**Review**

**Towards the Genomic Basis of Local Adaptation in Landraces**

Giandomenico Corrado * and Rosa Rao

Dipartimento di Agraria, Università degli Studi di Napoli, via Università 100, Portici (NA) 80055, Italy; rao@unina.it

* Correspondence: giandomenico.corrado@unina.it

Received: 27 September 2017; Accepted: 2 November 2017; Published: 4 November 2017

**Abstract:** Landraces are key elements of agricultural biodiversity that have long been considered a source of useful traits. Their importance goes beyond subsistence agriculture and the essential need to preserve genetic diversity, because landraces are farmer-developed populations that are often adapted to environmental conditions of significance to tackle environmental concerns. It is therefore increasingly important to identify adaptive traits in crop landraces and understand their molecular basis. This knowledge is potentially useful for promoting more sustainable agricultural techniques, reducing the environmental impact of high-input cropping systems, and diminishing the vulnerability of agriculture to global climate change. In this review, we present an overview of the opportunities and limitations offered by landraces' genomics. We discuss how rapid advances in DNA sequencing techniques, plant phenotyping, and recombinant DNA-based biotechnology encourage both the identification and the validation of the genomic signature of local adaptation in crop landraces. The integration of ‘omics’ sciences, molecular population genetics, and field studies can provide information inaccessible with earlier technological tools. Although empirical knowledge on the genetic and genomic basis of local adaptation is still fragmented, it is predicted that genomic scans for adaptation will unlock an intraspecific molecular diversity that may be different from that of modern varieties.

**Keywords:** genomics; differentiation; genome-environment association

---

1. **Crop Landraces**

Public awareness on the importance of biodiversity conservation is strengthening over time [1]. Climate change, pollution, environmental disasters, loss of natural habitats, environmental degradation, and overexploitation of resources regularly make front-page news. Without taking into consideration the impact of the measures implemented to avoid biodiversity loss, large attention is generally given to wild species, especially those at risk of extinction [1,2]. Agricultural biodiversity (i.e., the variety and variability of animals, plants, and microorganisms that are used directly or indirectly for food and agriculture [3]) is largely regarded as a subset of biodiversity. However, agriculture and biodiversity are closely tied. Their mutual dependence is crucial not only to ensure yield today, but also to contribute to a more resilient, sustainable agriculture. This includes the development of solutions for water-saving technologies and for minimizing the detrimental effects of global climate change on crops [4,5].

Plant genetic resources for food and agriculture (PGRFA) are the central components of agricultural biodiversity because they constitute the primary elements of the production process. Crop improvement relies on genetic diversity. Taking into account the trends and efforts of modern breeding, the main part of genetic diversity of cultivated species is expected to be found in traditional varieties, also known as landraces. It is not easy to provide an all-purpose definition of landraces...
because of their complex nature. Different classifications have been proposed in the literature [6–8]. For instance, the apparently simple distinction between autochthonous and allochthonous landraces is not easy to put in practice, considering that it is difficult to clearly identify geographical boundaries and define in quantitative terms a “recent introduction”. In spite of different definitions, human management is integral to the development and maintenance of landraces. Moreover, there is a consensus that plant landraces are dynamic populations that may be particularly adapted to certain environments. Other features that are usually attributed to landraces are yield stability, adaptability to sustainable farming, and resilience to stress, although these characteristics are expected to be evident especially in low-input agricultural systems [6–8]. For instance, submergence survival in rice landraces was retained as a beneficial trait in local ecosystems [9]. Similarly, a rice variety originating from regions with poor soil (e.g., phosphorus-deficient lowlands) was used to isolate the phosphorus-starvation tolerance 1 (PSTOL1) gene, which is absent in modern varieties [10,11].

While landraces are widely employed in low-income countries, in high-income countries they are associated with traditional or amateur farming, and niche products [12]. In advanced economies, landraces usually receive attention because of the consumers’ perception of food production of higher quality. Traditional agricultural products have a prominent role in supporting social, historical, and cultural identity, and are becoming increasingly appealing [13]. Trust, transparency, uniqueness, and authenticity are central drives of today’s consumers, especially those who have a wealth of resources at their fingertips. The perceived authenticity of a food product is usually connected with its origin and culture, including traditional cooking and specific industrial transformations [14].

The myriad of landraces, which constituted the cultivated genetic material for millennia, has been progressively displaced by modern cultivars in almost all the agricultural settings of the high-income countries [15]. The genetic erosion of cultivated material in southern Italy was estimated to be over 70% in terms of collected samples [16]. Modern agriculture, including plant breeding, has been frequently evoked to explain the incessant erosion of PGRFA [17]. In essence, plant breeding represents a fast evolutionary process to develop improved varieties, and in the last decades, it has been largely based on elite breeding pools [18]. Although new genetic diversity has introduced exploiting crop wild relatives, gene flow, or mutation, plant breeding is usually accompanied by loss of allelic diversity. Trait uniformity and stability are an essential target for breeders, not only for technical reasons related to current cropping systems, but also for protection purposes. Moreover, elite breeding populations move towards a reduced heterozygosity because of inbreeding and the random fluctuation in gene frequency (genetic drift) associated with small, effective population sizes [18].

Plant domestication and breeding have greatly increased food quality by removing unpleasant characteristics (e.g., excessive bitterness, sourness, pungency, toxic compounds, spikiness, hairiness, etc.) and favored others (e.g., sweetness, attractive color, relative amount of flesh or pulp, regular fruit shape, etc.) [19]. Although for different vegetables (e.g., tomato, summer squash, pepper, etc.), fruit size, color, or shape present a morphological variety absent in wild species [20], plant breeding may have unintentionally diminished fruit quality in exchange for production traits [21–23]. Selection for high yield may have reduced the relative amount of main components of fruit taste, such as sugar, aldehydes, and volatile organic compounds (VOC) [24]. Some Italian tomato landraces displayed higher level of metabolites related to fruit quality [25] and a potentially useful phenotypic variability that deserves a further genetic characterization [26]. Landraces can also have superior technological aptitude, which can be exploited for typical, highly-valued products, such as the San Marzano tomato tins and the Portuguese high quality maize bread [27,28].

2. Landraces as a Source of Local Adaptation

Excluding the cultural value, resilience, nutrition content, sensorial value, and compliance to low-input farming are traits that are controlled often by multiple genes. In landraces, these features constitute the bulk of their local adaptation (LA) because they are the target of main farmer-mediated evolutionary forces.
LA is a process by which a population becomes better suited to its local environment than other members of the same species [29]. Although LA is frequently linked to climate, there is ample evidence that plants adapt to different environmental elements, including biotic and non-climatic abiotic factors [30,31]. In ecology, LA is usually measured as the difference in fitness between a population in its environment or growing elsewhere, or by comparing the fitness of a local and an introduced population in one environment [29]. In agriculture, fitness may not necessary be a trait of interest, and LA usually relates to crop yield, or more generally, to a phenotype in response to certain environments and agricultural practices [32]. For instance, differences in maize landraces have been measured that considered phenological, morphological, or physiological traits related to yield [33–35].

While LA in plants has been found in a number of studies [36], relatively less is known on local adaptation due to farming. Farmer-mediated selection does not always go along with the more frequently reported environmental selection, although fertility, fitness, and yield are usually interdependent [37]. Adaptive divergence in quantitative traits is negatively correlated with the rate of population mixing [38], which suggests that LA should be also common in recently introduced landraces (i.e., a relatively reduced number of generations experiencing specific selective regimes). On the other hand, gene flow between populations, which is generally seen as a disruptive force for LA in the absence of strong adaptive selection [39], is more frequent for landraces within some centers of origin [40].

Establishing how populations respond to environmental conditions is not an easy challenge in agriculture as well as evolutionary biology [41]. The positive features of landraces are often linked to specific growing techniques and environmental conditions that cannot be easily reproduced in experimental stations, such as low-fertility soil [42]. For instance, PSTOL1 enhances grain yield in phosphorus-deficient soil [11]. In maize, LA is often associated with altitude, with highland landraces poorly performing in lowland areas, and vice versa [43,44]. For these reasons, reciprocal transplants to reveal underlying factors of LA may not be straightforward, especially for crop landraces adjusted to low-input farming or originating from marginal lands, which are of great interest as a source of adaptive traits useful to increase stress resistance [15]. Moreover, LA may also lead to an adaptability trade-off, and some landraces can adapt to a wide range of environments, whereas others can only adapt to a few environments [6,45].

To grasp adaptive diversity for current and future challenges, it is necessary to identify locations where agronomic and/or historical climatic conditions match predicted changes. The comparison of isolated landraces in similar yet geographically isolated environments would also provide the possibility to understand whether convergent solutions to a specific stress are established at the phenotypic and/or genetic level [46]. At least in natural populations, the probability of gene re-use in parallel or convergent phenotypic evolution was considered high [47].

The study of the phenotypic adaptation also requires establishing whether the observed phenotypic differences between populations are primarily genetically based or the result of phenotypic plasticity (i.e., a plastic response to the environment that does not require genetic changes) [48,49]. Adaptive phenotypic plasticity is typically associated with the magnitude of response of quantitative traits in relation to the environment [50]. However, yield stability across a range of conditions, which is a feature of many landraces [6], can be also a plastic response that is not always supported by a “distinct” phenotypic trait. Quantitative Trait Loci (QTLs), gene expression levels, and epigenetic mechanisms (i.e., those related to priming and acclimation) are likely to contribute to stress adaptation in plants [50,51]. For instance, DNA methylation has also been associated with drought resistance in upland rice, and to adaptation to higher altitude in maize landraces [52,53]. The adaptation of plant species and communities to global climate change is an important trait in breeding for a more climate-resilient agriculture, and is frequently associated with plasticity [54]. A meta-analysis indicated that evidence on the evolutionary adaptation to climate change is still relatively scarce [55].
3. Genomic Scans of Local Adaptation in Landraces

Landraces adaptive differentiation has been found in major crops, such as wheat, maize, rice, barley, and sorghum [56–60]. Uncovering the genetic basis of LA will shed light on the evolutionary forces acting on crops, and on the mechanisms underlying environmental adaptation, stress response, and yield. It is not fully clarified whether selection acts primarily on existing genetic variation within a crop (including introgression from modern varieties), or on new mutations specifically present in landraces’ populations [61,62]. In Arabidopsis, local climatic adaptation was associated with environmental-specific selection on existing variants, as well as hard-selective sweep (e.g., a rapid increase of the frequency of new beneficial mutations) [63,64]. In addition, it has also been shown that barley landraces have a mosaic ancestry, with multiple genomic segments from local wild populations that can contribute to adaptive variation [65].

The development of more cost-effective techniques to investigate genetic diversity at the large genomic scale makes it possible to identify sequence variations associated, and hopefully responsible, for superior crop performance in low-input and more sustainable farming systems. Briefly, two strategies are usually employed [66], and both are based on a comparative approach. The first relies on the identification of loci that display significantly different genetic differentiation among populations under the assumption that selection pressure differed [67–69]. This strategy can be applied irrespective of hypotheses about the causative role of the environment. Given the likely presence of some environmental constraint for the landrace, attention is usually given to outlier loci that are subject to positive selection, although positive directional selection may not necessarily increase intraspecific variability. A second strategy aims to correlate environments and genotypes under the assumption that a selective pressure creates associations between allelic frequencies (at the selected loci) and environmental variables. In essence, by analyzing allelic frequencies, it is tested whether a sequence variant, haplotype, or allele is significantly associated with a specific environment or environmental factor (if identified), while controlling for neutral genetic structure [58,59]. If the phenotype of interest and/or the genetic basis of the trait are known (e.g., stress resistance, metabolite production, etc.), the analysis of LA in crop landraces can be also carried out focusing on specific candidate regions or by genome-wide scans (e.g., linkage mapping or genome-wide association studies (GWAS), respectively) [59]. Association mapping in crop landraces (e.g., barley, common bean, soybean, durum, and common wheat) can reveal previously undescribed candidate regions associated with agronomic traits, including biotic stress resistance [70–75].

QTL mapping classically requires structured populations (e.g., recombinant populations deriving from phenotypically divergent inbred lines). The development of introgression lines (ILs) from rice landraces led to the identification of QTLs for yield components and the isolation, by map-based cloning, of an allele (NAL1) that increases grain productivity [76]. Scientific and technological advances have also enabled the exploitation of panels of unrelated cultivars or genotypes for QTL studies (i.e., non-candidate-driven association mapping approaches, such as GWAS). For instance, the molecular and phenotypic analysis of 723 wheat landraces revealed markers associated with previously unidentified QTLs relative to different traits [77].

Each method has its pros and cons, as discussed in the literature [66,78,79]. Essentially, neutral and demographic processes can generate correlations between the environment and the genotype that are difficult to distinguish from those arising from LA. Unfortunately, the availability of genomic data does not solve this problem.

For any genomic scan of LA, the type of DNA marker is invariably restricted to single-nucleotide polymorphisms (SNPs) because of the development of reasonably affordable high-throughput sequencing instruments. However, SNPs are not necessarily superior to other DNA markers, such as simple sequence repeats (SSRs) (also known as microsatellites or short tandem repeats, STRs), because the latter are more suitable to detect recent demographic events and private alleles due to their higher mutation rate and multiallelic nature, respectively. SNPs analysis based on next-generation sequencing (NGS) technologies, however, offers a possibility to investigate a number of polymorphisms that is
currently unmatched by any other approach, and strongly increases the possibility to identify adaptive loci [78]. The number of DNA polymorphisms under investigation is important, because it will affect the estimation of the population structure and the generation of the null distribution for statistical hypothesis testing. On the other hand, linkage and more generally, the non-independence of loci, are possible confounding sources that are likely to be more significant when analyzing a very large number of polymorphisms [66].

The genomic analysis of a landrace requires more extended sampling compared to a genetically uniform cultivated plant variety. Therefore, methods that ascertain sequence variations in a fraction of the genome are usually employed. Nonetheless, SNP genotyping chips and other reduced representation methods (for a list of methods, see [80]) are considered not fully adequate for identifying the genomic signatures of LA, because of their reduced genomic sampling power [66,78]. Moreover, SNP arrays are not very effective at capturing rare and previously undescribed variants in diverse genetic resources, and may suffer from ascertainment bias deriving from the SNP discovery process [80,81].

The availability of a reference genome for many crops [82] strongly facilitates the genomic analysis of landraces by resequencing [59,83]. Nonetheless, polymorphisms at the single nucleotide level cannot be considered sufficient to account for the whole LA, and it is necessary to analyze other more computationally demanding structural variations (SV), such as in/del, copy-number variation (CNV), and insertion of transposable elements (TEs). Adaptation to high boron concentration in wheat landraces is associated with multiple genomic changes, such as tetraploid introgression, gene duplication, and variation in gene structure and expression [84]. Different genes conferring resistance to stress (e.g., flooding and metal toxicity), firstly isolated in landraces, display gene CNV [9]. In maize landraces, more than half of the SNPs associated with altitude were within large structural variants (inversions, centromeres, and pericentromeric regions) [85]. A loss-of-function retrotransposon insertion led to adaptation to cultivation at high latitudes in a photoperiod-insensitive soybean landrace [86]. The de novo detection of SVs requires a deeper sequence coverage compared with the low-fold approaches usually employed in resequencing [87,88]. To overcome some of these limitations, a metagenome-like assembly strategy based on a low-coverage population sequencing data was employed for the construction of the dispensable rice genome as a more cost- and labor-effective strategy [89]. The recent availability of long-read sequencing technologies (also known as third-generation sequencing) can greatly improve the analysis of genome structure, not only for chromosome scaffolding and haplotype phasing, but also for the identification of long (e.g., >50 bp) structural variants [90].

The concept of “pan-genome” as the sum of the “core genome” (containing genes/sequences present in all strains) and the “dispensable genome” was first developed in microbiology, and later applied also to plants science [91,92]. At the least in some organisms, the “dispensable genome” significantly contributed to adaptation [91,93,94]. In maize and soybean, a substantial proportion of variation may lay in the “dispensable genome” [93,95]. It has been also suggested that the “dispensable genome” may have a role in the environmental adaptation in soybean [95].

For non-model and orphan species (i.e., those in which there has been little “omics” research), it is likely that genome-wide sequence analysis will be performed using a reduced representation method. Currently, genotyping by sequencing (GbS) represents one of the most affordable methodologies for SNP analysis in large populations. This approach is popular especially for GWA studies in crops because it can be also employed on plant species with complex and large genomes, including polyploids. A GbS-based survey of nucleotide diversity in soybean landraces revealed selective sweeps around starch metabolism genes; GWAS also provided insights into the origin and spread of haplotypes linked to agro-climatic traits [96].

When at least a reference transcriptome is available, exome sequencing, also known as whole-exome sequencing, may represent an affordable option for analyzing a well-characterized adaptive trait or very large plant populations [97], because of its reduced running cost when probes
are already available. For instance, an investigation of the barley genomic variability related to environmental conditions was carried out starting from the exome sequencing data of more than 250 georeferenced landraces and wild accessions [97].

The study of RNA molecules by the so-called transcriptomics technologies is also an opportunity to uncover genetic variants, with the added possibility of providing information on the molecular basis of adaptation because differences in transcript abundance are a component of phenotypic variation, especially in response to the environment. In maize, gene expression analysis by microarray has underlined that drought tolerant landraces more rapidly respond to stress compared with susceptible landraces [98]. RNA-Seq, also known as whole transcriptome shotgun sequencing (WTSS), is at present the most widely employed methodology for transcriptomics studies and it has overshadowed chip-based technologies. In a wheat landrace, RNA-Seq highlighted pathways and genes potentially related to resistance against *Fusarium* [99]. The analysis of transcribed or coding sequences also provides the possibility of detecting and coding landrace-specific allelic variations. Lastly, RNA-Seq can be also employed to detect allele-specific expressions in hybrids of cultivars and landraces, which can potentially contribute to adaptation [100]. However, RNA-based genomic scans cannot provide direct information on structural elements (e.g., regulatory sequences, as well as SVs), and may not have enough genomic resolution in very large genomes, unless linkage disequilibrium (LD) is high [78].

Finally, the integration of omics approaches (e.g., from proteomics and metabolomics to foodomics and nutrigenomics) can contribute to understanding the link between landrace-specific bioactive compounds, their importance, and DNA sequence variation. GWA mapping based on metabolomics data (mGWA) has been carried out in crops such as maize, tomato, and rice, which in some cases also exploit landraces [101–103].

4. Current Opportunities and Challenges

Conventional high-input agriculture faces diverse and complex challenges. It is necessary to promote the development and implementation of new agricultural techniques in order to mitigate the negative impact on soil conservation, water management, and biodiversity, as well as increase crop resilience to stress and adaptability to new areas. Agriculture will not overcome these challenges without modern (bio-) technology. Crop landraces represent a readily available resource to address these issues, because they are cultivated material already adapted to low-input agriculture, marginal lands, or stressful environments. For instance, maize landraces have an evolved adaptability to a wider range of environmental conditions than teosinte [104]. Identifying the loci involved in LA provides the possibility of not only defining the genomic basis of adaptation to specific conditions in crops, but also improving our understanding of some agriculturally important traits. Although genetic variability in landraces is considered lower than in wild relatives, different studies have underlined that it is higher than those of improved varieties [105,106]. More crucially, such variability should be readily available not only for germplasm improvement, but also for breeding. Different genes isolated from landraces have been successfully used for breeding programs in major crops [9].

The classic approach to detect LA is to verify whether the phenotypic divergence in candidate traits between populations cannot be explained by drift alone. Current advances in sequencing technologies encourage a genomic characterization of landraces. The genomic scan for LA represents today the first, most affordable step towards the exploitation of the positive features of crop landraces, mainly because of the resources needed for the phenotypic characterization of ample populations in different environments. Advances in high-throughput plant phenotyping facilities give reasons to believe that in the near future landraces characterization will be accelerated [107–109].

Irrespective of the approach and methodology employed, it is common that genomic scans for adaptation provide hundreds of candidate loci. Their identification is the foundation for understanding the physiological basis of adaptation. Therefore, a limiting factor towards the genomic basis of LA in landraces is the functional validation of these loci, an effort that should include the comparison of the trait of interest in near isogenic material in specific agricultural conditions. The correlation
between crop response in field and experimental conditions is a longstanding issue in agriculture, and it may be even more relevant for the validation of adaptive genetic variations [75]. The functional validation of candidate loci requires resources that often exceed the ones needed for the genomic analysis of LA, making necessary the selection of a limited number of variants by using additional computational methods [110,111]. Recent developments in plant biotechnology, including genome editing, give reason to believe that the validation of specific variants could be more easily achieved compared with more classic approaches, such as mutagenesis, genetic transformation, or the screening of natural and artificial populations [112]. However, the trait of interest may be highly polygenic, and LA may be the result of a number of relatively modest changes in allelic frequency that underlie or contribute to (unknown) phenotypic traits, as it is likely to occur for the natural variation of plant metabolites [113].

The identification of genetic loci and sequences responsible for LA will unlock the landraces’ diversity for precision breeding and plant science. Adaptation to rapidly changing climate conditions and to low-input sustainable agriculture will also require new varieties with, for instance, modified planting time or increased resilience, as well as the (assisted) migration of crops. Landraces are central for developing high-value plant varieties better suited to local conditions, especially for cropping systems that evolve towards a reduced use of off-farm inputs. For all of these reasons, understanding the genomic basis of LA in landraces has the potential to alleviate the environmental impact of agriculture in the near future.

Acknowledgments: This work was supported by the “Salvaguardia della biodiversità agroalimentare in Campania” (SALVE) project, Programma di Sviluppo Rurale per la Campania 2007–2013, misura 214 az. f2.

Author Contributions: Giandomenico Corrado and Rosa Rao wrote and reviewed the article.

Conflicts of Interest: The authors declare no conflict of interest.

References


23. Murphy, K.M.; Reeves, P.G.; Jones, S.S. Relationship between yield and mineral nutrient concentrations in historical and modern spring wheat cultivars. *Euphytica* **2008**, *163*, 381–390. [CrossRef]


43. Lafitte, H.; Edmeades, G. Temperature effects on radiation use and biomass partitioning in diverse tropical maize cultivars. *Field Crops Res.* 1997, 49, 231–247. [CrossRef]


70. Miklas, P.N.; Coyne, D.P.; Grafton, K.F.; Mutlu, N.; Reiser, J.; Lindgren, D.T.; Singh, S.P. A major QTL for common bacterial blight resistance derives from the common bean great northern landrace cultivar Montana No. 5. *Euphytica* **2003**, *131*, 137–146. [CrossRef]


75. Mamo, B.E.; Barber, B.L.; Steffenson, B.J. Genome-wide association mapping of zinc and iron concentrations in barley landraces from Ethiopia and Eritrea. *J. Cereal Sci.* **2014**, *60*, 497–506. [CrossRef]


112. Cardi, T. Cisgenesis and genome editing: Combining concepts and efforts for a smarter use of genetic resources in crop breeding. Plant Breed. 2016, 135, 139–147. [CrossRef]