*, shown in **c**).

exotic resources of wheat and which increased the yield of cultivated wheat varieties. For example, independent lines with the short arm of chromosome 1B of wheat (*Triticum aestivum*) replaced by the homologous arm from rye (*Secale cereale*), produce higher yields both in optimal and stress environments⁹. Yield improvements have also been associated with an *Agropyron elongatum* (a tall wheat grass) chromosomal segment that carries a rust resistance gene¹⁰ (*Lr19*), and with a high-grain protein QTL from *Triticum dicoccoides* (wild emmer wheat) that improves the quality of pasta made with flour from wheat that carries the QTL¹¹.

Tomato. New-world founder cultivars of tomato were brought to Europe in the sixteenth century and, after a history of gardening and selection, were disseminated to many areas of the world. This pattern of domestication is responsible for the very narrow genetic basis of the cultivated crop and underlies the early interest in incorporating traits from wild species into modern varieties. At present, commercial hybrids include different combinations of 15 independently introgressed, disease-resistance genes that originate from various wild resources¹². These genes were introduced through recurrent backcrossing, and each resides on a small independent chromosome segment from one of the diverse donor species. This situation facilitated the development of polymorphic molecular markers that are being used to bring about

the PYRAMIDING of multiple traits on a single genetic background. The tomato has also been a pioneer crop in the directed introgression of wild chromosome segments that are associated with QTL that improve fruit quality¹³. An important gene that was introduced from the wild tomato species *Lycopersicon pennellii* is *B*, which increases the level of provitamin A (β -carotene) in the fruit by more than 15-fold¹⁴.

Rice. Owing to the narrow genetic basis of cultivated rice compared with its wild relatives, more than 20,000 wild rice accessions are stored in seed banks awaiting use¹⁵. Traits for resistance to more than seven pathogens have been introgressed into rice germplasm from wild species, and some introgressed lines are in commercial cultivation¹⁶. The mapping of yield-associated QTL in interspecific populations has indicated that wild rice species harbour genes of interest for the improvement of rice cultivation in diverse environments¹⁷. The rice genome will soon be completely sequenced, and mapping information for traits of breeding value will facilitate the development of informative markers. These markers should help to eliminate linkage drag and to isolate the genes that are responsible for selected phenotypes¹⁸.

Soybean. Despite the extremely narrow basis of modern soybean varieties, which trace back to a handful of introductions from China, there are no wild introgressions present in the leading varieties cultivated in the United States. However, interesting traits have been identified in wild species that are held in seed banks, including many disease-resistance traits, and traits that confer tolerance to herbicides and salinity, and increased protein levels. Some of this new variation is being introduced through interspecific hybridization^{19,20}. It is possible that more-advanced soybean varieties, into which exotic genes have been introduced, are being developed in breeding companies, which usually keep such information confidential.

Cotton. *Gossypium hirsutum*, the most important cotton species, was domesticated within the last 5,000 years for its seed fibre. Modern cotton varieties trace back to a few Mexican lines, and their genetic diversity is therefore very low^{21,22}. Despite the fact that many wild cotton species show agronomic traits of potential interest, very little use has been made of these exotic resources in breeding. The only documented case is the use of *Gossypium barbadense*, which apparently

contributed some fibre-quality traits to present-day varieties²³. As with soybean and maize, transgenic cotton varieties with insect and herbicide resistances are in widespread cultivation.

Maize. The direct ancestor of maize, which was domesticated in Central America, is teosinte. It is estimated that five mutations are responsible for the principal differences between maize and teosinte²⁴. A detailed analysis of one of the mutations, *teosinte branched 1* (*tb1*), showed that both maize and teosinte are highly polymorphic; however, the upstream region of *tb1* in maize was fixed through selection²⁵. This indicates that the mutation that was selected during the process of domestication might alter *tb1* regulation. Maize is unique because it is a naturally outcrossing species that has undergone more gene flow between cultivated and wild species than have inbred varieties of the aforementioned crop species. This has resulted in a highly polymorphic genome in which the value of exotic resources has yet to be explored.

In summary, in contrast to the breeding of wheat, tomato and rice, very few wild-species introgressions are present in modern cultivated soybean and cotton varieties. However, the fact that many breeding studies in the field have indicated that exotic germplasm contains useful genes that can be practically exploited, has stimulated intensive genetic research in a wide range of crop species²⁶. Even in maize, crosses to teosinte populations produce progeny with traits of some agricultural potential²⁷.

Exotic libraries

The observation that wild genetic resources can contribute to crop improvement, combined with the alarming rate at which locally adapted landraces are being lost and at which natural habitats are being damaged, has led to the establishment of large germplasm collections in the form of seed banks. The curators of these seed banks initiate collection missions, maintain and characterize accessions, and make them available to the breeding community. The task facing us is to devise the tools and concepts that would allow us to rapidly use the genetic potential that exists in these seed banks and in wild species^{28,29}. This could be achieved by developing a permanent seed resource of introgressed lines that would facilitate the rapid screening of traits that are harboured in wild genomes. This resource would overcome a limitation of early wild germplasm experiments — that they did not create a

PERSPECTIVES

lasting germplasm resource. At present, if a breeder wants to re-screen the progeny of a particular interspecific cross that has been explored in the past, for a new trait, he or she has to start from the parents and then develop the required generations — a time-consuming undertaking.

With the development of informative genetic markers and high-density marker maps, it became possible to map not only monogenic traits, but also QTL that affect phenotypic variation. Numerous studies

have resolved major and minor QTL in populations that segregate for the entire genome — in F_2 , backcross, recombinant inbred and advanced backcross populations³⁰. The problem with the genetic analysis of such populations, which result from crosses between divergent genomes, is that the plants contain a large proportion of wild-germplasm-derived genes, which often results in partial sterility. For this reason, these populations are generally not suitable for the identification of QTL that improve agricultural yield.

An exotic library consists of a set of lines, each of which carries a single, defined chromosome segment that originates from a donor species in an otherwise uniform elite genetic background. The production of such a congenic resource takes about ten generations, and each line can be used directly for breeding (BOX 1a). These lines would provide an efficient tool for detecting and mapping valuable agronomic traits for several reasons. Because lines in the library differ from the elite variety by only a single,

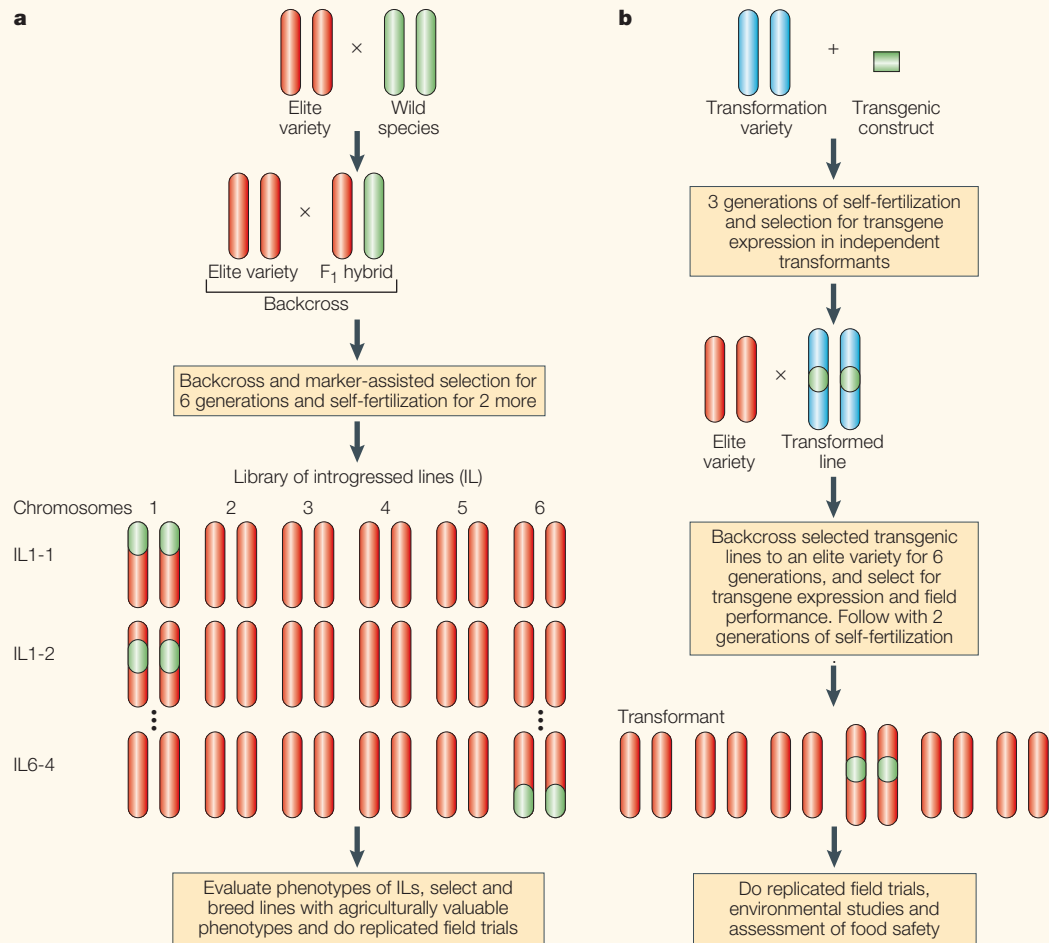
Box 1 | Exotic and genetic modification breeding

The figures show breeding schemes for generating and screening an exotic genetic library (a) and a transgenic variety (b). In both schemes, the elite variety chromosomes are shown in red and the wild-species introgressions or transgene are shown in green. Chromosomes of the transformation variety are shown in blue. The haploid chromosome number of the plant is six, and backcross generations are shown for only a pair of homologues.

In panel a, the wild species (green) is crossed as a male parent to a leading cultivated variety (red), and the F_1 hybrid is backcrossed to the elite parent. Through recurrent backcrosses, the average proportion of the wild species genome is reduced by 50% in each generation. Chromosome segments are traced through crosses by genotyping the lines with a genome-wide panel of polymorphic markers that can distinguish between parental alleles. By the sixth backcross generation, independent plants can be isolated, each heterozygous for a different

segment of the wild-species genome. Further self-pollination and selection leads to homozygosity at targeted introgressed segments. The resulting exotic library, shown for the first two lines (IL1-1 and IL1-2), and for the last line (IL6-4), can be used in many screens for different agriculturally important traits. Typically, three to five generations are required to identify wild-chromosome segments that improve the phenotype of the elite line and to eliminate the negative traits that are caused by linkage drag. Once a particular exotic library has been developed, it can be used to rapidly identify other traits of agricultural importance.

In the transgenic approach shown in panel b, intensive basic research is first required to identify new genes of potential benefit that have not been patented. Once this is achieved, the gene is introduced onto a genetic background that is most suitable for transformation. Many plants are usually evaluated for appropriate transgene expression levels in the tissue or organ of interest; selected transformants have to be tested over several generations, as transgene silencing is a common problem for some crops and when using some types of gene-transfer technique. When the desired expression level is obtained, the transformed line is backcrossed to an elite variety, and classical breeding is used to select the optimal background for expression of the transgene and the associated phenotype. At this stage, the transgenic variety can be introduced and tested, although product-marketing rights need to be applied for through the legal framework that regulates genetically engineered foods^{44,45}. Once such a transformant has been produced, it represents a variety that has been improved for only one trait.



defined chromosomal segment, the resulting plants will generally resemble the cultivated variety, which reduces the sterility problems that occur in other breeding-population structures and allows progeny to be assayed for yield-associated traits. The EPISTATIC effects that are mediated by other regions of the exotic genome are also removed in introgressed lines because all the phenotypic variation between a line in the library and the nearly isogenic cultivated variety is associated with the introgressed segment. As a result, the ability to statistically identify small phenotypic effects is increased. Because exotic libraries provide a permanent resource with a characterized genotype, they can be tested by several research groups over time, and the phenotypic data that arise from these studies can be collected and stored in curated databases for general access. Furthermore, homozygous lines in a library can be crossed to different tester lines, which allows the effects of heterozygosity on the phenotype to be explored, and which might lead to the identification of the chromosome segments that are associated with heterosis. This resource would also allow the phenotypic effects of QTL to be mapped to smaller intervals, through the recombination-mediated reduction in size of QTL-carrying segments into smaller, overlapping regions that could be used to overcome the effects of linkage drag. Finally, once introgressed chromosome segments have been sub-divided and targeted, and QTL-containing lines have been created, crosses between the lines can be used to examine the phenotypic effects of QTL interactions, to better understand the nature of epistasis³¹.

One of the earliest uses of such a library was that by Kuspira and Unrau³², who analysed complex traits in common wheat using whole-chromosome substitution lines. In tomato, RFLP (restriction fragment length polymorphism) markers have been used to develop a full-coverage exotic library in the form of introgression lines from a cross between the wild green-fruited species *L. pennellii* and the cultivated tomato *Lycopersicon esculentum* cv. M82 (REF. 33 and FIG. 3). This population allowed us to identify yield-associated QTL, and to examine their epistatic and environmental interactions. It also allowed us to eventually map, to high resolution, a QTL to a region that spans a single exon and an intron³⁴. The development of such populations³⁵ was a difficult task in the early days, when molecular markers were still being developed. However, now, with the availability of a range of marker-screening

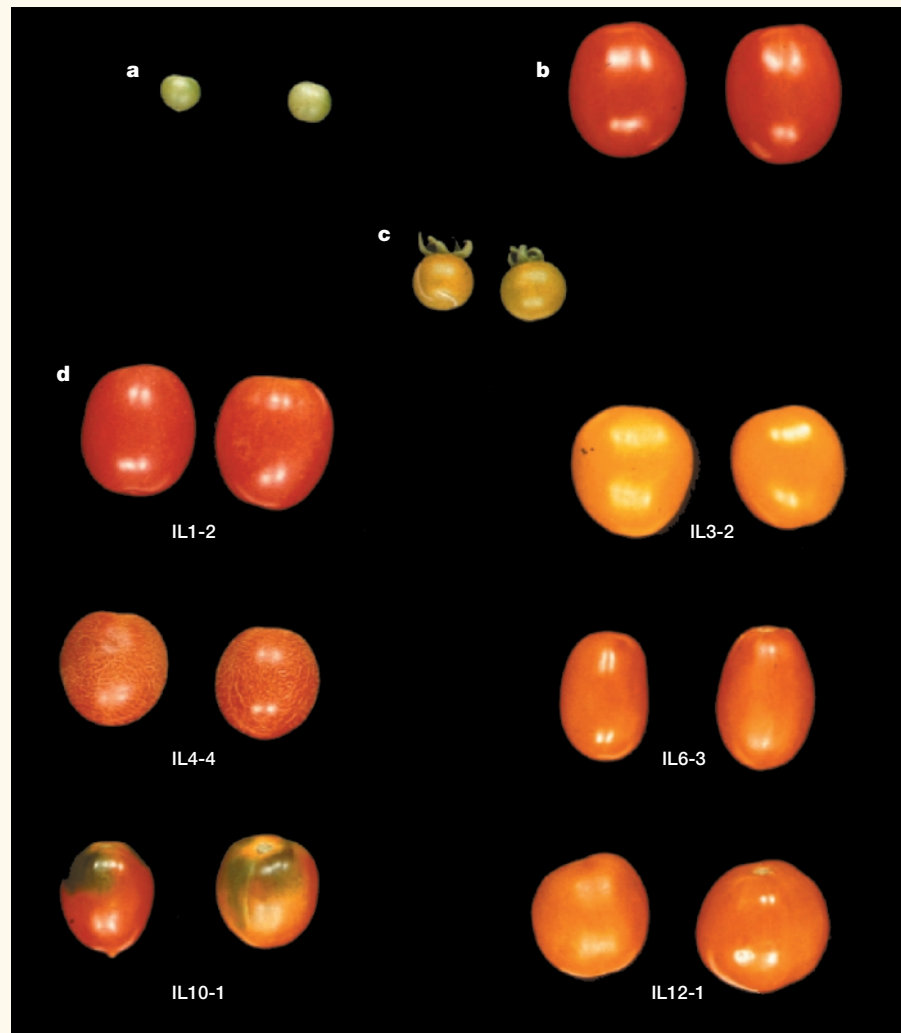


Figure 3 | **The *Lycopersicon pennellii* library.** **a** | Green fruits of the wild species *Lycopersicon pennellii*, **b** | the lycopene-rich red fruits of *Lycopersicon esculentum*, **c** | their F_1 hybrid progeny, and **d** | six introgression lines (ILs) that show their different fruit colour and carotenoid-content phenotypes. Each line contains a single, marker-defined, wild-species-derived chromosome segment that harbours a gene that affects fruit phenotype.

technologies, such libraries can be developed after ten generations of crossing and marker analysis³⁶.

Phenotyping is the rate-limiting step in large-scale germplasm-enhancement programmes. However, library resources can be rapidly phenotyped for a range of yield-associated traits, including biotic and abiotic stresses, as well as for the metabolic profiles of hundreds of distinct compounds³⁷. The establishment of a user-friendly bioinformatics management system is an essential component of a programme that aims to handle this type of complex biological data. The challenge for the coming years is to develop methodologies that will enable genomic information to be associated with phenotypes of interest for crop improvement. The

framework for this data organization could be a genetic map that links all available biological data to a plant locus in the form of chromosomal addresses. In plants, the use of comparative mapping has shown unexpectedly high levels of synteny conservation in grasses, Solanaceae (the family to which tomatoes belong), crucifers and legume crops³⁸. Maps that show regions of conserved synteny provide the means to compare phenotypes between species that cannot be crossed, and to establish more comprehensive and useful breeding databases for linking exotic traits³⁹. Naturally selected genetic variation provides the opportunity to explore developmental networks that lead to higher yields, some of which are controlled at the gene-regulation level^{14,25}.

The key proposal of this paper is to strengthen the resources of the research community, which is positioned between the seed banks and the commercial plant breeders, so that this community can bring about germplasm enhancement through the development of exotic introgression lines. The wild accessions to be included in such an effort will depend on the specific evolution, reproductive patterns and breeding strategies of each crop, and should include representatives of the spectrum of variation that can be crossed to a cultivated parent. Establishing such resources will require investing a considerable amount of time and money into developing the many introgressed lines and into an integrated bioinformatics framework that can collate and make available the phenotypic and mapping information generated from these lines. Investment will also need to be made into complex, high-throughput, marker-assisted selection technologies³⁶. However, such resources could facilitate the rapid application of new genetic variation in breeding.

Glossary

ACCESSION

A sample of plant material that is collected at a specific location and maintained in a seed bank.

ELITE VARIETY

A variety that excels under conditions of modern intensive agriculture.

EPISTASIS

An interaction between non-allelic genes, such that one gene masks, interferes with or enhances the expression of the other gene.

HETEROSIS

Hybrid vigour that leads to superior crop varieties.

INTROGRESSION BREEDING

The incorporation of selected traits from an unadapted exotic resource through a succession of crosses (backcrosses) to a commercially elite variety.

LANDRACE

A locally adapted, cultivated variety that is selected by farmers.

LODGING

The collapse of top-heavy plants, particularly grain crops.

PYRAMIDING

The accumulation of several independent traits in the same genotype through introgression breeding.

QUANTITATIVE TRAIT LOCI

(QTL). Genetic loci that are identified through the statistical analysis of complex traits (such as plant height or body weight). Quantitative traits are typically affected by more than one gene and by the environment.

“Considering the problems of consumer acceptance of GM products ... we should apply GM technologies in plant breeding only in cases in which we have no other classical genetic alternative.”

Genetically modified plants

GM technologies have been highlighted as the most significant, recent development in plant breeding, because they allow plant genomes to be enriched with the DNA of any organism, introducing new variation into the gene pool. As transgenic technology becomes an integral part of plant breeding, it is important to define the traits that have been targeted for introduction and the time that it takes to develop a transgenic product⁴⁰ (BOX 1b). Typically, ten generations are required to create a transgenic variety for testing, which is similar to the time frame for developing and testing a line developed through exotic breeding. It is also important to note that once an exotic library has been generated, the time required for its use in breeding for other traits is considerably shortened.

The identification of a natural insect toxin in the bacterium *Bacillus thuringiensis* made it possible to introduce a new insect tolerance trait into crop plants, which has significantly reduced the amount of insecticide released into the environment. Resistance to the herbicide glyphosate, which was introduced into plants from *Agrobacterium tumefaciens* — a bacterium that causes crown gall disease in some plants — provides an efficient method of weed control and therefore increases productivity. These GM plants are widely cultivated, and the expectation is that the rate at which unique traits are introduced, particularly those that benefit the end consumer, will increase in the next century⁴¹. One of the first GM varieties to enter the market was the Flavr Savr tomato, which was modified to soften at a slower rate than normal tomatoes, allowing it to be picked at a later stage of maturity, when the flavour of the tomato was more developed. Flavr Savr tomatoes were created using antisense technology that inhibits the expression of the enzyme polygalacturonase, which is associated with cell wall softening. At the same time, tomato varieties that were heterozygous for the naturally occurring, single-gene mutation *rin* (*ripening inhibitor*) took the tomato market by storm

and still dominate it, whereas Flavr Savr now has little-to-no market share. This is because the effect of *rin* on extending shelf life is much more significant. This and other examples indicate that some of the variation that is introduced into plants by transgenesis to suppress certain gene products occurs naturally or is induced in mutants, and can be exploited in a cost-effective manner. Considering the problems of consumer acceptance of GM products, the take-home message from the above cases is that we should apply GM technologies in plant breeding only in cases in which we have no other classical genetic alternative.

Conclusion

Although animals and plants are highly evolutionarily divergent, the strategies used to dissect complex traits in them are often quite similar. Owing to favourable biological attributes (such as short generation times and tolerance to inbreeding), new concepts in breeding are often first proposed and tested in plants. So, it is possible that the exotic library approach might also be useful in breeding agriculturally important animals by providing the means to identify potentially orthologous natural alleles across diverse genotypes.

This paper challenges the view that we are living in a time of decreasing opportunities for crop improvement through the use of classical breeding approaches^{42,43}. Exotic libraries can make a wide array of previously unexplored genetic variation rapidly available to plant breeders and geneticists. Either in combination with GM technology or without it, exotic genetic libraries represent a dynamic new resource that could substantially enrich traditional crop improvement programmes for many years to come.

Dani Zamir is at the Otto Warburg Center for Agricultural Biotechnology, Faculty of Agriculture, The Hebrew University of Jerusalem, PO Box 12, Rehovot 76100, Israel. e-mail: zamir@agri.huji.ac.il

1. Khush, G. S. Green Revolution: the way forward. *Nature Rev. Genet.* **2**, 815–822 (2001).
2. Peng, J. *et al.* ‘Green Revolution’ genes encode mutant gibberellin response modulators. *Nature* **400**, 256–261 (1999).
3. Lev-Yadun, S., Gopher, A. & Abbo, S. Archaeology. The cradle of agriculture. *Science* **288**, 1602–1603 (2000).
4. Ladizinsky, G. *Plant Evolution under Domestication* (Kluwer Academic, Dordrecht, The Netherlands, 1998).
5. Jarret, R. L. & Newman, M. Phylogenetic relationships among species of *Citrullus* and the placement of *C. rehmii* De Winter as determined by internal transcribed spacer (ITS) sequence heterogeneity. *Genet. Res. Crop Evol.* **47**, 215–222 (2000).
6. Koornneef, M. & Stam, P. Changing paradigms in plant breeding. *Plant Physiol.* **125**, 156–159 (2001).
7. Bessey, C. E. Crop improvement by utilizing wild species. *Am. Breed. Assoc.* **11**, 112–118 (1906).
8. Fedak, G. Molecular aids for integration of alien chromatin through wide crosses. *Genome* **42**, 584–591 (1999).

9. Villareal, R. L., Del Toro, E., Mujeeb-Kazi, A. & Rajaram, S. The 1BL/1RS chromosome translocation effect on yield characteristics in a *Triticum aestivum* L. cross. *Plant Breed.* **114**, 497–500 (1995).
10. Hoisington, D. *et al.* Plant genetic resources: what can they contribute toward increased crop productivity? *Proc. Natl Acad. Sci. USA* **96**, 5937–5943 (1999).
11. Kovacs, M. I. P., Howes, N. K., Clarke, J. M. & Leisle, D. Quality characteristics of durum wheat lines deriving high protein from *Triticum dicoccoides* (6b) substitution. *J. Cereal Sci.* **27**, 47–51 (1998).
12. Pan, Q. *et al.* Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: tomato and *Arabidopsis*. *Genetics* **155**, 309–322 (2000).
13. Rick, C. M. High soluble-solids content in large-fruited tomato lines derived from a wild green-fruited species. *Hilgardia* **42**, 493–510 (1974).
14. Ronen, G., Carmel-Goren, L., Zamir, D. & Hirschberg, J. An alternative pathway to β -carotene formation in plant chromoplasts discovered by map-based cloning of β - and old-gold color mutations in tomato. *Proc. Natl Acad. Sci. USA* **97**, 11102–11107 (2000).
15. Plunknett, D. L., Smith, N. J. H., Williams, J. T. & Murthi-Anishetty, N. *Gene Banks and the World's Food* (Princeton Univ. Press, Princeton, New Jersey, 1987).
16. Brar, D. S. & Khush, G. S. Alien introgression in rice. *Plant Mol. Biol.* **35**, 35–47 (1997).
17. Moncada, P. *et al.* Quantitative trait loci for yield and yield components in an *Oryza sativa* x *O. rufipogon* BC2F2 population evaluated in an upland environment. *Theor. Appl. Genet.* **102**, 41–52 (2001).
18. Yano, M. Genetic and molecular dissection of naturally occurring variation. *Curr. Opin. Plant Biol.* **4**, 130–135 (2001).
19. Singh, R. J. & Hymowitz, T. Soybean genetic resources and crop improvement. *Genome* **42**, 605–616 (1999).
20. Sebolt, A. M., Shoemaker, R. C. & Diers, B. W. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Sci.* **40**, 1438–1444 (2000).
21. Small, R. L., Ryburn, J. A. & Wendel, J. F. Low levels of nucleotide diversity at homoeologous *Adh* loci in allotetraploid cotton (*Gossypium* L.). *Mol. Biol. Evol.* **16**, 491–501 (1999).
22. Iqbal, M. J., Reddy, O. U. K., El-Zik, K. M. & Pepper, A. E. A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theor. Appl. Genet.* **103**, 547–554 (2001).
23. McCarty, J. C. & Percy, R. G. in *Genetic Improvement of Cotton: Emerging Technologies* (eds Jenkins, J. N. & Saha, S.) 65–79 (USDA — Agricultural Research Service: Science, Enfield, New Hampshire, 2001).
24. Doebley, J. & Stec, A. Inheritance of the morphological differences between maize and teosinte: comparison of results for two F_2 populations. *Genetics* **134**, 559–570 (1993).
25. Wang, R. L., Stec, A., Hey, J., Lukens, L. & Doebley, J. The limits of selection during maize domestication. *Nature* **398**, 236–239 (1999).
26. Mauricio, R. Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nature Rev. Genet.* **2**, 370–381 (2001).
27. Ray, J. D., Kindiger, B. & Sinclair, T. R. Introgressing root aerenchyma into maize. *Maydica* **44**, 113–117 (1999).
28. Tanksley, S. D. & McCouch, S. R. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**, 1063–1066 (1997).
29. Lee, M. Genome projects and gene pools: new germplasm for plant breeding? *Proc. Natl Acad. Sci. USA* **95**, 2001–2004 (1998).
30. Fulton, T. M. *et al.* Advanced backcross QTL analysis of a *Lycopersicon esculentum* x *L. parviflorum* cross. *Theor. Appl. Genet.* **100**, 1025–1042 (2000).
31. Eshed, Y. & Zamir, D. Less than additive epistatic interactions of QTL in tomato. *Genetics* **143**, 1807–1817 (1996).
32. Kuspura, J. & Unrau, J. Genetic analysis of certain characters in common wheat using all chromosome substitution lines. *Can. J. Plant Sci.* **37**, 300–326 (1957).
33. Eshed, Y. & Zamir, D. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* **141**, 1147–1162 (1995).
34. Fridman, E., Pleban, T. & Zamir, D. A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. *Proc. Natl Acad. Sci. USA* **97**, 4718–4723 (2000).
35. Ramsay, L. D. *et al.* The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative trait loci. *Genome* **39**, 558–567 (1996).
36. Young, N. D. A cautiously optimistic vision for marker-assisted breeding. *Mol. Breed.* **5**, 505–510 (1999).
37. Roessner, U. *et al.* Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **13**, 11–29 (2001).
38. Gale, M. D. & Devos, K. M. Plant comparative genetics after 10 years. *Science* **282**, 656–659 (1998).
39. Patterson, A. H. *et al.* Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* **269**, 1714–1717 (1995).
40. Zhong, G. Y. Genetic issues and pitfalls in transgenic plant breeding. *Euphytica* **118**, 137–144 (2001).
41. Uzogara, S. G. The impact of genetic modification of human foods in the 21st century: a review. *Biotechnol. Adv.* **18**, 179–206 (2000).
42. Mann, C. C. Crop scientists seek a new revolution. *Science* **283**, 310–314 (1999).
43. Academy, N. *Transgenic Plants and World Agriculture* (National Academy Press, Washington DC, 2000).
44. Mitten, D. H., MacDonald, R. & Klonus, D. Regulation of foods derived from genetically engineered crops. *Curr. Opin. Biotechnol.* **10**, 298–302 (1999).
45. Kuiper, H. A., Kleter, G. A., Noteborn, H. P. J. M. & Kok, E. J. Assessment of food safety issues related to genetically modified foods. *Plant J.* **27**, 503–528 (2001).

Acknowledgements

I thank E. Fridman, Y. Eshed and S. Abbo for helpful discussions. This work was supported by the United States–Israel Binational Research and Development Fund (BARD).

 Online links

DATABASES

The following terms in this article are linked online to: **Maize Database:** <http://www.agron.missouri.edu/browser.html> *tb1*

FURTHER INFORMATION

C. M. Rick Tomato Genetics Resource Center (TGRC):

<http://tgrc.ucdavis.edu/>

Consultative group on international agricultural research:

<http://www.cgiar.org/>

Encyclopedia of Life Sciences: <http://www.els.net/>

Plant breeding and crop improvement

***L. pennellii* tomato introgression lines:**

<http://tgrc.ucdavis.edu/>

Maize genome database: <http://www.agron.missouri.edu/>

Map of *L. pennellii* introgression lines:

<http://soldb.cit.cornell.edu/>

Rice genome research program: <http://rgp.dna.affrc.go.jp/>

The Solanaceae Genome Network:

<http://soldb.cit.cornell.edu/>

Access to this interactive links box is free online.