CORRESPONDENCE

Hervé Vanderschuren

Department of Biology, Plant Biotechnology, ETH Zurich, Zurich, Switzerland.
e-mail: hvanderschuren@ethz.ch

1. Ezezika, O.C. *et al. Nat. Biotechnol.* **30**, 38–40 (2012).
2. Turok, N. *Nature* **474**, 567–569 (2011).
3. Juma, C. *Science* **334**, 1323 (2011).
4. Vasil, I.K. *Nat. Biotechnol.* **14**, 702–703 (1996).
5. Tripathi, L., Mwaka, H., Tripathi, J.N. & Tushemereirwe, W.K. *Mol. Plant Pathol.* **11**, 721–731 (2010).
6. Shepherd, D.N. *et al. Plant Biotechnol. J.* **5**, 759–767 (2007).
7. Bull, S.E. *et al. Nat. Protoc.* **4**, 1845–1854 (2009).
8. Bull, S.E., Ndunguru, J., Gruissem, W., Beeching, J.R. & Vanderschuren, H. *Plant Cell Rep.* **30**, 779–787 (2011).
9. Chetty, C.C., Rossin, C.B., Gruissem, W., Vanderschuren, H. & Rey, M.E.C. *New Biotechnol.* published online, doi:10.1016/j.nbt.2012.04.006 (6 June 2012).
10. Vanderschuren, H., Moreno, I., Anjanappa, R.B., Zainuddin, I. & Gruissem, W. *PLoS ONE* **7**, e45277 (2012).

Can genomics boost productivity of orphan crops?

To the Editor:

Advances in genomics over the past 20 years have enhanced the precision and efficiency of breeding programs[1] in many temperate cereal crops[2,3]. One of the first applications of genomics-assisted breeding has been the introgression of loci for resistance to biotic stresses or major quantitative trait loci (QTLs) for tolerance to abiotic stresses into elite genotypes through marker-assisted backcrossing (MABC)[4]. For instance, introgression of a major QTL for submergence tolerance (*Sub1*) into widely grown rice varieties has substantially improved yield in >15 million hectares of rain-fed low-land rice in South and Southeast Asia[5]. Despite this success story, the overall adoption of genomics-assisted breeding in developing countries is still limited especially for complex traits like yield under environmental stress in several other crops[6,7].

Although maize, rice and wheat dominate global food production, several other crops are of great importance for some communities in developing countries (**Supplementary Table 1**). This group includes sorghum and millets, groundnut, cowpea, common bean, chickpea, pigeonpea, cassava, yam and sweet potato (**Table 1**). As they are not extensively traded and receive little attention from researchers compared to the main crops, these important crops for marginal environments of Africa, Asia and South America are often referred to as 'orphan crops'. Breeding for orphan crops is lagging behind major crops although they are key staple crops in many low-income countries where small-holder farmers cannot afford to buy improved seed. The magnitude of the breeding effort for those orphan crops and the capacity of adopting modern technologies is extremely variable across

developing countries and generally directly related to the health of the national economy.

In any crop, many of the key traits affecting crop performance are under complex genetic control and show quantitative variation[8]. These traits are of growing importance in breeding as favorable alleles for most simply inherited traits are already prevalent and often fixed in elite germplasm, suggesting that genetic gain for such traits may be difficult to achieve. The overall objective for most breeders is to improve crop productivity in a target environment. Yield is a direct reflection of biomass or the proportion of biomass that is converted to the harvestable commodity. Yield is a result of the integration and interaction of many physiological and metabolic processes over time and in environments that are increasingly variable due to climate change[9].

The prediction of phenotype on the basis of genotypic composition is challenging, as crop performance can be profoundly influenced by weather conditions, soil composition, pathogens and pests and trial management (e.g., fertilizer input, weed control, water supply). Interaction of genotype with the environment can be evaluated through suitable experimental design, multiple site evaluation and careful measurement of the environmental variables. Accurate prediction of plant phenotype from genotype through genomics-assisted breeding is exacerbated when landraces or wild germplasm, representing different gene pools, are used as sources of favorable alleles for target traits. These sources have been important in improving disease and pest resistance traits, some traits influencing nutritional and sensory quality traits, and in some cases polygenic traits such as yield[10]. Such

success, however, is yet to be achieved in routine practicethe in the above mentioned orphan crops.

Here we provide a short overview of both the potential and the challenges in implementing genomics-assisted breeding in orphan crops. Additionally, we present a critical appraisal of the application of association mapping or genome-wide association studies (GWAS)[11], marker-assisted recurrent selection (MARS)[7] and genomic selection (GS)[12,13] in improving grain yield of orphan crops in developing countries.

Until recently, only major commercial crop species have benefited from the application of next-generation sequencing and high-throughput tools in molecular biology. Today, however, the use of such high-throughput technologies to assay molecular markers and transcript sequences, genome structural variation, gene space or even genome sequences is now feasible for orphan crops[14–16]. The low sequencing cost has motivated Elshire *et al.*[17] to develop a robust, cost-effective, highly multiplexed sequencing approach known as 'genotyping by sequencing', which is expected to replace high-density-marker genotyping (**Supplementary Fig. 1**). Importantly, these new sequencing technologies for characterizing germplasm collections will help overcome 'ascertainment bias', which has hampered genotypic evaluation of diverse collections (**Supplementary Fig. 2**).

One of the prerequisites of molecular breeding has been the identification of molecular markers associated with traits of interest[18]. Linkage mapping to detect QTLs generally focuses on the analysis of families from crosses between two inbred lines. Although the statistical power for QTL detection is high, the genetic resolution of these QTLs is poor. Conversely, GWAS or association mapping takes advantage of historic recombinations that reveal chromosomal regions of linkage disequilibrium (LD) where markers remain associated with traits of interest over many generations. GWAS, widely used in human genetics, is being increasingly adopted in crop species like maize[19–23] and rice[24–27]. The best features of linkage and association approaches can be combined through nested association mapping populations[28,29] that enable high-power and high-resolution analysis through joint linkage mapping and association mapping. However, many breeders feel that the association mapping approach has failed to deliver beneficial alleles or haplotypes for crop improvement.

Table 1 Details on production of three world's major food crops and selected orphan crop species

Crop name	Botanical name	Chromosome number (*n*), ploidy level (*x*), (genome size in Mbp)	Major producing countries[a] (production in million tons (MT))				Total world production (MT)
			1st	2nd	3rd	4th	
World's major crops							
Maize	*Zea mays*	$2n = 2x = 20$ (2,500)	USA (333.01)	China (164.10)	Brazil (51.23)	Mexico (20.14)	818.82
Wheat	*Triticum aestivum*	$2n = 6x = 42$ (16,000)	China (115.11)	India (80.68)	Russian Federation (61.73)	USA (60.31)	685.61
Rice	*Oryza sativa*	$2n = 2x = 24$ (430)	China (196.68)	India (133.70)	Indonesia (64.39)	Bangladesh (47.72)	685.24
Main orphan crops of developing countries							
Cassava	*Manihot esculenta*	$2n = 2x = 36$ (760)	Nigeria (36.80)	Thailand (30.08)	Brazil (24.40)	Indonesia (22.03)	233.79
Sweet potato	*Ipomoea batatas*	$2n = 6x = 90$ (2,200-3,000)	China (76.77)	Uganda (2.76)	Nigeria (2.74)	Indonesia (2.05)	102.29
Coconut	*Cocos nucifera*	$2n = 2x = 32$ (3,600)	Indonesia (21.56)	Philippines (15.66)	India (10.14)	Sri Lanka (2.09)	61.7
Sorghum	*Sorghum bicolor*	$2n = 2x = 20$ (750)	USA (9.73)	India (7.25)	Nigeria (5.27)	Sudan (4.19)	56.09
Yam	*Dioscorea alata*	$2n = 2x = 40$ (550)	Nigeria (29.09)	Côte d'Ivoire (6.90)	Ghana (5.77)	Benin (2.37)	49.18
Groundnut	*Arachis hypogaea*	$2n = 4x = 40$ (2,890)	China (14.76)	India (5.51)	Nigeria (2.96)	USA (1.67)	36.45
Millets Pearl millet (*Pennisetum glaucum*), $2n = 2x = 22$ (2,450) Finger millet (*Eleucine coracana*), $2n = 4x = 36$ (3,000) Foxtail millet (*Setaria italica*), $2n = 2x = 18$ (423)			India (8.81)	Nigeria (4.88)	Niger (2.67)	Mali (1.39)	26.7
Common bean	*Phaseolus vulgaris*	$2n = 2x = 22$ (600)	Brazil (3.48)	Myanmar (3.00)	India (2.44)	China (1.49)	20.69
Chickpea	*Cicer arietinum*	$2n = 2x = 16$ (740)	India (7.06)	Pakistan (0.74)	Turkey (0.56)	Australia (0.44)	10.46
Cowpea	*Vigna unguiculata*	$2n = 2x = 22$ (620)	Nigeria (2.36)	Niger (1.55)	Burkina Faso (0.32)	Myanmar (0.18)	5.24
Cacao	*Theobroma cacao*	$2n = 2x = 20$ (430)	Côte d'Ivoire (1.22)	Indonesia (0.80)	Ghana (0.66)	Nigeria (0.37)	4.08
Pigeonpea	*Cajanus cajan*	$2n = 2x = 22$ (858)	India (2.27)	Myanmar (0.76)	Malawi (0.20)	Uganda (0.09)	3.48

[a]Source: http://faostat.fao.org/site/567/default.aspx#ancor (accessed 12 October 2011).

A critical analysis of this table highlights the importance of orphan crops in developing countries. For instance, although rice, maize and wheat remain the leading crops of the world, crops such as cassava are the biggest crops in the developing countries. In countries like Benin, Ghana, Malawi, Nigeria and Uganda, cassava is the most widely produced crop, whereas it is the second biggest crop grown in Côte d'Ivoire, Kenya and India. The case is similar for many other so-called orphan crops in developing countries. Furthermore, molecular breeding is almost an integral part of crop improvement for the world's major crops, whereas, molecular breeding has either just started or is yet to be initiated in the main orphan crops. It is therefore imperative to deploy molecular breeding in crop improvement programs for these orphan crops.

Limited population size, low heritability, lack of candidate genes, low marker density and the difficulty in identifying beneficial alleles are the main limiting factors. Furthermore, variation in phenology greatly reduces the effectiveness of association mapping in singling out the effects of loci that influence yield potential *per se* as opposed to those affecting plant development or phenology. Clearly, many or, indeed, most, functional genes in the genome will contribute directly or indirectly to yield. Therefore, it is unreasonable to expect that any genetic approach will unequivocally resolve the plethora of polymorphisms that contribute directly or indirectly to yield, even with proper experimental design[30] and the use of appropriate GWAS statistical models[31,32].

Cloning of QTLs is becoming increasingly feasible for manipulating quantitative traits by means of marker-assisted selection or genetic engineering[33]. Over 50 major QTLs of agronomic interest have been cloned[34]. QTL cloning has also revealed the important role of noncoding sequences in modulating gene expression[35]. In particular, the availability of gene sequences allows a direct and targeted search of novel alleles, hence expanding the pool of variability available for breeding purposes.

One of the difficulties for developing superior genotypes for complex traits such as 'yield in stressful environments' is that these traits are often controlled by multiple, small-effect QTLs and/or several epistatic QTLs. When selecting for those complex traits, MARS or GS appears to be best suited for stacking beneficial QTL alleles (**Fig. 1**).

MARS[7] allows the improvement of polygenic traits by stacking favorable alleles at a large number (10–40) of the most significant loci involved in the expression of the target traits. Although MARS is being used routinely in private sector breeding programs[3], no published reports that discuss its use by the public sector are available. In GS[12,13] the breeding value of genotypes is calculated by incorporating data for all marker loci. Implementation of GS indicates that the quality of prediction decays with generational distance from the training population[36]. Part of this decay is the result of basic GS models accurately capturing kinship or broad-scale pedigree relatedness, but failing to identify markers in extremely high LD with the causal loci. Such loci are likely to remain informative through many generations, particularly when

Figure 1 Schematic representation of three molecular breeding approaches for crop improvement. (**a**) MABC is the most extensively used molecular breeding approach, largely deployed for introgressing transgene or major loci for resistance to biotic stress, grain quality traits or a major QTL explaining higher phenotypic variance for abiotic stress tolerance. In the absence of a predictive gene-based marker, the first step in this approach is identification of loci based on analysis of genotypic and phenotypic data from populations segregating for the traits of interest. Subsequently, the elite variety lacking the desired allele(s) at the target locus or loci is used for backcrossing with the donor genotype. Molecular markers associated with the QTL are used for screening the backcross populations for identification of the superior lines that possess the favorable QTL allele at each cycle of backcrossing. Depending on the population size and considering one or two target loci, two to three backcrosses are usually sufficient for recovering most of the recipient genome[4]. Progress can be monitored with the help of randomly selected molecular markers for monitoring the background selection. After generating the backcross progenies with the desired genome coverage from the elite genotype, one cycle of selfing is done and homozygous MABC lines are used for selecting the superior lines for field evaluation. (**b**) The MARS approach is deployed to accumulate favorable alleles for 10–40 loci in a set of complex and simple traits including yield. In the first step,

segregating populations are generated by crossing elite varieties that are superior for the targeted traits but presenting favorable alleles at different loci. These populations are genotyped and phenotyped at a suitable inbreeding level in the targeted environments and these data are used for identification of QTLs for the target traits. Looking at allelic complementarity at target loci, selected genotypes are crossed to stack favorable alleles across successive cycles of recurrent selection. After completing two to three recombinant cycles, superior lines selected based on ideal genotypes are self-fertilized for field evaluation. (**c**) The GS approach, rather than relying on mapped loci, uses the breeding values that are calculated based on high-density genotypic data and historical phenotypic data from a 'training population' usually made up of breeding lines. Based on genomic-estimated breeding values (GEBVs), superior lines are selected for new crosses. The progeny lines from these crosses are genotyped and, based on the model developed for calculating GEBVs with the training populations, GEBVs are calculated for the progeny lines. Subsequently, the progeny lines having higher GEBVs are used for the next cycle of crossing. At any point in the cycle, progeny lines with higher GEBVs can be extracted for the field evaluation.
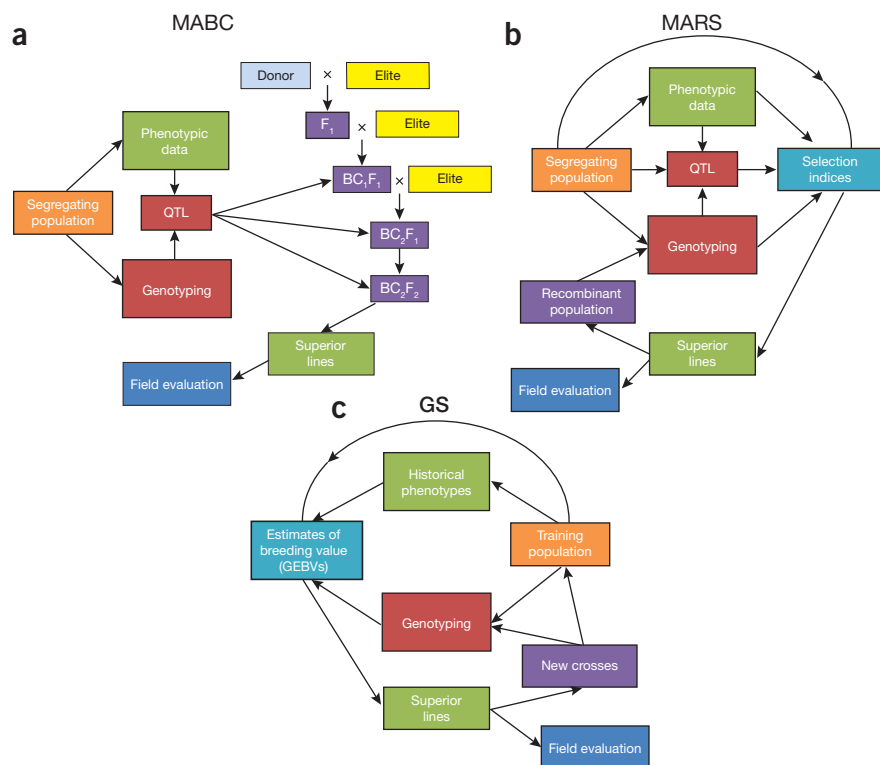
only a moderate number of markers are used. This is exactly what GWAS excels at. The challenge is now to bridge these models and approaches (GWAS and genomic selection) to accelerate genetic gains in orphan crops.

Although several success stories have been published, mainly based on the MABC approach[4], the overall impact of genomics-assisted breeding on crop development programs in developing countries remains very limited[6,7], especially for complex traits. In the short term, the adoption and successful application of this approach outside of developed countries requires the following: first, scientists trained in modern breeding technologies; second, improved local infrastructure capacity for accurate and relevant phenotyping; third, local access to marker technologies for efficient genotyping; and fourth, the deployment of suitable data-management systems (**Box 1**).

High-throughput genotyping and sequencing facilities could benefit from

economies of scale, as large-scale service laboratories are also preferred by large seed companies. With the availability of sequencing service laboratories, the same approach can be adopted for breeding in developing countries, although sample shipment across national borders sometimes poses logistical and quarantine challenges, especially in the developing world, and raises concerns about germplasm protection. Even so, the availability of some medium-scale genotyping and sequencing facilities in developing countries that have stable economies will be useful from a capacity-building perspective, because it will expose scientists to modern technologies while allowing some of the projects to be run locally. Physically accessible facilities are often also important to build confidence by breeders in new technologies.

An essential component of molecular breeding infrastructure is facile access to computational tools used in every step of the process, from sample handling to phenotype

prediction. Some steps of the process require high-power computers and associated services. Such facilities could be shared, and perhaps located in genotyping centers, as long as high bandwidth network access is easily available. Computational tools needed at every breeding station are much more modest and a good laptop and internet access are usually sufficient.

Facilitated access to high-throughput genotyping and sequencing technologies, as mentioned above, is expected to enhance breeding progress in developing countries in at least three ways: first, high-resolution germplasm fingerprinting by low-coverage sequencing or single-nucleotide polymorphism genotyping will facilitate selection of parents for new crosses and introgression of novel alleles from exotic germplasm; second, accurate mapping of loci associated with traits of interest using either bi- or multiparental mapping or GWAS methodology will enhance identification of

Box 1 Strengthening molecular breeding and delivery capacity in developing countries

To succeed, molecular and genomics approaches must build on well-structured breeding programs, which are often not established in developing countries with poor economies, and, therefore, limited infrastructure and human resources are real bottlenecks to the effective use of genomics-assisted breeding approaches in those countries. To overtake those bottlenecks, training in modern plant breeding skills and fostering integrated breeding strategies through systemic collaboration among partners (e.g., national programs and CGIAR centers) and collaborative groups to share knowledge and expertise (e.g., communities of practice) remain a high priority. In addition, access to emerging platforms like the Integrated Breeding Platform (IBP; http://www.integratedbreeding. net/), a one-stop shop for information, analytical tools and related services to design and carry out integrated breeding projects will provide access to information and the latest technologies for integrated breeding in developing countries.

Reliable phenotypic data are a must for high-quality genetic studies[39] and, most importantly, the application of the results toward cultivar improvement. One major, often underestimated aspect that limits the value of the collected data is their relevance to the conditions prevailing in the target environment[40]. An excellent example is provided by the improvement of drought tolerance: in this case, the dynamics of drought (e.g., timing, frequency, intensity and duration of the drought episodes) prevailing in different target environments can vary substantially. Most developing countries lack suitable field infrastructure for collection of accurate and reliable phenotypic data. Uncoordinated efforts among donor agencies, or one–shot funding, also resulted sometimes in buying of sophisticated pieces of equipment that

end up rusting in experimental stations in national programs. Some of the main reasons for this include lack of support to train researchers and/or technicians on the usage and maintenance of those equipments, or just lack of operational support to run the experiments that will require those equipments.

Guidelines on best practice must also be provided on how to design and run a trial and conduct precise phenotyping for genetic studies. This is even more critical when breeding is carried out in marginal environments, as is generally the case in developing countries, where genetic-environment interactions are generally large, thus reducing traits heritability. Under such conditions, selection based on yield only might not always be most efficient and genetic gain might be improved considering secondary traits as proxy to define a selection index (e.g., in MARS experiments). In those instances, secondary traits that are simple, cheap, nondestructive and fast to assay should be genetically highly correlated with grain yield in the target environment. Furthermore, these traits should be genetically variable and stable throughout the measurement period with a high level of heritability and should not be associated with any yield loss under nonstressful conditions.

Governmental and institutional commitment is critical for adoption of biotechnologies in developing countries. Strengthening the breeding and delivery capacity in national programs will also help address several other problems that have hampered biotechnology-based product development and delivery in the past. Of particular importance is ensuring that the technology is delivered to farmers in well-adapted and market-ready lines. Bottlenecks to the efficient delivery of improved varieties, such as weak extension or seed systems, must also be anticipated.

marker-trait associations of interest to the breeders; third, high-throughput genotyping or sequencing approaches will enable genome-assisted breeding, including for orphan crops. All of these advances will require enhanced capacity for precision phenotyping in target environments (**Box 1**). The phenotyping capacity could be enhanced by training and collaborative programs with Consultative Group on International Agricultural Research (CGIAR) institutions and advanced research organizations in tight collaboration with partners from developing countries.

The above approaches have the potential to convert orphan crops into genomic resource–rich crops. Even the definition of an 'orphan crop' is rapidly changing. For instance, a few decades ago access to mutants was critical for genetic studies and peas, maize, barley and tomato were regarded as model crops, whereas rice was considered an orphan crop. In fewer than 20 years, however, rice has moved from being an orphan crop to become a core model for the cereals. Similarly, as a result of investment from several public initiatives, many orphan crops and legumes in particular are presently becoming genomic resource-rich[37]. For example, a genome sequence has

become available for pigeonpea[38], whereas pea, the model legume crop from Mendel's time, is yet to be sequenced.

The remaining 'orphan crops' are thus those that have genomes too complex to sequence or characterize. Complexity can be due to polyploidy or large genome size, as is the case for wheat. In wheat, many of the genomic resources taken for granted in other species are only now being developed (http://www. wheatgenome.org/). Similar problems hamper resource development for sugarcane, onion, faba bean and many other important crops, horticultural and forestry species with large and complex genomes.

The ultimate aim of genomics-assisted breeding is to increase the rate of genetic gain across target environments, in less time and at lower cost compared with conventional selection based exclusively on phenotype. Nonetheless, it is critical to remember that even with the availability of the best genotyping resources, genomics-assisted breeding may not be successful in the absence of quality phenotypic data.

In our opinion, development and availability of genomic resources, due to advances in technology, should not be an issue

anymore in the orphan crops. Centralized service facilities for high-throughput sequencing and genotyping, together with access to genomics and analytical breeding tools, should enhance implementation and adoption of molecular breeding in staple crops in developing countries. Continued training of breeders and geneticists in modern genomics and molecular breeding approaches and their retention in developing countries coupled with adequate institutional and governmental support will be critical for the sustainable and effective integration of genomics-assisted breeding in crop improvement programs for ensuring food security in developing countries.

CORRESPONDENCE

Rajeev K Varshney[1–3], Jean-Marcel Ribaut[2], Edward S Buckler[4,5], Roberto Tuberosa[6], J Antoni Rafalski[7] & Peter Langridge[8]

[1]*Center of Excellence in Genomics (CEG), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India.* [2]*CGIAR Generation Challenge Programme (GCP), c/o International Maize and Wheat Improvement Center (CIMMYT), Mexico, DF, Mexico.* [3]*School of Plant Biology (M084), Faculty of Natural and Agricultural Sciences, The University of Western Australia (UWA), Crawley, Australia.* [4]*United States Department of Agriculture–Agriculture Research Service, Ithaca, New York, USA.* [5]*Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York, USA.* [6]*Department of Agroenvironmental Sciences and Technology, University of Bologna, Bologna, Italy.* [7]*DuPont Pioneer, Wilmington, Delaware, USA.* [8]*Australian Centre for Plant Functional Genomics (ACPFG), University of Adelaide, Adelaide, Australia.*
e-mail: r.k.varshney@cgiar.org

1. Varshney, R.K., Graner, A. & Sorrells, M.E. *Trends Plant Sci.* **10**, 621–630 (2005).
2. Varshney, R.K., Hoisington, D.A. & Tyagi, A.K. *Trends Biotechnol.* **24**, 490–499 (2006).
3. Eathington, S.R., Crosbie, T.M., Edwards, M.D., Reiter, R.S. & Bull, J.K. *Crop Sci.* **47**, 154–163 (2007).
4. Ribaut, J.M. & Hoisington, D. *Trends Plant Sci.* **3**, 236–239 (1998).
5. Septiningsih, E.M. *et al. Ann. Bot. (Lond.)* **103**, 151–160 (2008).
6. Tester, M. & Langridge, P. *Science* **327**, 818–822 (2010).
7. Ribaut, J.M., de Vicente, M.C. & Delannay, X. *Curr. Opin. Plant Biol.* **13**, 213–218 (2010).
8. Collins, N.C., Tardieu, F. & Tuberosa, R. *Plant Physiol.* **147**, 469–486 (2008).
9. Varshney, R.K., Bansal, K.C., Aggarwal, P.K., Datta, S.K. & Craufurd, P.K. *Trends Plant Sci.* **16**, 363–371 (2011).
10. Tanksley, S.D. & McCouch, S.R. *Science* **277**, 1063–1066 (1997).
11. Rafalski, A. *et al. Curr. Opin. Plant Biol.* **13**, 174–180 (2010).
12. Bernardo, R. & Yu, J. *Crop Sci.* **47**, 1082–1090 (2007).
13. Heffner, E.L., Sorrells, M.E. & Jannink, J.-L. *Crop Sci.* **49**, 1–12 (2009).
14. Varshney, R.K., Nayak, S.N., May, G.D. & Jackson, S.A. *Trends Biotechnol.* **27**, 522–530 (2009).
15. Davey, J.W. *et al. Nat. Rev. Genet.* **12**, 499–510 (2011).
16. Varshney, R.K. & May, G.D. *Brief. Funct. Genomics* **11**, 1–2 (2012).
17. Elshire, R.J. *et al. PLoS ONE* **6**, e19379 (2011).
18. Bernardo, R. *Crop Sci.* **48**, 1649–1664 (2008).
19. Belo, A. *et al. Mol. Genet. Genomics* **279**, 1–10 (2008).
20. Yan, J. *et al. Nat. Genet.* **42**, 322–327 (2010).
21. Tian, F. *et al. Nat. Genet.* **43**, 159–162 (2011).
22. Poland, J.A., Bradbury, P.J., Buckler, E.S. & Nelson, R.J. *Proc. Natl. Acad. Sci. USA* **108**, 6893–6898 (2011).
23. Chia, J.-M. *et al. Nat. Genet.* **44**, 803–807 (2012).
24. Huang, X. *et al. Nat. Genet.* **42**, 961–967 (2010).
25. Xu, X. *et al. Nat. Biotechnol.* **30**, 105–111 (2012).
26. Zhao, K. *et al. Nat. Commun.* **2**, 467 (2011).
27. Huang, X. *et al. Nat. Genet.* **44**, 32–39 (2012).
28. McMullen, M.D. *et al. Science* **325**, 737–740 (2009).
29. Cook, J.P. *et al. Plant Physiol.* **158**, 824–834 (2012).
30. Myles, S. *et al. Plant Cell* **21**, 2194–2202 (2009).
31. Korte, A. *et al. Nat. Genet.* **44**, 1066–1071 (2012).
32. Segura, V. *et al. Nat. Genet.* **44**, 825–830 (2012).
33. Salvi, S. & Tuberosa, R. *Trends Plant Sci.* **10**, 297–304 (2005).
34. Xing, Y. & Zhang, Q. *Annu. Rev. Plant Biol.* **61**, 421–442 (2010).
35. Salvi, S. *et al. Proc. Natl. Acad. Sci. USA* **104**, 11376–11381 (2007).
36. Heslot, N., Yang, H.-P., Sorrells, M.E. & Jannink, J.-L. *Crop Sci.* **52**, 146–160 (2011).
37. Varshney, R.K., Close, T.J., Singh, N.K., Hoisington, D.A. & Cook, D.R. *Curr. Opin. Plant Biol.* **12**, 202–210 (2009).
38. Varshney, R.K. *et al. Nat. Biotechnol.* **30**, 83–89 (2012).
39. Salekdeh, G.H., Reynolds, M., Bennett, J. & Boyer, J. *Trends Plant Sci.* **14**, 488–496 (2009).
40. Walter, A., Studer, B. & Kolliker, R. *Ann. Bot. (Lond.)* **110**, 1271–1279 (2012).

A prototrophic deletion mutant collection for yeast metabolomics and systems biology

To the Editor:

Auxotrophic markers—mutations in genes encoding enzymes in pathways for the biosynthesis of metabolic building blocks, such as amino acids or nucleotides—are used as selection markers in most yeast genetics and genomics experiments[1–3]. The nutritional deficiency caused by the mutation (auxotrophy) can be compensated by supplying the required nutrient in the growth medium. This compensation, however, is not necessarily quantitative because such mutations influence several physiological parameters and may act in combination[2,4,5]. The construction of a prototrophic derivative of the parent strain of the widely used genome-scale yeast deletion collection[1] has confirmed the need to remove auxotrophic markers to reduce bias in physiological and metabolic studies[2]. Moreover, flux balance analyses using a genome-wide metabolic model (Yeast 5)[6] indicate that the activity status of some 200–300 reactions changes between different auxotrophic strains and the wild type. To alleviate this bias, we have constructed a version of the haploid deletant collection restoring prototrophy in the genetic background and thus preventing any influence of auxotrophy on the phenotype of a given gene deletion. This new deletant library is based on the popular S228c (MAT**a**) knock-out collection[1], and facilitates the exploitation of prototrophic yeast in both functional genomics and quantitative systems biology.

We assessed the physiological effect of auxotrophy by monitoring the growth of 16 yeast strains carrying all possible combinations of the markers (histidine (*his3Δ*; *his3Δ306-495, his3Δ1*), leucine (*leu2Δ*), methionine (*met15Δ*) and uracil (*ura3Δ*)) used in the MAT**a** version of the yeast deletion collection[1]. All markers and their combinations affected yeast growth,

but without altering the adenylate (ATP, ADP and AMP) energy charge (**Fig. 1a**). As the most critical phenotypic quantity, the maximum specific growth rate (μ_{max}) varied between 0.125 ± 0.003 (s.d.) h^{-1} (*leu2Δ*) and 0.207 ± 0.007 h^{-1} (*ura3Δ his3Δ*), rendering quantitative comparisons among these strains impossible (**Fig. 1a** and **Supplementary Table 1**). These growth differences could not be explained by the different medium supplementations for the following reasons: first, prototrophic yeast exhibited a different and substantially less diverse growth pattern in the 16 minimal media (**Fig. 1b**, left; medium recipes are given in the **Supplementary Methods**); and second, growth differences were altered, but not abrogated, when other proteogenic amino acids were included as well (synthetic complete medium; **Fig. 1c**). On both types of medium, we observed complex (epistatic) interactions among all auxotrophic mutations. For instance, restoring *MET15* had a negative effect on μ_{max} in *leu2Δura3Δhis3Δmet15Δ* (0.185 ± 0.004 $h^{-1} \rightarrow 0.164 \pm 0.003$ h^{-1}) or *leu2Δura3Δmet15Δ* (0.162 ± 0.005 $h^{-1} \rightarrow 0.149 \pm 0.001$ h^{-1}) but, unexpectedly, promoted growth in *leu2Δhis3Δmet15Δ* (0.136 ± 0.006 $h^{-1} \rightarrow 0.173 \pm 0.009$ h^{-1}; **Fig. 1a**); restoring *LEU2* had a positive effect in *leu2Δura3Δhis3Δ* (0.164 ± 0.003 $h^{-1} \rightarrow 0.185 \pm 0.006$ h^{-1}) or *leu2Δhis3Δmet15Δ* (0.136 ± 0.006 $h^{-1} \rightarrow 0.161 \pm 0.004$ h^{-1}) but not in *leu2Δura3Δhis3Δmet15Δ* (0.185 ± 0.004 $h^{-1} \rightarrow 0.186 \pm 0.007$ h^{-1}; **Fig. 1a** and **Supplementary Table 1**). Thus, although blocking different pathways, all markers affect one another, indicating that they have a wide-ranging and combinatorial influence on the metabolic network.

In batch culture experiments, further problems arise from the unequal consumption of amino acid supplements, resulting in cultivation phase–dependent